

Doctoral Dissertation

Bennour Mohamed Hsin

Pécs, 2026

UNIVERSITY OF PÉCS
FACULTY OF BUSINESS AND ECONOMICS
Doctoral School of Regional Policy and Economics

Bennour Mohamed Hsin

Regional Technological Diversification:
Contextualising Relatedness

DOCTORAL DISSERTATION

Supervisors: Dr. habil. Kehl Dániel & Dr. Gyimesi András

Pécs, 2026

Table of contents

Abstract	1
Acknowledgement	3
List of Abbreviations	4
List of Figures	4
List of Tables	5
1 Introduction	7
1.1 General background	7
1.2 Problem statement	9
1.2.1 Relatedness, relatedness density, and diversification	10
1.2.2 Empirical consequences	10
1.2.3 Research problem	12
1.3 Objective	14
1.4 Research questions and hypotheses	15
1.4.1 RQ1: Is regional technological potential a better predictor of technology entry than relatedness density?	15
1.4.2 RQ2: How do technology-specific characteristics shape regional technological potential?	15
1.4.3 RQ3: Does industrial alignment condition the influence of technology characteristics on potential?	16
1.4.4 RQ4: Do national ecosystem characteristics explain variation in regional technological potential beyond dyadic portfolio mechanisms?	17
1.4.5 RQ5: How do aggregate portfolio characteristics and spatial interdependence structure regional potential at the portfolio level?	17
1.5 Structure	18
2 Literature review	19
2.1 Regional diversification and resilience: Overview from an Evolutionary perspective	19
2.2 Why place (still) matters?	22
2.3 Relatedness as the baseline, context as the extension	27
3 The interconnectedness between regions and technologies	34
3.1 Methodological motivation	34

Table of contents

3.2	The mainstream approach	37
3.2.1	Relatedness through co-occurrence	37
3.2.2	Structural limitations of co-occurrence-based relatedness	40
3.2.3	Relatedness density through linear aggregation	41
3.2.4	Structural limitations of linear aggregation	42
3.2.5	Implications for contextualisation	44
3.3	An alternative approach: Machine learning for contextualised relatedness .	45
3.3.1	Machine learning as a prediction framework	45
3.3.2	Why Random Forest	46
3.3.3	The Random Forest algorithm	48
3.3.4	Training procedure	51
3.4	Regional technological potential	52
3.5	Feature Importance Technology Space (FITS)	53
3.6	Coherence: Bridging technology networks and regional portfolios	57
3.7	Summary	64
4	Empirical design	67
4.1	Overview and identification goals	68
4.2	Research design	69
4.3	Data	72
4.3.1	Outcome: Regional technological potential	74
4.3.2	Technology–region level variables	76
4.3.3	Technology level variables	76
4.3.4	National level variables: GEI pillars	80
4.3.5	Region level variables	83
4.4	Model 0: Entry Benchmark	84
4.5	Model 1: Dyadic fixed effects models	85
4.6	Model 2: Hierarchical models	92
4.7	Model 3: Spatial panel models	93
4.7.1	Spatial Autoregressive (SAR) panel - primary specification	94
4.7.2	Spatial Error Model (SEM) - robustness	95
4.8	Summary	95
5	Results	97
5.1	Entry as a benchmark	98
5.2	Network effects and coherence moderation	101
5.3	National context and robustness	104
5.4	Hierarchical structures and spatial implications	107
5.5	Role of space	110
5.6	Summary	117
6	Summary	121
6.1	Discussion	121
6.1.1	On the methodology	121
6.1.2	On the results	126

Table of contents

6.1.3	Policy implication	130
6.2	Future research agenda	131
6.3	Conclusion	133
References		135

Abstract

In this dissertation we investigate the patterns of technological diversification in European regions. Specifically we inquire on the factors that contribute to the feasibility of a diversification option. Our focus in this work is to provide a context rich assessment that bridge different gaps in the literature namely relatedness, network positions, institutions, and the role of space. To operationalise this contextualised assessment we first address methodological and conceptual issues in the relatedness constructs.

Our methodological alternative is a machine learning training strategy that shift relatedness from a descriptive exercise to a predictive exercise. Our alternative approach enables us to construct a bidirectional network that's inherently more granular and informative than conventional technology space approaches. Similarly, our approach also enables us to quantify the feasibility of a specific technology for each region, we call this quantity regional technological potential.

With these two constructs, we divide technologies into two categories based on their role in the network structure, bridging technologies and enabling technologies. Then we investigate which technologies are more feasible, and whether the industrial alignment between the region's portfolio of technologies and a target technology moderate that relationship. Additionally, we incorporate national ecosystem co-variates that describe institutional and market conditions into our analysis to investigate their contribution in the feasibility of a diversification option. Lastly, we investigate whether or not these dynamics are space dependent and whether this dependency is stable at different effect channels.

Our results shows four main conclusions. First, we validate that our regional technological

Abstract

potential is an adequate predictor of entry to new technologies. Second we confirm that different network structures are strong predictor or technological potential. Third industrial alignment is our strongest predictor here and moderates the relationship between network structures and potential. Forth, we show that national entrepreneurial ecosystem characteristics are substantial predictors of regional technological potential and that their effect is stable but heterogeneous for different pillars. Fifth, and lastly, we confirm the spatial dependency at the regional portfolio level, and show that the average characteristics of these portfolios in terms of their endowment of bridging and enabling technologies are similarly strong and consistent predictors of aggregate regional potential and that industrial alignment remains a substantial moderation mechanism that's stable across all effect channels.

Acknowledgement

I dedicate this work to my parents. Thank you for raising me right. Thank you for being patient with me. Thank you for always being on my side. To my partner and best friend, without you by my side, this wouldn't have been possible, thank you.

To my Supervisors, thank you for the support, the encouragement, the guidance, and for believing in me. I'm immensely grateful, without your commitment and engagement this wouldn't have seen the light. Thank you for challenging me, for making me a better person, and a better researcher.

I'm equally grateful for the faculty, teachers and staff whose support was essential and who made – through their kind words and acts – even the darkest days shine bright.

I'm thankful for my friend who paved the way for me to get here. And for my colleagues, who made my journey full of memories I will forever cherish.

To my past self who was full of doubt, this is a testament of your hard work and dedication. The result of your struggles, discipline, and perseverance . Always move forward, that's the only way, everything else is temporary.

List of Abbreviations

Abbreviation	Definition
EEG	Evolutionary Economic Geography
REC	Relatedness and Economic Complexity
FITS	Feature Importance Technology Space
GPT	General Purpose Technology
KET	Key Enabling Technology
GEI	Global Entrepreneurship Index
NSE	National Systems of Entrepreneurship
RTA	Revealed Technological Advantage
RCA	Revealed Comparative Advantage
NUTS	Nomenclature of Territorial Units for Statistics
IPC	International Patent Classification
EPO	European Patent Office
SAR	Spatial Autoregressive Model
SEM	Spatial Error Model
LPM	Linear Probability Model
OLS	Ordinary Least Squares
RF	Random Forest
ML	Machine Learning
ICC	Intraclass Correlation Coefficient
S3	Smart Specialisation Strategy

List of Figures

3.1	Node correlation - Density vs ML approach Tacchella et al. 2023	48
3.2	FITS Technology Network (based on 2018 patent data)	54
3.3	Relatedness vs FITS Networks for H10L (based on 2018 patent data) . . .	57
3.4	Relatedness vs FITS Networks for D06Q (based on 2018 patent data) . . .	58
3.5	Bidirectional Embedddness: profile of a region vs a technology	61
4.1	Average Regional Technological Potential (2008 vs 2018)	75
4.2	Average Coherence Across Regions (2008 vs 2018)	77
4.3	Average Relatedness Density Across Regional Technology Portfolios	86
4.4	Average Betweenness Centrality of Regional Technology Portfolios	88
4.5	Average Eigenvector Centrality of Regional Technology Portfolios	89
4.6	Total Patent Count by Region (2008–2018 Averages)	90
4.7	Geographic Distribution of Regional Complexity Across European NUTS2 Regions (2008 vs 2018)	91
5.1	Local Indicators of Spatial Association (LISA)	114

List of Tables

3.1	Heterogeneity of coherence at similar potential deciles (2009)	60
4.1	Hypotheses, Variables, and Empirical Tests	73
4.2	Summary statistics of the Dyad level data	78
4.3	Correlation matrix of the Dyad level data	79
4.4	Correlation matrix of the GEI pillars	81
4.5	GEI Pillars Retained in Analysis	82
4.6	Descriptive statistics of the GEI pillars	83
4.7	Descriptive statistics of the aggregated variables	83
4.8	Correlation matrix of the aggregated variables	84
4.9	Spatial Autocorrelation Diagnostics	94
5.1	Regional Potential vs Relatedness density – Entry(year + 1, LPM with fixed effects)	100
5.2	Network Interactions – Regional Potential (with fixed effects)	102
5.3	Network Interactions and National Ecosystem – Regional Potential (with fixed effects)	106
5.4	Hierarchical Random Effects Model – Regional Potential	109
5.5	Spatial Autoregressive Model (SAR) - Regional Potential	111
5.6	Heterogeneity of average betweenness at different average coherence deciles	113
5.7	SAR effect decomposition - Regional Potential	116
5.8	Spatial Error Model (SEM) - Regional Potential	118

1 Introduction

1.1 General background

The Covid 19 pandemic showed structural challenges of national economies all over the world, specifically the fragility of neoliberal policies in times of crisis and the lack of industrial and economic resilience. Six years after the fact, our societies are now confronted with inevitable novel challenges and looming shocks. We are already witnessing the consequences of AI development as it paves the way for a new technological revolution that would render most local economies obsolete and cause massive unemployment in white collar sectors. Sadly, that doesn't seem to be the end of the shocks the world has seen recently as the war in Ukraine, Gaza, Iran, and most importantly the current, chaotic trade wars, seem to foster ever increasing uncertainties. Facing all this, policy makers are confronted with a simple choice; strategise and plan for more resilient local economies.

In different streams of literature, resilience is directly related to diversification/variety, whether in portfolio management in finance, trade partnerships/linkages, industrial activities, or in terms of knowledge as well. Thus we can equivalently say that resilience is the capacity to resist and/or adapt to external shocks by relying on existing internal capabilities that evolve in the face of such shocks. This means that for an economy to survive uncertainties, it needs to evolve, change, and innovate its way to the other end of the bleak challenges it's confronted with. However, to evolve, change and innovate, a baseline of knowledge should be leveraged since the consensus is that variety is a buffer against external shocks and shields uncertainties. This chain of thoughts, takes us back to the conceptual basics of sustainable economic development and growth; the knowledge

1 Introduction

fabric is what facilitates any long term strategy, and has been shown in the literature as clear catalyst for societal prosperity and economic resilience. For this reason the study of knowledge is essential for policy making, and understanding how to increase its diversity is more relevant than ever before.

Knowledge is a set of information that covers one or many topics, and its characteristics are contingent on the different forms it can take or how it was created, generally speaking, academia and businesses are the main knowledge creators in any economy through research and patents. Essentially, knowledge can be codified (accessible by anyone through any medium), or tacit (personal information based on social connections, intuition, experience, etc, that's hard to share with others). The consensus in the literature is that the main driver of competitive advantage for firms is the tacit form of knowledge, which is also widely acknowledged that it's space dependent. However, the knowledge produced by firms can be reliably seen in patents, although they capture codified information, they also reveal tacit knowledge and its geographic footprint in space. This means that its essential to assess the knowledge in patenting activities (we refer to this knowledge as technologies), and more so to focus on the local aspect of these activities i.e: sub-national regions.

This framing, however, is not at all new or novel. In fact, this is the entire aim of the literature of the geography of innovation; to study how innovation is created and diffused to different actors in different geographical contexts. Specifically, relatedness and economic complexity (REC) is one of the main streams of literature that focus on the relationships between activities and geographies. The conceptual and methodological framework that REC provides is widely used and adopted in academia and among policy practitioners and was one of the main contributors to the smart specialisation policy literature. The ideas embedded in this framework, to put it simply, rely on the premise of spatial dependence of tacit knowledge in local/regional economies/geographies and focus on simplifying these relationships using network science to model the relationships between knowledge and regions. These simplifications provide valuable scope of analysis, interpretation and empirical insights and are widely used for policy design. But the resulting loss of granular information limits how much regional and technological contexts can be incorporated

within these assumptions, and implies that there's much more conceptual, methodological, and empirical work needed. The reason for this is because the loss of information bias the empirical interpretation in the sense that we end up with a homogeneous implication with weak regards to the regional and national contexts as well as the technological characteristics. This work is motivated by this gap and aims to simply contextualise the study of knowledge (technological) diversification using the same granular information publicly available and commonly used in the REC literature. The idea is simple, account for endogenous and exogenous contexts using the same granular data often used in this regard in order to understand the contexts and contingencies that drives regional diversification. In here we limit the definition of diversification as what is possible for a region to do. With that, a diversification strategy/option would mean how feasible is a technology for a region to diversify into.

1.2 Problem statement

The main problem that this works aims to solve is directly embedded in the methodological and empirical framework of REC. REC models diversification through network aggregation based on co-location and co-occurrence patterns. Using these patterns, different aggregations are used to quantify relatedness (the frequency of observing a pair of activities in the same region), and relatedness density (how much of the activities frequently observed together a region has). However, these measures are often interpreted not as aggregations of frequent observations but rather as relationship models. Empirically, diversification is studied using these constructs as predictors of entry, that is a region's entry to a new specialisation in a new technology, often considered as a binary outcome. In here we briefly outline the big picture in the REC methodological and empirical framework, its conceptual issues, its empirical consequences, and highlight the research gap.

1.2.1 Relatedness, relatedness density, and diversification

Relatedness and relatedness density are essentially measures of proximity. In a sense they describe how close two technologies are close to each other, or how close a given technology is to a region given its portfolio of technologies. To further decompose the problem here, we will first establish the methodological constructs for proximity measures. For relatedness that's co-occurrence, and for relatedness density that's the linear aggregation of relatedness.

First, co-occurrence is essentially the frequency of observing two activities together. In the REC literature, this frequency describes the strength of the relationship. Activities frequently observed together are more related than the pairs rarely observed together.

Second, linear aggregation of relatedness essentially measures the percentage of co-located technologies in a region that are related to a reference technology. Thus, we can think of relatedness density as the link between related technologies and co-located technologies. The idea in the REC literature states that relative to a given technology, the more related technologies a region has, the more likely that it can develop that technology.

These two constructs are used together to predict the probability that a region will enter a new technology. The REC literature shows that relatedness density is consistently associated with higher probabilities in almost all studies. These results among others were one of the major latent contributors to smart specialisation strategy (S3) policy. Thus the consensus in the literature was clear: In order to diversify into new technologies with the highest likelihood of success, regions must prioritise investment in related technologies.

1.2.2 Empirical consequences

The idea of resilience is not a main focus for the REC literature nor it is ours. However, falling back to this concept allows us to further assess the empirical consequences of the mainstream interpretation of relatedness and relatedness density. The idea is that in order to be resilient to external shocks and subsequent uncertainties diversification is

1 Introduction

key. But what kind of diversification is required and feasible and how to achieve it is the focus here. The REC literature tells us that the most likely successful diversification strategy is the one that targets related capacity in the regions, often referred to as related variety. However, generalising this recommendation is not that straight forward. Aggregate regional capacities and their national and broader geographic contexts differ significantly. The initial landscape of the regional technological portfolio is essential here because this strategy could favour regions with already diverse portfolio but it's questionable that regions with limited portfolios would equally benefit. This is aligned with the concept of path dependency, related variety without context enforces that dependency and locks regions within their limited capacities. This brings us back to the core focus of in this work; context is key. However, context on its own here might not be enough because adding contextual factors to the existing framework would not resolve the underlying problem, because path dependency in related variety is a direct consequence of how relatedness and relatedness density are constructed. These constructs and their underlying methodological assumptions (co-occurrence and linear aggregation) are the core problem that we're highlighting here.

Co-occurrence assumes that technologies frequently observed together are likely related. Although this is within the boundaries of common sense, it's highly unlikely that's actually the case. A frequency measure is only informative when we have more observations than items – that is, more geographies than technologies. Almost always, we will have more technologies than geographies, this means that relatedness is at best noisy. Additionally, relatedness as interpreted in the literature quantifies the relationship between pairs of technologies. However, as it stands, there's no differentiation in the direction of that relationship thus, assuming that the relationship between two technologies is symmetrical. Albeit this assumption in itself is not problematic, it exacerbates the linearity issue when we measure relatedness density as we loose information in co-occurrence, symmetry, and linear aggregations. This takes us to the final issue we would like to highlight; relatedness density. Simply put, relatedness density measures the sum of technologies related to a reference technology present in a given region. The implicit assumption in here is that technologies are linked through linear combinations, and those combinations predict the likelihood of successful diversification. However, relatedness density is often interpreted as a value that

quantifies the existing requirements a region has relative to a technology, whereas the sum of existing related technologies do not inform us on the actual requirements.

In summary, relatedness and relatedness density measures suffer from diverse methodological issues embedded in the implicit assumptions in their core constructs. Co-occurrence and linear aggregation of observed frequencies are misinterpreted, accrue information loss, and poorly handles the granular data often used. This means that the empirical and methodological work ahead must account for these issues to further contextualise the study of diversification strategies. Our work here is a step in that direction by reforming these constructs in a first methodological phase then use them within a broader empirical application.

1.2.3 Research problem

In the light of all the mentioned in this section, we fall back again on the core idea that we started this text with; How can we contextualise diversification strategies? The answer to this question is multi-layered and complex. In this section we started by outlining the importance of diversification strategies for regions into new technologies via patenting activities. We explained that the REC literature provides interesting methodological and conceptual framework of analysis and showed that despite their usefulness they suffer from structural issues that limit the advantages of the used granular data, thus limit the incorporation of a broader context empirically. Essentially, the research problem we focus on here, is both methodological and empirical in nature. We highlighted the structural methodological issues as the research gap which will be the focus of our methodological and empirical contribution.

To address these methodological gaps, we first reform the measurement foundation itself. Rather than relying on co-occurrence and linear aggregation, we adopt a machine learning framework that treats relatedness as a predictive exercise — producing asymmetric, capability-based measures that preserve the directional dependencies and combinatorial structure that conventional relatedness discards. This reformed foundation yields two core

1 Introduction

constructs: a regional technological potential measure that replaces relatedness density, and a directed capability network that replaces the conventional technology space. These constructs are richer by construction because they capture what the region’s portfolio actually predicts about future specialisation, rather than what it co-occurs with. It is on this reformed foundation that the empirical investigation is built. Rather than asking what relatedness density predicts, we ask what conditions determine the feasibility encoded in regional technological potential.

Moreover, reforming the measurement foundations opens a richer set of empirical questions. Once we can characterise where a technology sits in the directed capability space, a natural question follows: does the value of that position depend on the region’s portfolio characteristics? The intuition here is that a region’s existing portfolio may be more or less organised around the capability dependencies of a given technology at the industry level. A region and a technology both have a profile. To compare their profiles, we measure how similar the industrial embeddedness of a technology (how strongly it’s linked to each industry) and the industrial embeddedness of a region (how strongly that region is linked to each industry). These measures are then summarised in single coherence metric which is measure of industrial alignment between a region and a technology.

The empirical implication of this industrial alignment is one of the central questions this work investigates. We refer to this alignment as coherence (or knowledge infrastructure), and return to it in detail in the next chapters.

Another point parallel to the above mentioned concerns context beyond technologies and regions. National institutional and market characteristics, mirrored by the entrepreneurial ecosystem are one such context. Whether this context contribute independently to regional diversification feasibility, beyond what portfolio mechanisms already explain, is another question this work investigates.

A final question concerns the geographic context of regions. Even after accounting for technology characteristics, portfolio alignment, and national ecosystem conditions, regions do not operate in isolation. Especially in the European context where (open) borders are merely an administrative construct. Thus the regional spatial dependencies and whether

and how they structure regional diversification potential at the portfolio level is the last contextual layer this work investigates. In here, we focus at the aggregate portfolio level, because this enables us to understand how the different characteristics of a portfolio influence the aggregate portfolio potential for a region, at this level, this relationship can also be interpreted as a mechanism in which the portfolio characteristics act as a catalyst for the feasibility of all other technologies in a region.

With this in mind, we propose a multilayered contextual framework in which we distinguish between the endogenous (technology-region fit) and exogenous (national, spatial factors) conditions. With that, we can identify which combinations enable feasible diversification across different regional contexts operationalised through this structured design:

Endogenous factors (within-region):

- Technology-specific network position; whether a technology plays a bridging or enabling role in the directed capability space.
- Portfolio alignment (coherence); the degree to which a region's industrial structure is organised around that of a given technology.

Exogenous factors (beyond-region):

- National entrepreneurial ecosystem; the institutional and market characteristics that influence diversification feasibility.
- Spatial portfolio structure; the aggregate technological endowment of a region and its interdependence with neighbouring regions' technological potential.

1.3 Objective

The objective from this work is to extend the methodological framework of relatedness starting from the structural issues it suffers from. In order to show how our methodological contribution benefit the study of regional technological diversification we rely on an empirical study that shows how broader context can be included and how such context inform granular

policy insights. The broader context we aim to include empirically relied on relaxing the implicit relatedness assumptions, explicitly include the endogenous characteristics of the technologies and regions contingent on the regional knowledge infrastructure while simultaneously account for the characteristics of the national ecosystems.

We organise the empirical investigation around four contextual layers. The first concerns the technology’s structural position in the capability network, whether a technology plays a bridging or enabling role shapes how feasible it is for a region to diversify into it. The second concerns portfolio alignment, whether the region’s existing industrial structure is coherent with the target technology’s conditions how much value those network positions deliver. The third concerns the national ecosystem, institutional and market conditions at the national level may amplify or constrain the feasibility landscape independently of portfolio mechanisms. The fourth concerns space, regional potential may not be locally contained, and the aggregate characteristics of a region’s portfolio, together with the potential of its neighbours, could structure feasibility at the portfolio level.

1.4 Research questions and hypotheses

1.4.1 RQ1: Is regional technological potential a better predictor of technology entry than relatedness density?

H1: Regional technological potential is a better predictor of entry than relatedness density is.

1.4.2 RQ2: How do technology-specific characteristics shape regional technological potential?

Technologies occupy structurally distinct positions in the capability network, bridging and enabling roles imply different relationships to diversification feasibility. We ask whether these structural roles predict how feasible it is for regions to diversify into them.

H2a: Technologies with a bridging role are positively associated with regional technological potential, meaning that bridge technologies represent more accessible diversification pathways for regions.

H2b: Technologies with an enabling role are negatively associated with regional technological potential, meaning that foundational technologies represent structurally more difficult diversification targets.

1.4.3 RQ3: Does industrial alignment condition the influence of technology characteristics on potential?

The value of a technology's structural position may depend on whether the region's portfolio is aligned with the industrial level capabilities that position implies. We ask whether coherence moderates the relationship between network position and potential, and whether this moderation operates differently for bridging and enabling technologies.

H3a: Industrial alignment (coherence) between a region and a technology is associated with higher regional technological potential.

H3b: Technologies with a bridging role are positively associated with higher potential for regions with low coherence in them, than for regions with high coherence, meaning that bridge technologies matter most as diversification pathways when direct alignment is limited.

H3c: Technologies with an enabling role are positively associated with higher potential for regions with high coherence in them, than for regions with low coherence, meaning that enabling technologies matter most as upgrading pathways when direct alignment is high.

1.4.4 RQ4: Do national ecosystem characteristics explain variation in regional technological potential beyond dyadic portfolio mechanisms?

Portfolio configuration describes the endogenous opportunity structure of a region, but whether those opportunities are realised may depend on enabling conditions at the national level.

H4: National entrepreneurial ecosystem characteristics shape regional technological potential beyond what dyadic portfolio and network mechanisms account for. Meaning that the national institutional and market context is an independent layer of the feasibility landscape.

1.4.5 RQ5: How do aggregate portfolio characteristics and spatial interdependence structure regional potential at the portfolio level?

Moving from the technology–region dyad to the regional portfolio as a whole, we ask whether the aggregate composition of a region’s portfolio predicts average regional potential, and whether this relationship operates through spatial interdependence with neighbouring regions.

H5a: Regions with higher average potential tend to neighbour other high-potential regions, meaning that regional feasibility landscapes are spatially interdependent rather than locally contained.

H5b: Regions with portfolios concentrated in bridging technologies have higher average potential when their portfolio coherence is sufficient, meaning that the portfolio-level value of bridging technologies depends on whether the aggregate industrial alignment supports them.

H5c: Regions with portfolios concentrated in enabling technologies have higher average potential which increases as coherence increase, meaning that enabling technology endowments are a structural portfolio-level advantage, amplified further by coherence.

1.5 Structure

We structure this work around our methodological and empirical contributions. Given the complex nature of the problems we aim to solve, we will first start with more literature context that underlines economic geography, geography of innovation, and more importantly relatedness and economic complexity literatures. We then outline our approach in the methodology chapter where we discuss in more details measures of relatedness and proximity and provide alternative conceptualisation to modeling the relationships between technologies and regions. In the empirical chapter we outline the research design and the modeling procedures that implements our methodological contribution. The results chapter outline the results of our work and its empirical consequences and interpretation, which we discuss in more detail in the discussion and conclusion of this work in which we also outline different further extensions and future research directions.

2 Literature review

2.1 Regional diversification and resilience: Overview from an Evolutionary perspective

In the previous chapter, we framed this work’s core idea around the need to study technological diversification patterns of regions within a broader context. To justify this framing we grounded this core idea on the ever increasing importance of resilience facing an increasingly uncertain global and geopolitical world order (Xiao, Boschma, and Andersson 2018). Although resilience in itself is not a core focus of this work, we will use it an adjacent conceptual paradigm to justify our core ideas conceptually, methodologically and empirically through the literature.

Within the literature of evolutionary economic geography (EEG), P.-A. Balland, Rigby, and Boschma (2015) summarize this turn toward resilience and highlight that, while interest has grown, “the resulting debate suggests [that economic geographers] are far from a common understanding”. This is to say that resilience as a concept is attracting more attention, however, there’s still no consensus on a generalised ontological framework that drives further research in this area. The different studies that addressed this concept (eg: Christopherson, Michie, and Tyler (2010); Fingleton, Garretsen, and Martin (2012); Martin (2012); Pendall, Foster, and Cowell (2010); Pike, Dawley, and Tomaney (2010); Simmie and Martin (2010)) relatively converge in refuting equilibrium based approaches that consider resilience as an immediate reaction to shocks, and calling for further focus on an evolutionary approach. The argument for this proposed agenda is related to robust

results showing that precedent regional capabilities are recycled over time with different configuration driving different regional growth paths (Boschma, Balland, and Kogler 2015; Neffke, Henning, and Boschma 2011). However, P.-A. Balland, Rigby, and Boschma (2015) also criticises the literature in this context for ignoring key explanatory dimensions that are central to evolutionary accounts of regional change; technological relatedness, network structures, and institutions.

Recent syntheses also stress that capability accumulation is a necessary basis for adaptation, but not a sufficient explanation of successful upgrading, which is conditioned by technological trajectories, shifts in demand, and the evolving capabilities of competitors (C. Pinheiro 2025; Perez and Soete 1988; K. Lee and Malerba 2017). We adopt this triad as an organising device: relatedness and network structure describe the endogenous opportunity structure of regional capabilities, while institutions shape the exogenous enabling conditions. A useful way to make that second part operational without reducing institutions to a single rule of law proxy, is to treat entrepreneurship and institutions as a coupled system that governs how knowledge is mobilised into new ventures and new activities. This is precisely the motivation behind the “National Systems of Entrepreneurship” (NSE) perspective, which defines entrepreneurship systems as resource-allocation systems driven by individual opportunity pursuit but regulated by country-specific institutions (Acs, Autio, and Szerb 2014).

This agenda matters for the present study because our goal is not to address regional resilience per se, but to explain the mechanisms that shape diversification feasibility in a way that is consistent with an evolutionary view of regional adaptation. Some work has been done in this context notably Tóth et al. (2022) who address the relatedness and network structures gaps. The authors anchor their view on resilience by considering it an “adaptive ability of regions to reconfigure their economic structure” and that this ability could be observed in diversification patterns. This intuition was addressed and confirmed in Rocchetta et al. (2022) for technological diversification patterns and for industrial diversification patterns in Xiao, Boschma, and Andersson (2018). However, Tóth et al. (2022) retain a density-based operationalisation of relatedness. This choice,

as Tacchella et al. (2023) demonstrate systematically, generates measures that perform worse than trivial autocorrelation strategies (i.e., simply predicting that a region’s current specialisation portfolio will remain unchanged) in predicting genuine new specialisations—a major point we discuss further in subsequent chapters. Moreover, neither Tóth et al. (2022) nor Rocchetta et al. (2022) incorporate institutional or ecosystem conditions as contributing factors¹. This leaves open the question of whether network structure shapes feasibility unconditionally or in interaction with the enabling environment.

Against this background, we adopt an evolutionary framing that treats diversification as a constrained, path-dependent process of recombination. Empirically, this motivates a more contextualised assessment of regional capabilities and a decomposition of diversification drivers along three dimensions that the literature highlights as consequential: relatedness, network structure, and higher-level context (including national entrepreneurial ecosystems and spatial interdependence). This framing carries particular policy urgency: as C. Pinheiro (2025) demonstrates, the dominant tools currently deployed to advise regional and national governments on diversification strategy—Relatedness and Economic Complexity (REC)—carry theoretical and empirical limitations that are consequential precisely in the kinds of catching-up or restructuring contexts where resilience is most needed. Crucially, this framing also carries methodological implications. If diversification is conditioned by portfolio-level configurations of capabilities rather than pairwise proximities, then measuring feasibility as a function of the whole capability basket—rather than as a sum of dyadic relatedness scores—is not merely a technical refinement but a theoretical necessity (Tacchella et al. 2023; C. Pinheiro 2025). This is the foundation on which the present study’s empirical strategy rests. Understanding what shapes diversification feasibility in more granular context, therefore becomes not only an academic concern but also a critical policy question.

¹Xiao, Boschma, and Andersson (2018) incorporated a quality of governance measure that showed no consistent significant effect in their study

2.2 Why place (still) matters?

The agenda for assessing regional resilience through an evolutionary lens through relatedness, network structures and institutions, falls on the theoretical convergence that regional economies can be considered as capability systems (Hidalgo and Hausmann 2009; Hausmann and Hidalgo 2011; Martin and Sunley 2007) where a region’s economy can be represented as a structured set of complementary capabilities that constrain which new activities are feasible. To approximate the interconnectedness between these capabilities in a regional economy, interpreting this system as a network is the consensus here (Tóth et al. 2022). In this sense, a region’s economy can be described as a capability network where nodes are technologies linked through how related they are (Boschma 2015; Whittle and Kogler 2020) and where the links represent empirical co-occurrence, co-specialization, or inferred dependence. Consequently proximity, centrality, and distance measures (among others) encode potential recombination pathways. The idea here is that such networks reflect the knowledge resources and dependencies of the region and how they are used in different configurations operationalised via a constrained search process (Napolitano et al. 2018).

From here, the direction is clear, the information contained in these networks should predict how these configuration result in different outcomes for different regions, as from this perspective place matters. Because of these place-dependent constraints of the accumulated regional knowledge structure, this search process (of new diversification opportunities) is inherently path dependent, i.e. future technological capabilities occurs near previous accumulated ones. This means that path dependency for related diversification is a feature of these network structures.

In light of this logic, Bathelt and Storper (2023) acknowledge the theoretical merits of these arguments, but provide a set of critical points often omitted in the literature. First, related diversification via relatedness measures are “based on a statistical artifact” and inferred via “another artifact: the industrial [technological] classification systems”. Second, this local related variety of diversification patterns predict the local infrastructure, but do not capture the dimensions of the learning process that produced them. Third, these

patterns as measured via relatedness tend to confound scale and become instead a proxy of the size of the regional economy itself. These concerns were also highlighted in Whittle and Kogler (2020) where the authors qualify them as a definitional bottleneck that confounds different mechanisms “making an all-purpose interpretation problematic”.

These critiques matter for the present study, not because they invalidate the relatedness tradition, but because they clarify what relatedness can and cannot do on its own. Measures based on observed co-occurrence in patent or industry classifications can successfully summarize regularities in diversification patterns, yet their meaning is partly constructed by data and taxonomy choices (Bathelt and Storper 2023; Boschma 2017; Whittle and Kogler 2020). In particular, if relatedness proxies are sensitive to classification “artifacts” and scale effects, then a purely dyadic “proximity $\rightarrow$ entry” story risks conflating (i) a region’s size and measurement resolution with (ii) the underlying learning and capability-building processes that actually generate diversification (Bathelt and Storper 2023; Boschma 2017). Additionally, from an industrial-policy perspective, this implies that relatedness-based recommendations should be treated as context sensitive rather than universal, especially in catch-up settings where historically successful upgrading sometimes required assembling capabilities beyond locally ‘related’ trajectories (C. Pinheiro 2025; Amsden 1989; K. Lee and Malerba 2017; Wade 2018). This motivates a shift from treating relatedness as a self-sufficient explanation toward treating it as a useful but incomplete representation of constraints—one that must be complemented with clearer mechanisms and context.

A key step in that direction is to reintroduce agency explicitly. Regions do not diversify by themselves: “a city and its surrounding region do not act for themselves, but [...] depend on firms and entrepreneurs to introduce new activities” (Neffke et al. 2018). This agent-centered view is consistent with evolutionary accounts of change (variation, selection, retention) (Nelson and Winter 1985; Martin and Sunley 2007), and it clarifies why the same relatedness structure can lead to different outcomes across places. Neffke et al. (2018) frame this using a resource-based view applied to regions, where places are endowed with capability bases that shape feasible paths of development (Lawson 1999; Boschma 2004; Penrose 2009). While branching evidence supports relatedness-constrained recombination,

novelty and structural change often arise through new establishments (especially those with nonlocal roots) rather than incumbents, implying that feasibility depends on both opportunity structure and the actors attempting entry (Neffke, Henning, and Boschma 2011; Boschma, Balland, and Kogler 2015; Hidalgo et al. 2007; Neffke et al. 2018). In this view, relatedness and network position describe an opportunity structure, what is plausibly feasible given a region’s existing capabilities. Whereas higher-level context governs whether feasible options become realised diversification outcomes. This distinction aligns with recent industrial-policy discussions that treat capability proximity as informative but incomplete unless combined with institutional and market conditions (C. Pinheiro 2025; Andreoni and Chang 2019).

This perspective also sharpens the link between diversification and adaptation: beyond incremental “branching,” structural change is “unrelated regional diversification that requires a transformation of the local capability base” (Neffke et al. 2018). Put differently, adaptive diversification is not only about whether a region adds new technologies, but about whether it can renew or extend its capability base to support activities that are less aligned with its inherited structure (Martin 2012).

This agent-based framing also helps reconcile “path dependence” with the empirical reality of novelty. The branching literature shows that diversification frequently occurs near what regions already do, consistent with relatedness-constrained recombination (Neffke, Henning, and Boschma 2011; Boschma, Balland, and Kogler 2015; Hidalgo et al. 2007). Yet innovation and new paths can also emerge from recombinations that bridge previously distant knowledge domains, and from new entrants that introduce different routines and connections (Martin and Sunley 2007; Isaksen and Trippel 2017). Neffke et al. (2018) provide evidence for this asymmetry: incumbents tend to reinforce existing specializations, while new establishments contribute disproportionately to unrelated diversification and structural change (Neffke et al. 2018). In this sense, diversification feasibility is not only a property of regional relatedness patterns; it depends on who attempts diversification (incumbents vs entrants, local vs nonlocal actors) as well as on whether they can mobilize the resources needed to scale and on whether local environments can support the survival

and scaling of novel activities (Neffke et al. 2018).

At this point, the institutional dimension becomes central in a way that remains compatible with the network framing. Institutions matter because they shape how agents access and mobilize capabilities. Neffke et al. (2018) make the mechanism explicit: access depends on embeddedness in “economic, social, and trust relations with governments, institutions, and other firms”. This implies two channels: institutions shape access to local resources via embeddedness, and they shape reliance by affecting whether actors can substitute extra-local capabilities (e.g., through subsidiaries or external networks(Neffke et al. 2018)).

This is consistent with long-standing arguments that localized institutions, norms, and relational infrastructures condition learning, coordination, and innovation in places (Storper 1995; Cooke and Morgan 1993; Gertler et al. 1995; Saxenian 2008). It is also consistent with the geography of innovation view that knowledge spillovers, interactive learning, and absorptive capacity are uneven across space (Cohen, Levinthal, et al. 1990; Audretsch and Feldman 1996; Visser and Boschma 2004). The implication here is straightforward: networks describe feasible recombinations, but institutions condition whether those recombinations can be enacted and/or sustained. This is precisely why entrepreneurship and institutions are treated jointly: entrepreneurs “provide the transmission mechanism transferring advances in knowledge into economic growth”, yet that transmission “may be either hampered or facilitated by the institutional environment (Acs et al. 2018). This ecosystem perspective also aligns with our framing because it provides a disciplined way to think about which parts of the institutional environment matter for mobilisation, rather than treating institutions as a single latent factor. The measurement framework explicitly highlights three features: a systemic approach (components interact), a “Penalty for Bottleneck” logic (weakest links constrain performance), and contextualization (entrepreneurship is embedded in institutional frameworks). This is the rationale behind the Global Entrepreneurship Index (GEI) (Acs, Autio, and Szerb 2014): it is built from pillars that combine individual and institutional variables, and it operationalises bottlenecks by penalising underperforming pillars rather than allowing them to be fully offset by strong ones. Importantly, the GEI

(and its different derivatives) is among the few ecosystem measures explicitly designed to integrate both individual and institutional dimensions, precisely because ecosystem performance is not reducible to institutions alone.

This is where the entrepreneurial ecosystem perspective becomes analytically useful for the present study: it provides a structured way to approximate the institutional and systemic conditions that make diversification feasible beyond what dyadic relatedness captures. Innovation-systems research makes the same point in functional terms: “innovation systems are a very important determinant of technological change” (Hekkert et al. 2007). Rather than treating institutions as a static backdrop, the functional approach focuses on what the system does—supporting experimentation, knowledge creation, diffusion through networks, resource mobilization, market formation, and legitimacy building (Hekkert et al. 2007). In this view, entrepreneurs are essential because they turn “the potential of new knowledge, networks and markets into concrete actions” (Hekkert et al. 2007), but their effectiveness depends on the broader system functions that provide resources and reduce resistance to new activities (Hekkert et al. 2007; Sabatier 1998). Seen this way, ecosystem conditions are not just background because they influence whether regions can mobilise finance, legitimacy, and complementary inputs fast enough to exploit shifting markets and technological transitions. This resonates with recent REC-for-policy reviews that emphasise timing and competition as central moderators of capability-based opportunity (Hekkert et al. 2007; C. Pinheiro 2025; Perez and Soete 1988). Entrepreneurial ecosystems translate these functional ideas into a place-based lens by focusing on how finance, human capital, universities, intermediaries, culture, infrastructure, and governance cohere locally to enable productive entrepreneurship (Stam and Welter 2020; Audretsch and Belitski 2017; Brown and Mason 2017). They also emphasize that ecosystems are spatially embedded but shaped by nested regional, national, and global structures, which matches the idea that innovation systems “cut through” geographical levels and interact across these nested hierarchies (Carayannis and Grigoroudis 2016).

Finally, this ecosystem framing directly addresses the concern raised by Bathelt and Storper (2023) that relatedness patterns can reflect outcomes (and size) more than the learning

processes that produced them. Ecosystem approaches push attention toward the inputs and conditions of learning and experimentation—such as absorptive capacity, network connectivity, institutional support, and access to resources—which vary systematically across places and can be targeted by policy in ways that pure relatedness measures cannot (Qian, Acs, and Stough 2013; Stam and Welter 2020; Varga et al. 2020). This is also consistent with the recent methodological move in the relatedness/complexity literature to evaluate measures by their predictive content and to complement proximity maps with richer, context-sensitive approaches, including machine-learning benchmarks and portfolio-based indicators of coherence and capability alignment—an approach we return to in later sections (Tacchella et al. 2023; Albora et al. 2023; Straccamore, Bruno, and Tacchella 2025).

Taken together, these arguments motivate the core move in this dissertation: a more contextualised assessment of diversification feasibility that keeps the strengths of the relatedness/network tradition (path dependence, structured opportunity sets) while explicitly incorporating (i) agency (firms and entrepreneurs as carriers of capabilities and novelty) (Neffke et al. 2018), and (ii) institutional–system conditions approximated through entrepreneurial ecosystems and higher-level contexts that shape whether potential translates into sustained diversification (Hekkert et al. 2007; Stam and Welter 2020; Audretsch and Belitski 2017). This provides a coherent bridge from an evolutionary resilience-adjacent framing to an empirical strategy that decomposes diversification into endogenous opportunity structure and exogenous enabling conditions—while remaining faithful to the EEG insight that regional change is path-dependent, multi-actor, and deeply place-contingent (P.-A. Balland, Rigby, and Boschma 2015; Boschma 2017; Neffke et al. 2018).

2.3 Relatedness as the baseline, context as the extension

The critiques raised by Bathelt and Storper (2023) and Whittle and Kogler (2020) find a rigorous empirical counterpart in recent methodological work. Tacchella et al. (2023) demonstrate systematically—across trade, patent, and scientific production data—that

the co-occurrence-based relatedness measures that underpin the product space and related approaches perform worse in predicting new specialisations than trivial autocorrelation strategies: simply predicting that a region’s current basket will remain unchanged outperforms the Product Space in the COMTRADE dataset in multiple classification metrics. The reason is structural rather than incidental. The number of activities in any such dataset vastly exceeds the number of locations from which co-occurrences are counted, meaning that “the correlation structure that emerges is mostly random” (Tacchella et al. 2023). Additionally, bipartite location-activity networks exhibit a strong nested structure, in which the relatedness signal is of second order relative to the drive toward diversification—causing more diversified regions to score systematically higher on co-occurrence measures, independently of any genuine capability similarity (Tacchella et al. 2023). This is precisely the “scale artifact” Bathelt and Storper (2023) identify: relatedness density conflates the size and diversity of a regional economy with the learning processes that generated it. The implication is that a purely dyadic proximity-to-entry account of diversification does not only miss mechanisms—it likely measures the wrong thing.

The methodological response to these limitations motivates a shift from two-body (pairwise) to many-body (portfolio-level) measures of feasibility. Tacchella et al. (2023) show that moving from co-occurrence networks to decision-tree-based machine learning models—which condition predictions on the full activity basket rather than summing independent pairwise relatedness scores—improves forecasting precision by up to 384% over the Product Space in out-of-sample link prediction tasks. The core insight is that “describing countries’ development paths as the sum of binary Relatedness relationships is an oversimplification” (Tacchella et al. 2023): what matters is not how close any single candidate technology is to any single existing one, but whether the portfolio as a whole provides the combinatorial conditions for a new specialisation to emerge. Alhora et al. (2023) operationalise a related approach via the Product Progression Probability (PPP), a random-forest-based method that uses the current basket as input to predict industrial upgrading, and show it substantially outperforms pairwise co-occurrence methods—while acknowledging that false positives remain common and statistical regularities in diversification are “simply too weak” for confident individual-level prediction (C. Pinheiro 2025). Taken together, these

methodological advances provide the empirical foundation for the present study’s use of a machine-learning-based potential measure trained on portfolio configurations rather than pairwise distances—treating feasibility as a property of the whole capability system, not of isolated technology-region pairs.

From an industrial policy perspective, the limitations of co-occurrence relatedness extend beyond measurement to theory. C. Pinheiro (2025) argues that recommending related diversification as a general strategy rests on three flawed premises: that related diversification is more common than unrelated diversification (empirically contested); that higher frequency implies higher success rates (not demonstrated); and that the observed pattern reflects a “natural law” of diversification rather than the accumulated effect of past policies that primarily incentivised related activities (Andreoni and Chang 2019). Most consequentially for the present study, C. Pinheiro (2025) draws on developmental state and innovation scholarship (Amsden 1989; K. Lee and Malerba 2017; Wade 2018) to show that the most successful catching-up trajectories in history—South Korea, Ireland, Taiwan—involved precisely the “long jumps” into unrelated productions that REC-based frameworks classify as near-impossible. As Hidalgo et al. (2007) themselves acknowledged, “it is precisely these long jumps that generate subsequent structural transformation, convergence and growth.” This creates a direct tension for policy: relatedness-based recommendations may be appropriate for already-advanced regions with sufficient related capabilities to draw on, but they risk locking less developed regions into low-complexity trajectories. Flavio L. Pinheiro et al. (2025) confirm this empirically for European regions, showing that related diversification has disproportionately benefited regions that were already advanced, generating a spatial inequality feedback loop. The implication is not that relatedness is irrelevant, but that its role as a constraint on feasibility is strongly conditioned by a region’s position in the capability distribution—and that measuring feasibility without accounting for this conditioning risks systematically underestimating the diversification potential of lagging regions while reinforcing the advantages of leading ones.

If relatedness captures the proximity dimension of the capability constraint, network structure captures the architectural dimension—specifically, the role that different positions

in the technology space play in shaping which regions can access which diversification pathways. Tóth et al. (2022) operationalise this intuition using the technology space as a capability network, and Napolitano et al. (2018) show that constrained search processes through these networks generate systematic differences in regional development outcomes. The relevant insight here is that not all technologies are equivalent as diversification targets, even when controlling for relatedness: technologies that serve as bridges between otherwise disconnected knowledge domains—what the innovation literature terms General Purpose Technologies (Bresnahan 2003; Helpman 1998)—may generate qualitatively different diversification opportunities than technologies that are highly embedded in specific knowledge clusters. The former enable recombination across domains; the latter deepen existing specialisations. This distinction matters because, as C. Pinheiro (2025) argues, supply-side capability assessments alone cannot identify which diversification options are strategically valuable—and strategic value depends partly on where in the network architecture a technology sits. Mapping network positions (centrality, betweenness, eigenvector) onto feasibility outcomes therefore provides a mechanism layer that relatedness density cannot supply: it specifies not only whether a technology is reachable from a region’s current portfolio, but what kind of recombination potential it carries and for which types of regions it is most accessible.

A further dimension absent from standard relatedness accounts is the coherence of the regional capability portfolio itself—the degree to which a region’s existing specialisations form a structurally consistent bundle rather than an eclectic collection. Teece et al. (1994) introduced coherence in the context of firm diversification, defining it as the degree to which a firm’s activities draw on a common underlying set of capabilities. The concept translates directly to regions: a region whose technological portfolio is coherent—in the sense that its existing specialisations are embedded in the technology network in similar ways—has a capability base that is internally consistent and potentially more effective at supporting related recombination than a region whose portfolio spans disparate network positions, even if both have similar levels of relatedness density. This distinction matters because coherence captures a property of the whole portfolio that pairwise relatedness cannot: two regions can have the same average density vis-à-vis a candidate technology while differing substantially

in whether their existing specialisations share the combinatorial infrastructure needed to support entry. This is not merely a refinement of the relatedness concept but a qualitatively different moderating condition—one that conditions how network positions translate into feasibility outcomes. As Tacchella et al. (2023) argue, “describing development paths as the sum of binary Relatedness relationships is an oversimplification”—coherence is precisely the kind of higher-order portfolio property that many-body approaches can detect and that pairwise approaches necessarily miss.

Building on the ecosystem/innovation-systems framing introduced above, the institutional dimension can further be justified as an enactment mechanism. While networks describe feasible recombinations; institutions condition whether those recombinations can be enacted and sustained. Neffke et al. (2018) make this mechanism explicit since access to these capabilities depends on embeddedness in “economic, social, and trust relations with governments, institutions, and other firms”. This implies two distinct channels where first, institutions shape local resource access through different types of embeddedness, and second, they also shape reliance on extra-local capabilities through connectivity to external networks. Consequently, this aligns with many studies that argue that localised institutions, norms, and relational infrastructures, condition learning, coordination, and innovation in places (Storper 1995; Cooke and Morgan 1993; Gertler et al. 1995; Saxenian 2008), and with the geography of innovation view that knowledge spillovers, interactive learning, and absorptive capacity are uneven across space (Cohen, Levinthal, et al. 1990; Audretsch and Feldman 1996; Visser and Boschma 2004). This is one of the core arguments also in C. Pinheiro (2025): REC-based frameworks are biased toward domestic supply, ignoring the demand-side and competitive dynamics that condition capability accumulation.

This motivates treating entrepreneurship and institutions as a coupled system rather than treating institutions as a single latent factor that mirrors quality of governance. The Global Entrepreneurship Index (GEI) operationalises this coupling through a systemic approach that simultaneously combines individual and institutional aspects of a national ecosystem (Acs, Autio, and Szerb 2014; Varga et al. 2020). This is among the few ecosystem measures explicitly designed to capture both the individual and the systemic dimensions of

the institutional enabling environment. This justifies its empirical merit as an appropriate operationalisation of the higher-level context that the present study requires.

One dimension often absent from existing accounts of relatedness-constrained diversification is spatial interdependence. The standard capability network framing treats each region’s feasibility as a function of its own portfolio and the global technology network—implicitly assuming independence across regions. Yet evolutionary economic geography provides strong reasons to expect spatial clustering of diversification outcomes: knowledge spillovers are geographically bounded (Audretsch and Feldman 1996), learning processes are proximity-dependent (Boschma 2005), and the institutional and relational infrastructures that enable capability mobilisation are spatially embedded (Brown and Mason 2017; Audretsch and Belitski 2017). If neighbouring regions share similar capability bases, face similar institutional environments, or generate knowledge externalities that spill across borders, then their diversification trajectories will be correlated in ways that a purely dyadic, aspatial framework cannot capture. P.-A. Balland, Rigby, and Boschma (2015) identify this as one of the underexplored dimensions of evolutionary resilience research: spatial structure shapes not only what regions can do but how their options co-evolve with those of neighbouring economies. This has a direct implication for empirical strategy: estimating diversification feasibility without accounting for spatial interdependence risks attributing to internal capability configurations what is partly the product of regional clustering and spatial spillovers. Incorporating spatial models—specifically spatial autoregressive and spatial error specifications—therefore constitutes not merely a robustness check but a substantive test of whether the feasibility landscape is spatially structured and, if so, of how much of regional diversification potential is shaped by neighbourhood effects versus internal portfolio characteristics.

Taken together, these arguments motivate the core empirical move of this dissertation: a more contextualised assessment of diversification feasibility that keeps the strengths of the relatedness and network tradition—path dependence, structured opportunity sets, capability constraints—while explicitly incorporating the dimensions that this tradition has neglected. The first of these is the measurement dimension: moving from pairwise co-

occurrence proximity to portfolio-level, machine-learning-based feasibility assessment that captures higher-order capability configurations (Tacchella et al. 2023; Albora et al. 2023). The second is the network architecture dimension: decomposing technology positions into bridging (betweenness) and enabling (eigenvector) roles that carry qualitatively different implications for which regions can access which diversification pathways. The third is the coherence dimension: treating portfolio-level structural consistency as a moderating condition that shapes how network positions translate into feasibility outcomes (Teece et al. 1994). The fourth is the institutional-ecosystem dimension: approximating the enabling conditions for capability mobilisation through a systemic, bottleneck-sensitive measure of national entrepreneurial ecosystems (Acs, Autio, and Szerb 2014; Varga et al. 2020) which addresses C. Pinheiro (2025)’s concern that supply-side capability maps miss the demand, competition, and institutional conditions that determine whether feasibility becomes realised diversification. The fifth is the spatial dimension: testing whether the feasibility landscape is spatially structured, which P.-A. Balland, Rigby, and Boschma (2015) identify as a neglected but theoretically motivated dimension of evolutionary regional change. This context decomposition is not an arbitrary accumulation of controls, rather it is a theoretically grounded architecture that maps directly onto the gaps identified across the evolutionary, complexity, and industrial policy literatures, and that together allows a more complete account of what shapes regional technological diversification than what single-dimension approaches can provide.

3 The interconnectedness between regions and technologies

3.1 Methodological motivation

Traditional relatedness frameworks model diversification through symmetric co-occurrence matrices and linear aggregation of relatedness density (Hausmann and Hidalgo 2011). However, as established in our problem statement, these constructs suffer from three structural limitations: (1) symmetry assumptions that obscure directional dependencies between technologies, (2) linear aggregation that loses granular information about technology-specific requirements, and (3) noise when technologies outnumber regions.

We address these limitations by replacing traditional measures with a machine learning training strategy based on the Random Forest algorithm (RF). Specifically, we use RF to generate: (1) **Feature Importance Technology Space (FITS)** following Fessina et al. (2024), analogous to the mainstream Technology Space. FITS captures directional, hierarchical technology relationships, replacing symmetric relatedness measures; and (2) **predicted probabilities (technological potential)** following Albora et al. (2023), analogous to relatedness density. The technological potential estimates region-specific feasibility of technology adoption given the region’s own technological portfolio, replacing linear relatedness density. This approach enables the contextualization of diversification strategies by accounting simultaneously the heterogeneity in technology-specific characteristics, regional knowledge infrastructure, and the broader non-linear interconnectedness between regions and technologies that traditional measures cannot adequately capture.

3 The interconnectedness between regions and technologies

Before detailing our methodology and how we leverage the contributions in Fessina et al. (2024) and Albora et al. (2023), we will briefly situate this approach within the REC literature. Essentially, REC formalizes the empirical observation that shared input requirements (knowledge, resources, capabilities) determine diversification feasibility (Hidalgo et al. 2018; Hidalgo 2021). While this framework have proven valuable for policy (Zaccaria et al. 2018; E and A 2021), deriving granular, context-specific implications remains challenging (Hidalgo 2023; Li and Neffke 2024). Recent work on unrelated diversification (Flávio L. Pinheiro et al. 2022; Boschma et al. 2023), geographic inequalities (Flavio L. Pinheiro et al. 2025; Hartmann et al. 2017), and emerging technologies (C. Lee et al. 2018; Fessina et al. 2024) highlights the need for methodologies that capture contextual nuances beyond path dependency identification which we will elaborate on in subsequent sections.

Our approach is a response to the nuances expressed in the literature. But in order to properly situate our contribution we must first elaborate on the mainstream approach conceptually, methodologically and its empirical implications which we briefly outlined in the introduction. From there we expand on the mainstream approach and present our alternative constructs conceptually and mathematically, and showcase our training strategy. To this end, we also provide a brief example on how the mainstream approach and the machine learning approach to modeling the technology space compare. For the sake of clarity and consistency we will establish our notation system and the main definition in this paragraph and use them in the subsequent sections and chapter in this dissertation unless we specify otherwise.

We consider the following sets:

- $\mathcal{T} = \{t : 1 \leq t \leq N_T\}$: set of technologies
- $\mathcal{R} = \{r : 1 \leq r \leq N_R\}$: set of regions
- $\mathcal{Y} = \{y : 1 \leq y \leq N_Y\}$: set of years
- $\mathcal{C} = \{c : 1 \leq c \leq N_C\}$: set of technology categories¹
- $\mathcal{K} = \{k : 1 \leq k \leq N_K\}$: set of countries

¹we refer to categories, domains, industries interchangeably in this work

3 The interconnectedness between regions and technologies

where N_T , N_R , N_Y , N_C , and N_K denote the total number of technologies, regions, years, technology categories, and countries respectively.

These sets are mirrored by the European Patent Office data (1978-2021) classified at the 4-digit IPC level, yielding 641 distinct technologies across 345 NUTS2 regions in 34 European countries. IPC classifications provide hierarchical structure: section (letter, we also refer to this as categories), class (two digits), subclass (letter). For example, F16H encompasses Section F (Mechanical engineering), Class 16 (engineering elements for mechanical power transmission), and Subclass H (gearing systems).

We quantify regional specialization using the Revealed Comparative Advantage (Balassa 1965). Although originally designed for trade data, this metric has been widely adopted in innovation geography literature. Following the nomenclature in P.-A. Balland and Rigby (2017), we refer to it as Revealed Technological Advantage (RTA) in our patent data context. The RTA measures relative specialization, enabling simultaneous capture of expertise depth and portfolio diversity. Despite critiques regarding patent-based applications (P. Balland and Boschma 2019; Diodato et al. 2023), we retain it as the input measure for our machine learning models. The goal here is to learn meaningful technology relationships from portfolio patterns rather than from raw co-occurrence frequencies.

The RTA for region r in technology t during year y is defined as:

$$\text{RTA}_{r,t,y} = \frac{\frac{X_{r,t,y}}{\sum_{t'} X_{r,t',y}}}{\frac{\sum_{r'} X_{r',t,y}}{\sum_{r',t'} X_{r',t',y}}} = \frac{X_{r,t,y} \sum_{r',t'} X_{r',t',y}}{(\sum_{t'} X_{r,t',y}) (\sum_{r'} X_{r',t,y})}$$

Where, $X_{r,t,y}$: patent count for region r in technology t during year y

For each year y , we construct the RTA matrix $\mathbf{R}^{(y)}$ with entries $\text{RTA}_{r,t,y}$:

$$\mathbf{R}^{(y)} = [\text{RTA}_{r,t,y}]_{r=1,\dots,N_R}^{t=1,\dots,N_T}$$

3 The interconnectedness between regions and technologies

These yearly matrices form the foundation for all subsequent modeling. We then binarize specialization for classification tasks:

$$z_{r,t,y} = \begin{cases} 1 & \text{if } \text{RTA}_{r,t,y} \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

where $z_{r,t,y} = 1$ indicates region r has comparative advantage (specialization) in technology t at year y .

To further elaborate, let's consider $c_{t,t'}$, the count of patents containing both technologies t and t' ($t \neq t'$) at a given time period, such that $t, t' \in \mathcal{T}$, and c_t , $c_{t'}$ denote the total patent counts for each class individually.

In the following we will begin by outlining how the REC literature approach relatedness, how it's calculated and discuss in detail different technical contributions and their limitations. We will then present our approach and detail our methodological and conceptual contribution in response to these limitations.

3.2 The mainstream approach

3.2.1 Relatedness through co-occurrence

The REC literature operationalises relatedness through the observed frequency of co-occurrences. The intuition is appealingly simple: technologies that frequently appear together—whether in the same patent, the same firm, or the same region—likely share underlying knowledge requirements. If mastering one technology facilitates developing another, we should observe them co-located more often than chance would predict. This reasoning has proven influential, providing the empirical foundation for technology space visualisations and diversification recommendations across academic and policy settings.

Yet the inference from co-occurrence to capability-based relatedness rests on assumptions that warrant scrutiny. Co-occurrence is a frequency measure; it tells us how often two technologies appear together, not why they do so or whether that co-presence reflects shared inputs, complementary applications, or statistical artefact. The distinction matters: a measure intended to guide capability-building strategies must capture genuine knowledge relationships rather than patterns that emerge from data structure alone.

Let $c_{t,t'}$ denote the count of patents containing both technologies t and t' (where $t \neq t'$) within a given time period, and let c_t and $c_{t'}$ denote the total patent counts for each technology individually. The simplest relatedness measure would be $c_{t,t'}$ itself, but raw co-occurrence counts are problematic: frequently patented technologies will naturally co-occur more often simply due to their prevalence. Various normalisation schemes address this issue.

Association Strength compares observed co-occurrence to expected co-occurrence under independence:

$$\phi_{t,t'}^{\text{assoc}} = \frac{c_{t,t'} \cdot T}{c_t \cdot c_{t'}}$$

where $T = \sum_{t \in \mathcal{T}} c_t$ is the total patent count. Values exceeding one indicate that the technologies appear together more frequently than would be expected if their occurrences were independent—they are positively associated. This measure properly corrects for size effects: a rare technology and a common technology can still exhibit strong association if they co-occur more than their marginal frequencies would predict.

The Probability Index refines this probabilistic approach by accounting for the fact that a patent cannot co-occur with itself:

$$\phi_{t,t'}^{\text{prob}} = \frac{c_{t,t'}}{\frac{1}{2} \left[\frac{c_t}{T} \cdot \frac{c_{t'}}{T-c_t} + \frac{c_{t'}}{T} \cdot \frac{c_t}{T-c_{t'}} \right] \cdot T}$$

3 The interconnectedness between regions and technologies

This correction, while subtle, provides more accurate estimates for patent co-occurrence analysis and is implemented in standard software packages used in the field.

Cosine Similarity offers a set-theoretic alternative:

$$\phi_{t,t'}^{\text{cosine}} = \frac{c_{t,t'}}{\sqrt{c_t \cdot c_{t'}}$$

This measures relative overlap between technology profiles. While intuitive, it systematically favours frequent technologies over rare ones and does not properly correct for size effects, making comparisons across technologies with different prevalences problematic.

Minimum Conditional Probability takes a different approach by examining regional co-location rather than patent-level co-occurrence:

$$\phi_{t,t'}^{\text{min}} = \min \{P(\text{RTA}_{r,t}|\text{RTA}_{r,t'}), P(\text{RTA}_{r,t'}|\text{RTA}_{r,t})\}$$

This computes the minimum of the conditional probabilities: given that a region specialises in one technology, how likely is it to specialise in the other? The minimum operator ensures symmetry while avoiding inflation from imbalanced conditional relationships.

A clarification is warranted here regarding two distinct but often conflated concepts. Co-occurrence typically refers to technologies appearing together within the same patent or firm, while co-location refers to technologies appearing together within the same region. Both capture related but different phenomena: co-occurrence within patents reveals direct technical complementarities, while co-location within regions reveals shared capability requirements that may operate at greater conceptual distance. The literature sometimes uses these terms interchangeably, but the distinction matters for interpreting what relatedness measures actually capture.

3.2.2 Structural limitations of co-occurrence-based relatedness

Despite the variety of normalisation approaches, co-occurrence-based relatedness measures share structural limitations that affect their reliability and interpretation.

Dimensional mismatch produces noisy estimates. The reliability of frequency-based measures depends on having sufficient observations to estimate co-occurrence patterns accurately. In typical applications, the number of technologies substantially exceeds the number of locations—estimating a matrix of several thousand technology pairs from fewer than two hundred country observations, for instance, produces correlation structures that are largely random (Tacchella et al. 2023). The number of observations required for reliable estimation of such matrices easily reaches tens of thousands (Bun, Bouchaud, and Potters 2017), far exceeding available data. This problem is not resolved by moving to subnational levels, where harmonised data is often unavailable and the fundamental dimensional mismatch persists.

Nested structure dominates the relatedness signal. Location-activity networks typically exhibit strong nested structure: diversified regions produce almost everything while less diversified regions produce only ubiquitous activities (Mariani et al. 2019). This triangular pattern means that when counting co-occurrences, one primarily detects the diversification hierarchy rather than genuine capability-based relationships. The relatedness signal becomes second-order relative to the structural pattern that generates nestedness (Tacchella et al. 2023). This explains why even randomised relatedness matrices produce predictions comparable to co-occurrence-based measures—they still capture the fact that diversified regions are likely to add activities regardless of which specific activities are involved (Bustos et al. 2012).

Spurious associations are rarely filtered. Studies employing co-occurrence measures often fail to test the statistical significance of the relationships they identify (Saracco et al. 2015, 2017). Without appropriate null models, spurious co-occurrences due to sampling variability are mistaken for genuine relatedness. When such tests are applied to trade data, the results are sobering: the null hypothesis of random diversification—no path

dependence—is rejected for only approximately half of countries examined, with most high-income and large economies exhibiting diversification patterns that defy path dependence entirely (Coniglio et al. 2021). Whether these findings extend to patent-based technology data remains an open question, but they raise concerns about the generalisability of path-dependent diversification as a universal empirical regularity.

Symmetry obscures directional dependencies. All normalisation schemes above produce symmetric measures: $\phi_{t,t'} = \phi_{t',t}$. This symmetry assumes that the relationship between any two technologies operates bidirectionally with equal strength. Yet technological development often exhibits clear hierarchies (Boschma 2017). Expertise in semiconductor fabrication may strongly predict future development of integrated circuit design, but the reverse need not hold—circuit designers may not subsequently develop fabrication capabilities. The capabilities required to transition from technology i to technology j may differ vastly from those required for the reverse transition. Symmetric measures cannot distinguish prerequisite relationships from derived relationships, enabling technologies from their applications, or identify the stepping-stone pathways that matter for strategic planning.

3.2.3 Relatedness density through linear aggregation

Relatedness density (Hidalgo et al. 2007) bridges the gap between pairwise technology relationships and region-specific diversification potential. Where co-occurrence measures describe technology-to-technology proximity, relatedness density aggregates these proximities relative to a region’s existing portfolio, producing a single value meant to capture how feasible a new technology is for a given region. The measure has become central to diversification analysis and policy guidance, serving as the primary predictor in entry models and the basis for strategic recommendations.

The aggregation, however, imposes strong assumptions about how capabilities combine. By summing pairwise relationships, relatedness density treats the regional portfolio as a linear accumulation of independent components. Whether this adequately represents

3 The interconnectedness between regions and technologies

how knowledge and capabilities actually interact in enabling new activities is an empirical question that the measure’s structure forecloses.

For region r and technology t , relatedness density aggregates the relatedness between technology t and all technologies in which the region currently specialises:

$$\text{relatedness density : } \Phi_{r,t} = \frac{\sum_{t' \in \mathcal{T}_r, t' \neq t} \phi_{t,t'}}{\sum_{t' \neq t} \phi_{t,t'}} \times 100$$

where $\mathcal{T}_r$ denotes the set of technologies in which region r specialises (those with $\text{RTA}_{r,t} \geq 1$).

The interpretation is intuitive: the numerator sums relatedness values to technologies the region already has, while the denominator normalises by total relatedness to all technologies. High relatedness density indicates that technology t is “close” to the region’s existing capabilities—many of its related technologies are already present in the regional portfolio. Low relatedness density suggests that t requires capabilities the region currently lacks.

This measure has proven remarkably predictive: across numerous studies using different data, time periods, and geographic contexts, higher relatedness density consistently associates with higher probability of future entry. Regions are more likely to develop specialisations in technologies that relate to what they already do well.

3.2.4 Structural limitations of linear aggregation

The predictive success of relatedness density is not in question. The limitations concern what it cannot reveal and how its structure constrains interpretation.

Linear aggregation obscures combinatorial information. Relatedness density sums pairwise relationships without regard to their interactions. This implicitly assumes that having two related technologies is simply twice as beneficial as having one—capabilities combine additively. Yet capabilities may interact in ways that linear summation cannot

capture. Certain combinations may be more than the sum of their parts, creating complementarities where the joint presence of technologies A and B enables technology C in ways that neither alone provides. Other combinations may be redundant, contributing less than their individual relatedness values would suggest. By reducing many-body relationships to independent two-body terms, linear aggregation discards precisely the combinatorial information that may distinguish feasible diversification paths from infeasible ones (Tacchella et al. 2023). Empirically, prediction models that capture many-product interactions substantially outperform density-based predictions, suggesting that linear aggregation loses significant information (Albora et al. 2023).

Noise compounds through aggregation. The dimensional mismatch and spurious associations affecting pairwise co-occurrence estimates do not disappear when aggregated into relatedness density—they compound. Summing noisy pairwise values produces a noisy aggregate. The normalisation by total relatedness provides some correction but cannot recover signal that was lost at the pairwise level. For regions with portfolios concentrated in rare technologies, where pairwise estimates are least reliable, relatedness density inherits and potentially amplifies these estimation problems.

Frequency of entry conflates observation with feasibility. The empirical regularity that higher relatedness density associates with higher entry probability is often interpreted as evidence that related diversification is more feasible and therefore should be prioritised. But this inference conflates observed frequency with success rate as emphasised in C. Pinheiro (2025). We observe which entries occurred, not which were attempted; the denominator—attempts—remains unobserved. Related diversification may appear more common because it is attempted more frequently, perhaps precisely because existing frameworks recommend it, rather than because it succeeds more often when attempted. Coniglio et al. (2021) addresses precisely this issue and shows that related diversification attempts succeeded 80% in Germany in contrast to 40% in the United States and France. Moreover, the observed product and technology spaces may themselves reflect past policies promoting related diversification (Andreoni and Chang 2019), creating circularity when used to justify future policy.

3.2.5 Implications for contextualisation

The limitations identified above are not independent—they compound. Noisy co-occurrence estimates feed into symmetric matrices that are then linearly aggregated, producing context-free measures whose predictive success may reflect structural properties of the data (nestedness, diversification hierarchies) as much as genuine capability-based relationships. The resulting framework provides a supply-side instrument: it identifies what regions might produce given their existing portfolios, while remaining silent on demand dynamics, competitive pressures, and the institutional conditions that shape whether capability-based proximity translates into actual entry (C. Pinheiro 2025; Andreoni and Chang 2019).

This supply-side orientation has consequences. When applied to policy, relatedness-based recommendations may systematically favour already-diversified regions whose dense portfolios generate high relatedness density across many targets, while offering limited guidance to less-diversified regions whose sparse portfolios produce low density values regardless of strategic potential. The concern that such recommendations reinforce rather than ameliorate path dependency—locking peripheral regions into constrained trajectories while core regions accumulate further advantages—has been explicitly raised in recent literature examining regional inequality and the distributional consequences of relatedness-based strategies (Flavio L. Pinheiro et al. 2025; Mealy and Coyle 2022; Hidalgo 2023).

The issue is not that relatedness is empirically wrong—the predictive regularity is robust—but that the measures as currently constructed cannot inform us about the conditions under which relatedness matters more or less, about which technologies serve as stepping stones toward others, or about how regional knowledge infrastructure, national ecosystems, and spatial factors moderate diversification possibilities. Addressing these limitations requires reconstructing the core methodological constructs in ways that relax the problematic assumptions while preserving the fundamental insight that prior capabilities shape future possibilities. This reconstruction motivates the machine learning approach developed in the following sections.

3.3 An alternative approach: Machine learning for contextualised relatedness

The limitations outlined above—dimensional mismatch producing noisy correlations, symmetric measures obscuring directional dependencies, and linear aggregation discarding interaction effects—are not merely technical inconveniences. They represent fundamental obstacles to using relatedness for prediction and policy. If co-occurrence methods perform no better than trivial baselines, as demonstrated in Tacchella et al. (2023), then their assessment of diversification feasibility offers insufficient guidance for strategic decisions. This motivates a shift from descriptive correlation to predictive modelling.

3.3.1 Machine learning as a prediction framework

The core reframing here is methodological: rather than inferring relatedness from co-occurrence patterns and hoping this correlates with future outcomes, we directly model diversification as a prediction problem. This follows recent work positioning relatedness estimation within supervised learning frameworks (Tacchella et al. 2023; Albora et al. 2023; Fessina et al. 2024). The shift offers three advantages.

First, it provides a falsifiable evaluation standard. Co-occurrence measures proliferate without systematic comparison—proximity, association strength, minimum conditional probability, and numerous variants coexist with no principled basis for selection. Out-of-sample prediction performance offers an objective criterion: methods that better forecast actual diversification outcomes are, by definition, better capturing whatever underlying relatedness structure matters for real transitions (Albora et al. 2023). This addresses the circularity where relatedness is inferred from outcomes and then used to explain those same outcomes.

Second, prediction-based approaches can capture the many-body correlations that density-based measures discard. As Tacchella et al. (2023) emphasise, describing diversification paths as sums of binary relatedness relationships is an oversimplification. A region's

probability of developing technology i depends not just on pairwise similarities but on the full configuration of its existing portfolio—which combinations are present, which are absent, and how these interact. Tree-based algorithms learn precisely these complex conditional patterns.

Third, the framework naturally accommodates the asymmetric, directed relationships that symmetric co-occurrence measures cannot represent. When we train separate models for each target technology, we obtain directional importance scores: technology t may strongly predict entry into technology i while i weakly predicts entry into t . This asymmetry reveals hierarchical structure—prerequisites, enabling technologies, and developmental sequences—that bidirectional similarity metrics obscure.

We apply this framework through two complementary constructs:

Technological Potential estimates $p_{r,t,y}$, the probability that region r develops specialisation in technology t by year y , given its existing portfolio. This replaces relatedness density. Rather than linearly summing symmetric co-occurrence frequencies, we predict region-specific feasibility using the full structure of past portfolios. The measure captures non-linear interactions and regional heterogeneity that additive approaches miss.

Feature Importance Technology Space (FITS) extracts directional dependencies $I_{t \rightarrow t'}$ between technologies from the trained models. This replaces symmetric technology networks. Instead of co-occurrence frequencies, we identify which technologies predict others' future adoption—and crucially, which do not predict in the reverse direction. The resulting asymmetric network reveals hierarchical technological structure invisible to symmetric measures.

3.3.2 Why Random Forest

Among supervised learning approaches, tree-based ensemble methods—particularly Random Forest—offer properties well-suited to relatedness estimation. We briefly outline these before presenting the technical details.

High-dimensional sparse data. Regional technology portfolios are high-dimensional (hundreds of technology classes) and sparse (most regions specialise in few technologies). Tree-based methods handle this naturally through recursive partitioning: each split considers only a subset of features, and the hierarchical structure means irrelevant technologies simply do not appear in decision paths. Linear methods struggle with the collinearity and sparsity typical of specialisation data; neural networks require larger samples than regional patent data typically provide.

Non-linear interactions. The recursive splitting structure captures interactions without explicit specification. If technology t_1 matters for predicting t_3 only when t_2 is also present, this emerges naturally as a conditional split pattern. Density-based measures, by construction, cannot represent such conditionality—they assume each related technology contributes independently.

Interpretable feature importance. Unlike black-box models, Random Forests produce feature importance scores that aggregate each predictor’s contribution across all trees. These scores become the basis for FITS construction, translating prediction performance into an interpretable technology network. This maintains the explanatory value of relatedness measures while grounding them in predictive validity.

Ensemble variance reduction. Bootstrap aggregation across many trees reduces the variance that would plague individual decision trees, particularly given the moderate sample sizes (regions $\times$ years) available in patent data. The ensemble structure provides stability without sacrificing the flexibility needed to capture complex capability interactions.

The choice of Random Forest over alternatives thus reflects both practical constraints (sample size, interpretability requirements) and the specific structure of relatedness estimation (sparse, high-dimensional, interaction-rich). We now present the algorithm formally before describing our application.

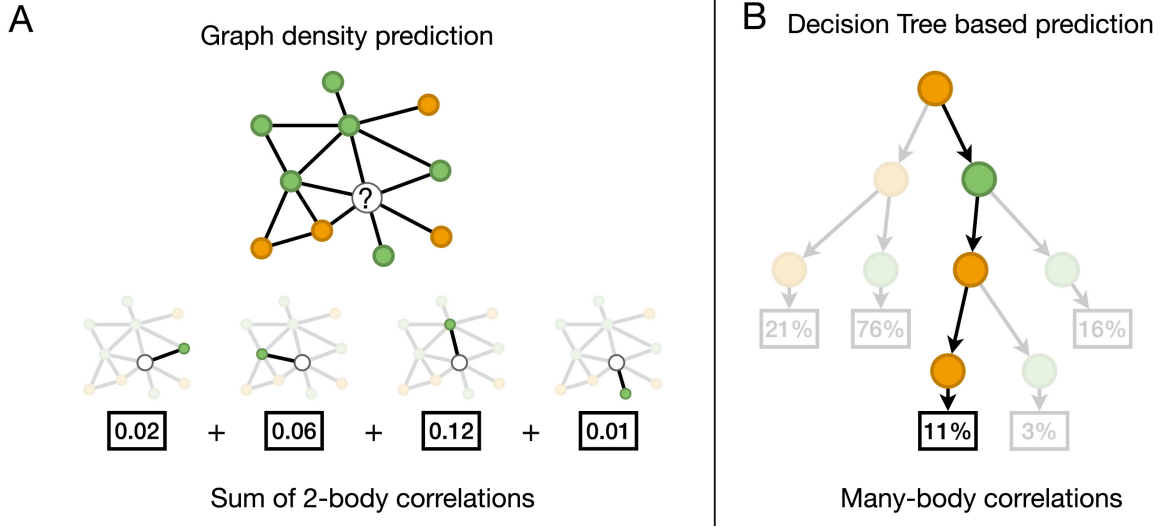


Figure 3.1: Node correlation - Density vs ML approach Tacchella et al. 2023

3.3.3 The Random Forest algorithm

Figure 3.1 from Tacchella et al. (2023) contrasts the fundamental difference between traditional relatedness density approaches (Panel A) and machine learning-based prediction methods (Panel B) in modeling diversification feasibility. Panel A illustrates how density-based predictions aggregate pairwise co-occurrence relationships linearly: the feasibility of entering the focal technology (white node with ‘?’) is computed as the sum of independent two-body correlations with each technology in the regional portfolio (green and orange nodes), yielding a simple additive score ($0.02 + 0.06 + 0.12 + 0.01 = 0.21$). Panel B demonstrates how decision tree algorithms capture many-body correlations by evaluating the full configuration of the portfolio through recursive partitioning: at each node, the algorithm tests combinations of technologies (e.g., presence of both green and orange technologies) and assigns predictions based on complete paths through the tree (21%, 76%, 11%, etc.), making the effect of any single technology dependent on the entire portfolio context. This structural difference explains why machine learning approaches can recover non-linear capability interactions that linear aggregation of symmetric relatedness measures cannot represent, as the value of possessing one technology changes depending on which other technologies are present, a form of complementarity that density treat as independent contributions. This is leads to explore further how and why random forest—our adopted tree

3 The interconnectedness between regions and technologies

algorithm—works, especially for our case which is the prediction of regional technological specialisation.

Random Forest constructs an ensemble of B decision trees through bootstrap aggregating (bagging) with random feature subsampling. Each tree T_b is built on a bootstrap sample $\mathcal{D}_b^*$ drawn with replacement from the original dataset.

Each tree recursively partitions the feature space through binary splits. At node t , we randomly select m features (typically $m = \sqrt{p}$) and evaluate all possible splits within this subset. For feature j and threshold τ , the split creates two child nodes: $t_L = \{i : x_{ij} \leq \tau\}$ and $t_R = \{i : x_{ij} > \tau\}$.

Split quality is assessed via Gini impurity:

$$G(t) = 1 - \sum_{k=0}^1 p_k^2(t) = 2p_0(t)p_1(t)$$

where $p_k(t) = n_k(t)/n(t)$ represents the proportion of class k observations at node t . Gini impurity quantifies node heterogeneity: $G = 0$ indicates perfect purity (homogeneous class), while $G = 0.5$ indicates maximum impurity (equal class distribution). The optimal split maximises weighted impurity reduction:

$$\Delta G(j, \tau) = G(t) - \left[\frac{n(t_L)}{n(t)} G(t_L) + \frac{n(t_R)}{n(t)} G(t_R) \right]$$

Weighting by relative node size prevents trivial splits that isolate single observations into pure but uninformative leaves.

Recursive splitting continues until predefined stopping criteria: node purity ($G = 0$), minimum node size threshold, or maximum tree depth. Terminal nodes are assigned the majority class of their constituent observations.

For prediction, observation $\mathbf{x}$ traverses all B trees. Final classification aggregates individual tree predictions via majority voting:

$$\hat{y}(\mathbf{x}) = \text{mode}\{\hat{y}_1(\mathbf{x}), \dots, \hat{y}_B(\mathbf{x})\}$$

Class probabilities are estimated as the proportion of trees predicting each class:

$$\hat{P}(y = 1|\mathbf{x}) = B^{-1} \sum_{b=1}^B \mathbb{I}[\hat{y}_b(\mathbf{x}) = 1]$$

This probability represents empirical vote share across trees. Values near 1 indicate strong consensus for class 1 (high confidence), while values near 0.5 reflect uncertainty. Unlike parametric models, these are data-driven vote proportions rather than model-based probability estimates.

The algorithm’s effectiveness stems from variance reduction through decorrelated predictions. Bootstrap sampling and random feature selection reduce inter-tree correlation ρ , yielding ensemble variance:

$$\text{Var}(\bar{y}) = \rho\sigma^2 + \frac{1-\rho}{B}\sigma^2$$

As B increases and ρ decreases, ensemble variance diminishes while maintaining the low bias of flexible tree models.

Feature importance quantifies each predictor’s contribution by aggregating Gini impurity reductions:

$$I(j) = \frac{1}{B} \sum_{b=1}^B \sum_{t \in T_b: v(t)=j} \Delta G(t)$$

where the sum runs over all nodes using feature j for splitting. Higher values indicate features consistently creating purer partitions. Feature importance measures predictive association rather than causal effect, and exhibits bias toward high-cardinality features and correlated predictor sets. It ranks predictive relevance but requires caution in causal interpretation.

3.3.4 Training procedure

We follow the methodology of Fessina et al. (2024), Albora et al. (2023) (and Tacchella et al. (2023)), originally developed for trade data. The key innovation lies in training separate models for each technology rather than constructing a single global model. For every technology $t, t' \in \mathcal{T}$, we train a binary classification model where:

- **Outcome:** $z_{r,t,y}$ indicates whether region r specialises in technology t at year y ($\text{RTA} \geq 1$)
- **Features:** $\{\text{RTA}_{r,t,y-\delta} : t \in \mathcal{T}, t \neq t'\}$ comprises RTA values for all other technologies δ years prior

This technology-specific approach enables each model to learn unique dependency patterns. The lag $\delta = 4$ years balances predictive horizon with data availability, following conventions in the literature on capability accumulation (Andreoni and Chang 2019).

We predict specialisation for target years $y^* \in \{2008, \dots, 2018\}$ using data from 1978 onward. For each target year y_t and technology i :

Training set:

$$X_{\text{train}} = \{\text{RTA}_{r,t,y} \mid y \in [1978, y_t - 2\delta], t \neq t'\}$$

$$Y_{\text{train}} = \{z_{r,t',y} \mid y \in [1978 + \delta, y_t - \delta]\}$$

Test set:

$$X_{\text{test}} = \{\text{RTA}_{r,t',y^*-\delta} \mid t \neq t'\}$$

$$Y_{\text{test}} = \{z_{r,t,y^*}\}$$

This temporal structure ensures strict separation: models predict future specialisation using only information available at the prediction date. The expanding training window incorporates historical diversification patterns while the fixed lag prevents information leakage.

The procedure yields 7,051 models (641 technologies $\times$ 11 years). Given computational constraints, we performed cross-validation on a stratified sample of four technologies (G06G, B67B, D02J, C08J) spanning different IPC sections following Bennour (2026), applying the modal optimal parameters across all models: `mtry = 139`, `trees = 100`, `min_n = 38`. Training was conducted in R using the Ranger package (Wright and Ziegler 2017), orchestrated via the targets pipeline (Landau 2021).

3.4 Regional technological potential

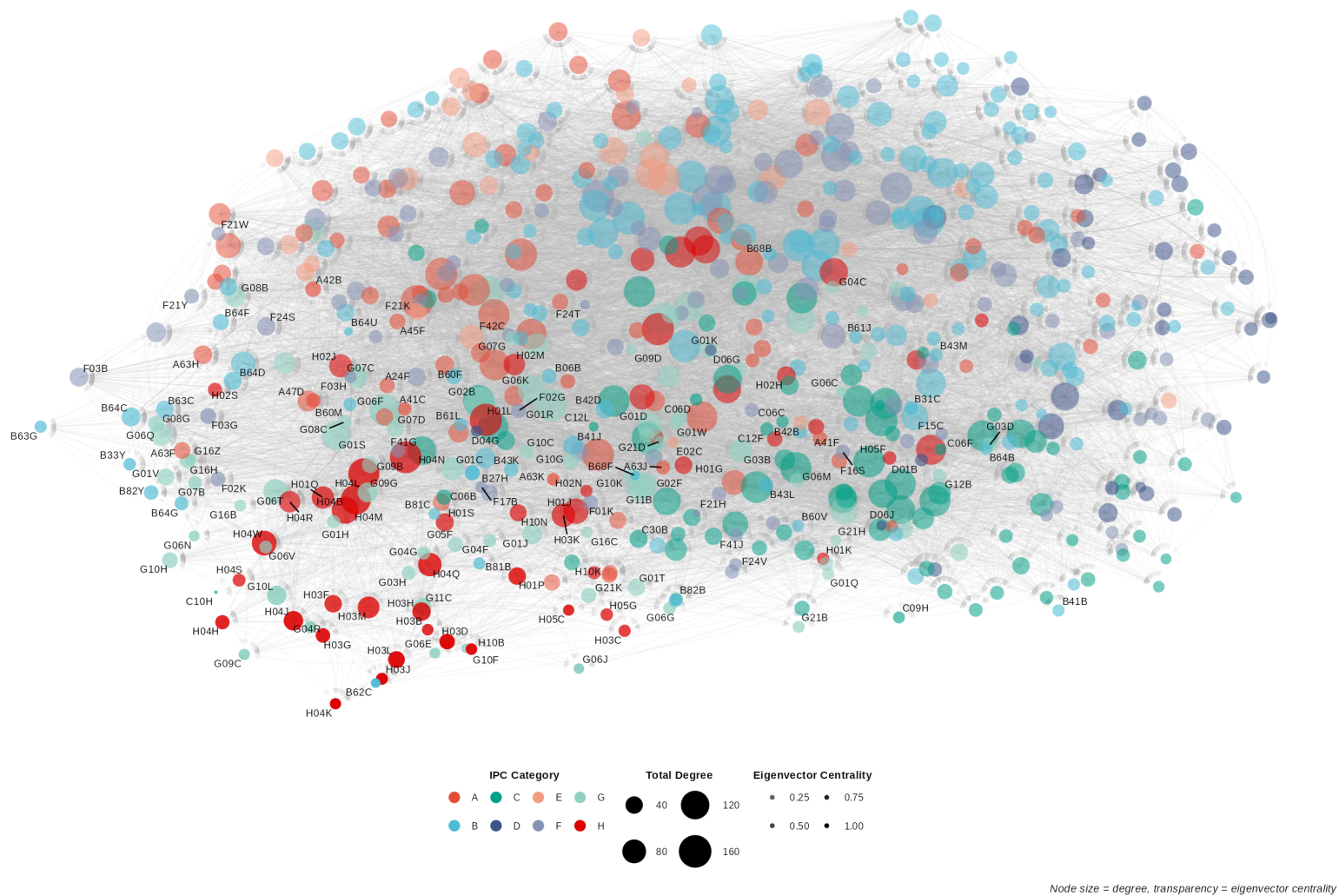
Each model produces probabilities $p_{r,i,y} = P(z_{r,i,y} = 1 \mid \text{RTA}_{r,.,y-\delta})$ representing the likelihood that region r develops specialisation in technology i by year y , conditional on its prior portfolio. These probabilities constitute **Technological Potential**—a forward-looking, region-specific measure of diversification feasibility.

The shift from correlation to prediction reframes what the measure captures. Density asks: “How similar is technology t to what region r already has?” Potential asks: “Given everything we know about how regions diversify, how likely is region r to develop technology t ?” The latter question—the one relevant for policy—requires a predictive approach.

High potential ($p_{r,t,y} \approx 1$) indicates a region’s existing portfolio strongly predicts future specialisation—the region likely possesses necessary complementary capabilities. Low potential ($p_{r,t,y} \approx 0$) suggests capability gaps that make diversification unlikely under current conditions. Intermediate values reflect genuine uncertainty, where diversification may be possible but contingent on other factors and conditions beyond current portfolio composition. Such factors could be industrial alignment, national ecosystem characteristics, or spatial context, which the subsequent empirical analysis investigates.

3.5 Feature Importance Technology Space (FITS)

Traditional technology networks rely on patent citations or co-occurrence patterns, each with limitations. Citations suffer from examiner additions that may not reflect actual knowledge flows, aggregation from patent-level to technology-level that obscures directionality, and backward-looking orientation that captures historical influence rather than predictive relationships (Fessina et al. 2024). Co-occurrence networks inherit the problems outlined in our critique: symmetry, noise, and linear assumptions.



3 The interconnectedness between regions and technologies

FITS addresses these by constructing an asymmetric, predictive network from the feature importance scores of our Technological Potential models. Rather than inferring relationships from co-occurrence, FITS extracts directional dependencies revealed by which technologies predict others' future adoption.

Recall that for each technology i , we trained a Random Forest model predicting $z_{r,i,y}$ using $\{\text{RTA}_{r,t,y-\delta} : t \neq i\}$ as features. The feature importance $I_i(t)$ quantifies how much technology t contributes to predicting future specialisation in technology i across all regions and time periods.

We formalise FITS as a directed, weighted network $G = (V, E, W)$ where:

- **Nodes** $V = \mathcal{T}$: the set of 641 technologies
- **Edges** E : directed connections $(t \rightarrow t')$ for all $t, t' \in \mathcal{T}, t \neq t'$
- **Weights** $W_{t \rightarrow t'} = I_t(t')$: feature importance of technology t in the model predicting technology t'

To ensure that the incoming edge weights sum to 1 for each technology and facilitate comparisons in later stages, we normalise as follows:

$$W_{t \rightarrow t'} = \frac{I_t(t')}{\sum_{t' \neq t} I_i(t')}$$

The defining property of FITS is asymmetry: $W_{t \rightarrow t'} \neq W_{t' \rightarrow t}$ in general. This asymmetry captures hierarchical technological dependencies invisible to symmetric measures:

- $W_{t \rightarrow t'} \gg W_{t' \rightarrow t}$: Technology t is a prerequisite or enabler for t' . Expertise in t predicts future development of t' , but not vice versa. This pattern identifies foundational technologies and developmental sequences.
- $W_{t \rightarrow t'} \approx W_{t' \rightarrow t}$: Technologies are complementary peers with mutual predictive relationships. This can indicate shared capability requirements or co-evolution.
- $W_{t \rightarrow t'} \ll W_{t' \rightarrow t}$: Technology t' enables t , this reverses the first hierarchical relationship.

3 *The interconnectedness between regions and technologies*

The contrast between relatedness and FITS is sharpest when examining specific technologies. Consider H01L (semiconductors) and D06Q (textile decoration) respectively shown in Figure 3.3 and Figure 3.4.

For semiconductors, relatedness identifies technologies that frequently appear together in regional portfolios. Some connections are sensible: semiconductor memories (H10B), packaging technologies (B68B). But the measure also surfaces implausible links—textile decoration (D06Q), woodworking (B27H), fusion reactors (G21B)—that probably appear not because of genuine technological affinity but because they happen to co-occur in diverse portfolios. This is exactly the noise problem we discussed earlier. Frequency-based aggregation treats all co-presence equally, regardless of whether they reflect shared capabilities or statistical coincidence.

Our FITS approach reveals a different structure. The models we developed asks a different question than relatedness. Rather than asking which technologies appear in proximity of H01L, we ask which technologies predict future development in H01L, and which technologies H01L predict. The asymmetry is clear: strong edges flow FROM application domains (digital computing G06F, telecommunications H04L, control systems G05F) TO semiconductors (H01L), which indicates that these fields depend on semiconductor capabilities. Weaker reverse edges show that semiconductor expertise does not particularly predict future specialisation into computing or telecommunications. In the FITS network, the importance of H01L’s position is not which technologies are close to it, but the asymmetric dependency map that predicts it. This is not possible to observe in relatedness based networks.

The contrast sharpens further with D06Q (textile decoration) the noisy node from the relatedness network. This is a sparse technology for which co-occurrence produces near-total noise: elevators (B64B), timing devices (G04D), chemical production (C13B). This is simply because D06Q is rare and co-occurs randomly with whatever else rare-technology regions happen to have. FITS filters this effectively. Our approach identifies genuine textile domain relationships—sewing (D05B), textile treatment (D06M), wall coverings (D06N), all share actual functional connections. We can also see from Figure 3.4 that the dominant

3 The interconnectedness between regions and technologies

direction of these links is To D06Q, which means that this technology at least for the top strongest links, is dependent on and have no influence on other technologies. Again, this information is obscured in the relatedness network.

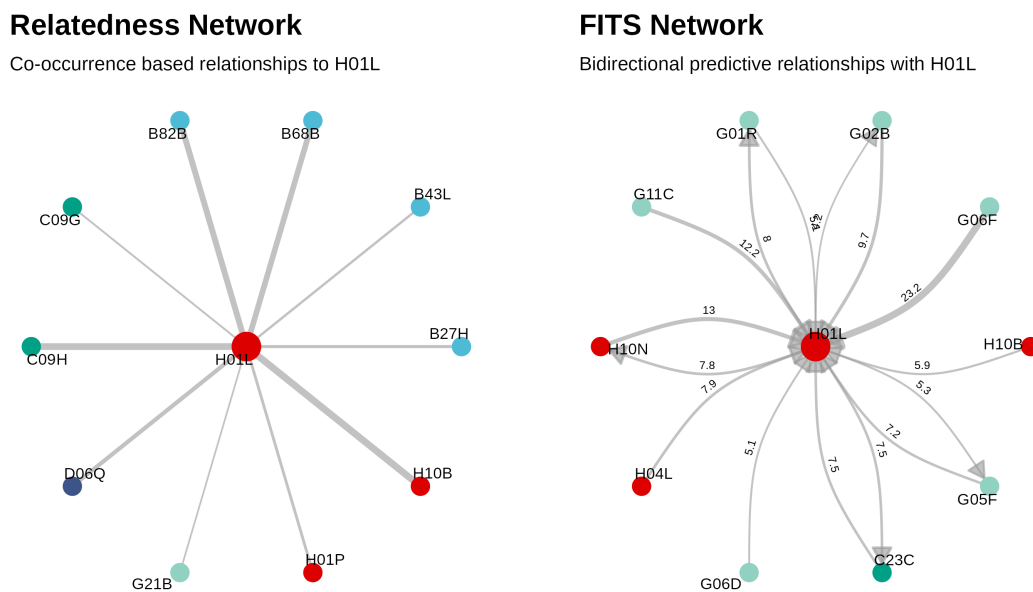


Figure 3.3: Relatedness vs FITS Networks for H10L (based on 2018 patent data)

Our approach reforms the constructs which relatedness is built around simply by shifting the paradigm from description to prediction. Our training strategy results in two alternative constructs: regional technological potential which provides region-specific feasibility estimates, and FITS, a network that reveals the directional, hierarchical structure of technological dependencies. Together, these constructs enable the contextualised analysis our research require.

3.6 Coherence: Bridging technology networks and regional portfolios

Till now we established two methodological constructs throughout the previous sections based on our machine learning approach. The first construct is FITS, which is our

3 The interconnectedness between regions and technologies

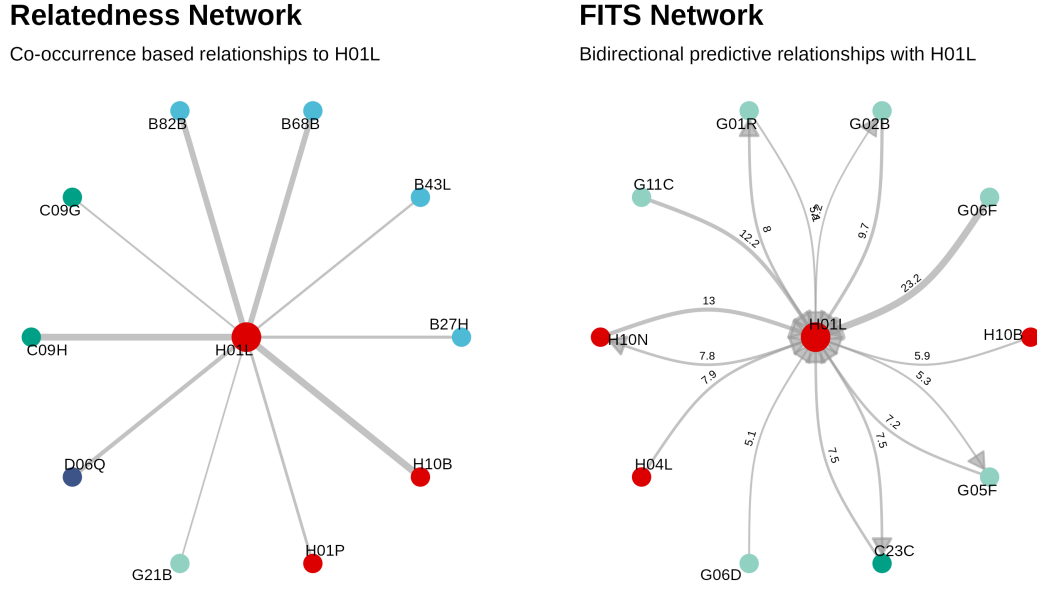


Figure 3.4: Relatedness vs FITS Networks for D06Q (based on 2018 patent data)

baseline for identifying directed technology-to-technology dependencies. Analogous to the relatedness approach, we use these dependencies to construct the technology network. The second construct, is Regional Technological Potential, which is our baseline for quantifying region-specific feasibility of future specialisation in a given technology. Similar to FITS, Potential is an analogue to the relatedness density construct. FITS and Potential each target a specific layer of information, the first targets technologies, and the second focuses on region's feasible specialisations.

To bridge these two layers of information we introduce here the Coherence measure. Coherence, links these two layers by quantifying the industrial alignment between a specific technology and a region's technological portfolio. In here we consider the first digit of the IPC classification as an industry (eg: semiconductor memories (H10B) belongs to section H, Electricity). To build the coherence measure we first examine industrial embeddedness. For a technology in the FITS network this is the (average) strength of it links (incoming and outgoing) with each industry. Essentially, we collapse the hierarchy of each node linked to a specific technology to their respective section numbers and that represents the industrial embeddedness of a given technology to each industry in both directions. For a

3 The interconnectedness between regions and technologies

region, we simply aggregate the industrial embeddedness of each technology present in its portfolio, again in both directions. This inform us on how much each technology and each regional portfolio is embedded in each industry for each direction. With that, the challenge becomes how to compare these 4 dimensions? How can we compare the embeddedness of a region's portfolio to that of a technology? In this section we formalise the answer to these questions based on our coherence measure.

The logic we just portrayed is not novel as it follows core ideas in the literature first established in Teece et al. (1994) and later in Nesta and Saviotti (2005). This idea of coherent diversification, argues that the performance of a diversification strategy is shaped not only by “how much” or “what” an actor does, but also by how the activities in its portfolio hang together. Regional coherence in this sense is a structured property of the knowledge base, or as we referred to in the introduction as knowledge infrastructure. This logic aligns with recombinant views of innovation, the idea that knowledge advances through combinations among related elements rather than as a homogeneous “stock” input (Quatraro 2010). This was also established at the firm level, two firms with the same number of technological fields can differ substantially depending on whether those fields form a coherent block or scattered knowledge infrastructure, coherence captures that internal structure (Pugliese, Cimini, et al. 2019).

However, in this context there's recurring open issue in the literature regarding the meaning and effect of coherence. Recent studies shows that coherence can correlate positively with growth at metropolitan scales but turn negative at larger regional or national scales (Straccamore, Bruno, and Tacchella 2025). This means that at broader scales coherence may signal path-dependency and lock-in rather than productive focus. This disagreement with other relevant studies (Quatraro 2010; Pugliese, Cimini, et al. 2019) raises another challenge as it creates a practical problem for regional diversification analysis: if coherence behaves differently across scales, we need a measure that keeps coherence informative at the regional level without letting it collapse into a purely macro “lock-in” indicator.

Our solution is to bridge the scale disagreement by measuring coherence for regions at a smaller internal scale: instead of computing coherence as a single region-wide property, we

Table 3.1: Heterogeneity of coherence at similar potential deciles (2009)

region	potential	coherence	decile pot	decile coh
FR10	0.001	0.926	1	10
SK01	0.001	0.742	1	8
B64U: UNMANNED AERIAL VEHICLES				

compute coherence within broad industries (IPC sections, i.e. 1st digit code). The intuition is simple: regional portfolios can be coherent in some industries and not others, and those within-industry alignments are precisely what should matter for whether a region can realistically leverage a technology’s prerequisites and downstream connections in FITS.

To illustrate this further, let’s consider two regions, both lacking specialization in technology t , both having similar potential $p_{r,t,y}$. However, region A’s portfolio is well connected to industries which the target technology t is well connected to. While Region B’s portfolio, is well connected to industries that the target technology t is not. Coherence captures this difference, region A should have high coherence with technology t because of the structural knowledge infrastructure in A and t align. Whereas region B would have low coherence with t because of the mismatch in their respective knowledge infrastructure. Let’s consider the example in Table 3.1 of two regions, both are capital regions: Île-de-France (FR10, France) and Bratislava (SK01, Slovakia), both in the lowest decile of potential (first decile) for technology B64U (unmanned aerial vehicles and equipments) in 2009. Despite their clear resemblance in terms of their potential, they have different coherence levels for this technology (0.926(D_{10}) for FR10, and 0.742(D_8) for SK10). To further understand this slight divergence, we can examine their embeddedness profiles as shown in Figure 3.5. The right side panel shown the bidirectional embeddedness of B64U, we can notice that this technology relies uniformly in both directions on industry sections H (Electricity), G (Physics), and F (Mechanical engineering), and similarly but only for the incoming edges on C (Chemistry and metallurgy). In the left side panel, we can compare the industrial profile of the regions. In here, we can see that FR10’s portfolio is strongly and broadly embedded across precisely these sections. In contrast, SK01’s profile is heavily embedded in B (Performing operations and transporting) which is not a focal industry for B64U. To put

3 The interconnectedness between regions and technologies

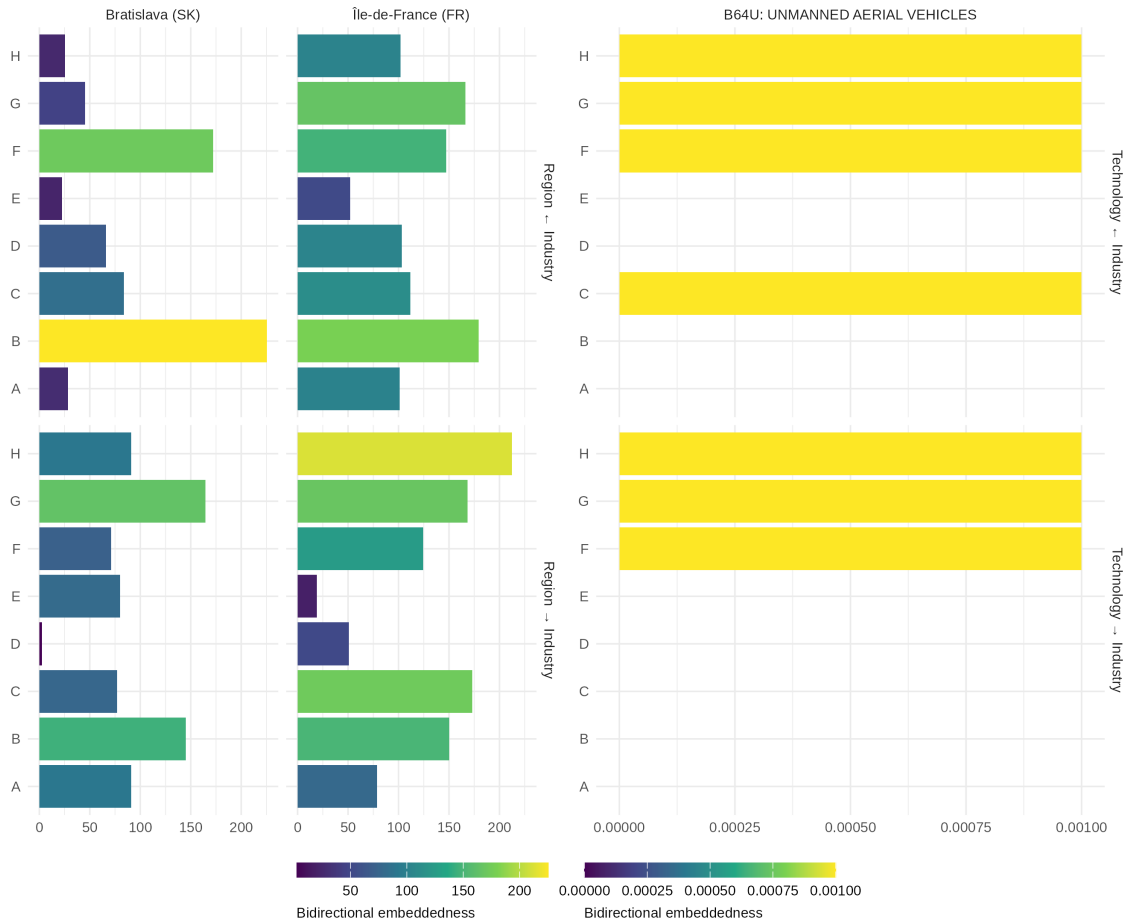


Figure 3.5: Bidirectional Embeddedness: profile of a region vs a technology

3 The interconnectedness between regions and technologies

it simply, FR10 coherence with B64U is higher than SK01’s because the former’s profile is more balanced and aligned around every focal industry B64U is embeded with in both directions. Our coherence measure, captures these asymmetries and provides a substantial layer of context.

In this sense, the coherence metric we propose here operationalizes the “knowledge coherence” and “cognitive proximity” concepts from innovation literature (Neffke, Henning, and Boschma 2011; Boschma 2015), and adds an extra layer of context in the information it captures by leveraging our directional network structure. With this in mind, we use this metric to test whether regional technological potential (predicted feasibility) depends on the alignment in knowledge infrastructure between regional portfolios and technologies.

To formalise our proposed metric, let $t \in \mathcal{T}$ denote the target technology, $r \in \mathcal{R}$ the region, and $c \in \mathcal{C}$ an IPC section (first digit). FITS weights $W_{t \rightarrow t'}$ are as defined in the FITS section. We use t' to index technologies distinct from t throughout. For each region r , technology t , and IPC category/industry c , we construct two embeddedness vectors.

For target technology t and IPC section c , the **incoming** and **outgoing** industrial embeddedness measure the average FITS edge weight between t and the technologies belonging to section c :

$$E_{t,c}^{\leftarrow} = \frac{\sum_{t' \in c, t' \neq t} W_{t' \rightarrow t}}{|\{t' \in c : W_{t' \rightarrow t} > 0\}|}$$

$$E_{t,c}^{\rightarrow} = \frac{\sum_{t' \in c, t' \neq t} W_{t \rightarrow t'}}{|\{t' \in c : W_{t \rightarrow t'} > 0\}|}$$

where $E_{t,c}^{\leftarrow}$ captures how strongly technologies in section c predict future specialisation in t , and $E_{t,c}^{\rightarrow}$ captures how strongly t predicts future specialisation in technologies belonging to section c . If no technology in section c connects to t in either direction, the corresponding scalar is set to zero. Since FITS weights are non-negative by construction, all embeddedness scalars satisfy $E_{t,c}^{\leftarrow}, E_{t,c}^{\rightarrow} \geq 0$.

3 The interconnectedness between regions and technologies

Further, we define the set of technologies in section c in which region r is specialised:

$$S_{r,c} = \{t' \in c : \text{RTA}_{r,t'} \geq 1\}$$

The **regional average embeddedness** in section c averages the embeddedness scores of those specialised technologies:

$$E_{r,c}^{*\leftarrow} = \frac{1}{|S_{r,c}|} \sum_{t' \in S_{r,c}} E_{t',c}^{\leftarrow}$$

$$E_{r,c}^{*\rightarrow} = \frac{1}{|S_{r,c}|} \sum_{t' \in S_{r,c}} E_{t',c}^{\rightarrow}$$

If $S_{r,c} = \emptyset$, then set $E_{r,c}^{*\leftarrow} = E_{r,c}^{*\rightarrow} = 0$. Meaning that the region has no foothold in section c and therefore no alignment signal in that domain.

The original formula computes cosine similarity between two vectors, each of which **pairs the technology's embeddedness with the region's embeddedness** – one vector for the incoming direction, one for the outgoing:

$$\text{Coherence}_{r,t} = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\| \cdot \|\mathbf{v}_2\|}$$

where:

$$\mathbf{v}_1 = [E_{t,c}^{\leftarrow}, E_{r,c}^{*\leftarrow}]_{c \in \mathcal{C}} \in \mathbb{R}_{\geq 0}^{2|\mathcal{C}|}$$

$$\mathbf{v}_2 = [E_{t,c}^{\rightarrow}, E_{r,c}^{*\rightarrow}]_{c \in \mathcal{C}} \in \mathbb{R}_{\geq 0}^{2|\mathcal{C}|}$$

So $\mathbf{v}_1$ concatenates the incoming embeddedness of technology t and region r for each industry, and $\mathbf{v}_2$ does the same for the outgoing direction. The cosine similarity between

them captures how consistently the incoming and outgoing industrial signals align across t and r jointly. Since all entries are non-negative, $\text{Coherence}_{r,t} \in [0, 1]$.

Conceptually, this is close to how coherent diversification measures compute “how surrounded a technology is” by related portfolio elements, but with two key differences. First, we ground proximity in predictive, directed dependencies (FITS) rather than symmetric co-occurrence similarity. Second, we compute coherence at an internal domain scale (IPC section) to keep the measure meaningful for regional analysis under known scale heterogeneity.

In here, our coherence measure ranges from 0 to 1. In our context, high coherence (≈ 1) means that for technology t ’s embeddedness signal in different categories/industries closely matches the average industrial embeddedness of region r ’s portfolio of technologies. The region’s existing industrial capabilities align with the structural industrial prerequisites and consequences of technology t . Alternatively, a neutral coherence (≈ 0) reflects a misalignment in industrial embeddedness between technology i ’s and the region.

To our aim, coherence serves two roles in this work. First, as dyadic fit measure: it is a direct indicator of portfolio–technology alignment in a given domain or industry. It captures whether predicted feasibility is supported by an internally coherent knowledge infrastructure. Second, as moderation channel: it allows us to test whether the effect of technology network position (e.g., bridging technologies, enabling technologies) depends on whether regions have the industry-level knowledge infrastructure to exploit those positions. This is an explicitly contextual mechanism that simple diversification counts cannot reveal (Quatraro 2010; Boschma 2015) and aligns with the core idea of this work.

3.7 Summary

This chapter develops the methodological foundation for contextualised analysis of regional technological diversification. We reconstruct two core constructs from the REC literature and introduce a novel bridging measure.

3 The interconnectedness between regions and technologies

We open by identifying three compounding structural limitations in the mainstream approach: co-occurrence measures produce noisy estimates when technologies outnumber regions, symmetric normalisation obscures the directional nature of technological dependencies, and linear aggregation of relatedness density discards the combinatorial interactions among capabilities that matter most for distinguishing feasible from infeasible diversification paths.

We respond to these limitations by shifting from descriptive correlation to supervised prediction using Random Forest. By training a separate classifier for each target technology, we derive two novel constructs. We propose Regional Technological Potential ($p_{r,t,y}$) as a replacement for relatedness density – a region-specific probability of future specialisation grounded in out-of-sample predictive performance rather than co-occurrence frequencies. We further propose FITS to replace symmetric technology networks with a directed, weighted graph that exposes hierarchical relationships, enabling technologies, and developmental sequences invisible to symmetric measures. Through the semiconductor versus textile decoration example, we show concretely how FITS filters noise and reveals asymmetric dependencies that co-occurrence conflates.

We conclude by introducing Coherence, a measure we propose to bridge the technology-level information in FITS and the region-level information in Technological Potential. Drawing on the coherent diversification literature, and motivated by evidence that region-wide coherence measures behave inconsistently across spatial scales, we compute Coherence within IPC sections rather than across the full portfolio. We design it to capture whether a region’s industrial embeddedness profile – as revealed by the directional FITS structure – aligns with the industrial embeddedness profile of a target technology. This makes Coherence both a dyadic fit indicator and a moderation channel: we use it to test whether the effect of a technology’s network position on diversification feasibility of a region depends on how well that region’s industrial structure aligns with the industrial capability dependencies of that technology.

Together, the three measures we develop operationalise a conceptual shift: from asking *what* a region might diversify into, toward asking under what structural context diversification

3 The interconnectedness between regions and technologies

is feasible. This is the core analytical problem we address in the subsequent chapters.

4 Empirical design

In the preceding chapters we discussed in more details the need for contextualised understanding of regional diversification and we argued that such extra context should account both endogenous and exogenous factors. Analogous to the research agenda in the EEG literature we highlighted that this contextualisation is mainly dependent of three channels: relatedness, network structures, and institutions.

In the previous chapter we went a step further and explained why we actually need to go beyond the classical proximity and relatedness paradigm since relatedness and network structure are almost identical constructs. The idea stems from their conceptual and methodological limitations which we summarised in three points: (1) symmetric co-occurrence measures obscure directional technological dependencies, (2) linear aggregation of relatedness density discards combinatorial information about how capabilities interact, and (3) co-occurrence measures yield noisy proximity metric because what's measured almost always have higher dimensions than what is observed.

Because of these issues, we proposed a machine learning training strategy that consider relatedness as predictive exercise where we use random forest to estimate alternative measures of proximity: regional technological potential our analogue to relatedness density, and the Feature Importance Technology Space (FITS) our analogue to the classical Technology Space (and relatedness). Additionally, we also defined a coherence measure following intuition in the literature that we previously discussed.

In this chapter we provide an empirical application to our methodological contribution following the same ideas we discussed in the literature review. In here we will first re-iterate

and further clarify our objectives and how they translate to empirical constructs, then we will explain progressively how we will to address the research questions and the test the hypothesis.

4.1 Overview and identification goals

The central question of this chapter is what explains variation in regional technological potential? The analogue to this question in the REC literature is what explains relatedness. The reason this question is asked at all stems from a large body of studies that show that relatedness density, is a major predictor of entry to new technologies. The more related a region is to a technology the higher the probability of future entry.

Similarly, relatedness density is also a major predictor of economic complexity. This is to say that the technological fabric of a region is extremely correlated with growth. In contrast, C. Pinheiro (2025) and Andreoni and Chang (2019) mentioned a structural gap in the way these ideas were addressed in the literature. Specifically, that conventional proximity assessment primarily provide a supply-side representation of what regions might do given their existing portfolios, while remaining comparatively silent on when and where (Hidalgo 2023) those feasibility conditions are strengthened or weakened by network structure, institutional environments, and spatial interdependence.

Moreover, because our contribution is analogue to relatedness we find it sensible to ask the same question again and extend our methodological contribution to an empirical one as well. We follow the agenda we discussed in the literature review: A contextualised assessment of diversification possibilities through a more contextualised proximity measure (potential), network structure (FITS), industrial alignment (coherence), institutions (Entrepreneurial Ecosystem), and spatial dependence.

To explore this question within the framework of this work, we provide two complementary designs through three modeling approaches each target a specific area of the question. The first empirical design is dyadic, which means it addresses the region-technology as a dyadic

structure. In the dyadic design we assess how the position of technologies in the FITS network influence regional technological potential, and whether this influence is universal or dependent on the regional industrial structure, where this dependence is operationalised by our coherence measure. We then extend this by assessing how the inclusion of the entrepreneurial ecosystem covariates affect these relationships. The second design addresses spatial dependence, where we aggregate our covariate into a regional panel and assess whether geographical proximity adds an additional dimension to the feasibility question, however, this layer is different from the dyadic layer and warrants different interpretation because we focus mainly on the aggregate portfolio mechanisms. In here we are mainly interested in understanding how the endowment of enabling or bridging technologies in a given regional portfolio influence the general feasibility of the entire portfolio, i.e.: in all other existing technologies. Finally, as a secondary benchmark, we also compare the influence of potential and relatedness as contrasting construct on regional entry into new technologies.

4.2 Research design

The guiding research question is what conditions determine the feasibility of regional technological diversification, and how do technology-specific characteristics, portfolio alignment, national ecosystem structure, and spatial interdependence jointly shape that feasibility. This decomposes into five testable claims which we re-iterate from the first chapter and summarise in Table 4.1.

H1 (Benchmark): Regional technological potential is at least as informative as relatedness density in predicting technology entry. The literature has established relatedness density as a robust predictor of entry (Hidalgo and Hausmann 2009; Neffke, Henning, and Boschma 2011). We test whether our capability-based potential measure subsumes, supplements, or displaces this predictive power. This is not a causal claim; it is a precision benchmark that validates the methodological contribution and motivates the subsequent decomposition.

H2 (Technology characteristics): Technologies occupy structurally distinct positions in

the directed capability network, and these positions predict how feasible it is for regions to diversify into them. We distinguish two archetypal roles following Antonietti and Montresor (2021) and Montresor and Quatraro (2017). First, technologies with high betweenness centrality act as bridges between otherwise distant capability domains, connecting clusters and enabling recombination across areas of the network. Second, technologies with high eigenvector centrality occupy authority-like positions, being well connected to other well-connected technologies, and represent capability-dense domains whose entry tends to require broader complementary competence. We therefore test two sub-claims:

- **H2a (Bridging role):** Technologies with a bridging role are positively associated with regional technological potential, meaning that bridge technologies represent more accessible diversification pathways for regions.
- **H2b (Enabling role):** Technologies with an enabling role are negatively associated with regional technological potential, meaning that foundational technologies represent structurally more difficult diversification targets.

H3 (Portfolio alignment): The value of a technology’s structural position may depend on whether the region’s portfolio is aligned with the capability demands that position implies. This draws on the recombination logic in evolutionary economic geography, which holds that regions diversify by recombining existing capabilities, and that feasibility depends on whether the regional portfolio provides the relevant interfaces needed to support entry (Frenken and Boschma 2007; Boschma 2017). Relatedness captures an important but incomplete dimension of this (Antonietti and Montresor 2021). Our coherence measure extends this logic by capturing the alignment between a technology’s position in the directed capability space and the region’s industrial structure. We test whether this alignment moderates the relationship between network position and potential, and whether this moderation operates differently for bridging and enabling technologies, through three sub-claims:

- **H3a (Alignment main effect):** Regions with more coherent portfolios have higher regional technological potential, meaning that portfolio alignment reduces feasibility

barriers across the technology space.

- **H3b (Bridging moderation):** Technologies with a bridging role are positively associated with higher potential for regions with low coherence than for regions with high coherence, meaning that bridge technologies matter most as diversification pathways when direct alignment is limited.
- **H3c (Enabling moderation):** Technologies with an enabling role are positively associated with higher potential for regions with high coherence than for regions with low coherence, meaning that enabling technologies matter most as upgrading pathways when direct alignment is high.

H4 (National ecosystem): National entrepreneurial ecosystem characteristics shape regional technological potential beyond what dyadic portfolio and network mechanisms account for. Portfolio configuration describes the endogenous opportunity structure of a region, but whether those opportunities are realised depends partly on enabling conditions that operate at the national level, including finance, competition, innovation capacity, and institutional characteristics (Acs, Autio, and Szerb 2014; C. Pinheiro 2025). We test whether national ecosystem measures explain variation in regional potential after accounting for portfolio structure and technology network mechanisms. Effects are expected to be heterogeneous across ecosystem dimensions.

H5 (Spatial structure): Moving from the technology-region dyad to the regional portfolio as a whole, we ask whether the aggregate composition of a region’s portfolio predicts average regional potential, and whether this relationship is spatially interdependent. If high-potential regions systematically neighbour other high-potential regions after conditioning on observed portfolio characteristics, this indicates genuine interregional interdependence rather than merely correlated endowments (Boschma and Iammarino 2009). We test this through three sub-claims:

- **H5a (Spatial dependence):** Regions with higher average potential tend to neighbour other high-potential regions, meaning that regional feasibility landscapes are spatially interdependent rather than locally contained.

- **H5b (Bridging portfolio):** Regions with portfolios concentrated in bridging technologies have higher average potential when their aggregate portfolio coherence is sufficient, meaning that the portfolio-level value of bridging technologies depends on whether the aggregate industrial alignment supports them.
- **H5c (Enabling portfolio):** Regions with portfolios concentrated in enabling technologies have higher average potential which increases as coherence increase, meaning that foundational technology endowments are a structural portfolio-level advantage, amplified further by coherence.

Finally, we note here that because H5 is a different level of analysis than the other dyadic level hypothesis, its interpretation is also different. We warrant this note given the contrasting phrasing between H3 and H5, we will get back in details to this in the results chapter.

4.3 Data

We use data very similar to many studies discussed in the literature review, namely the European patent applications from the EPO/PATSTAT database, covering 641 technologies (4-digit IPC classes) across 345 NUTS2 regions in 30 European countries over 11 years (2008–2018). The dyadic models (H1–H4) use approximately 1.9 million technology–region–year observations. The spatial models (H5) aggregate to approximately 3,200 observations. Patent data serve as a standard proxy for regional knowledge production and technological specialisation. We briefly described in the previous chapter how we use this data to construct our training strategy and build our empirical constructs (Regional Technological Potential, and the Feature Importance Technology Space (FITS)). In here we take this a step further and describe in more details the other variables we use in our design. Descriptive statistics and correlation matrix for all dyadic variables are reported respectively in Table 4.2 and Table 4.3.

Table 4.1: Hypotheses, Variables, and Empirical Tests

Hypothesis	Claim	Key variables	Model
H1	Potential is at least as informative as relatedness density in predicting entry	pred11 vs relatedness_density	Model 0
H2a	Technologies with a bridging role are positively associated with potential	betweenness	Model 1
H2b	Technologies with an enabling role are negatively associated with potential	eigenvector	Model 1
H3a	Coherence positively predicts potential	coherence	Model 1
H3b	Bridging technologies predict higher potential for low-coherence regions	betweenness x coherence	Model 1
H3c	Enabling technologies predict higher potential for high-coherence regions	eigenvector x coherence	Model 1
H4	National ecosystem characteristics shape potential beyond portfolio mechanisms	GEI pillars x8-x14	Model 2
H5a	Regional potential is spatially interdependent across neighbouring regions	Spatial lag (SAR)	Model 3
H5b	Bridging portfolio concentration predicts higher potential under sufficient coherence	avg_betweenness x avg_coherence	Model 3
H5c	Enabling portfolio concentration predict higher potential and is amplified by coherence	avg_eigenvector x avg_coherence	Model 3

Note: Model 0 is the entry benchmark (LPM). Models 1 and 2 are dyadic fixed effects and hierarchical random effects respectively. Model 3 is the spatial autoregressive (SAR) panel estimated on region-year aggregates.

4.3.1 Outcome: Regional technological potential

The primary outcome in this work is Regional Technological Potential $p_{r,t,y}$: the Random Forest-predicted probability that region r develops revealed technological advantage ($RTA \geq 1$) in technology t at year y , estimated from portfolio characteristics four years prior. Values are bounded between 0 and 1.

Because OLS on a bounded outcome is problematic and can generate out-of-range fitted values, we apply a logit transformation before estimation:

$$p_{r,t,y}^* = \text{logit}(p_{r,t,y}) = \log\left(\frac{p_{r,t,y}}{1 - p_{r,t,y}}\right)$$

with extreme values trimmed to avoid numerical instability. The transformed outcome $p_{r,t,y}^*$ is then unbounded and appropriate for OLS estimation. Ideally, a beta or a fractional logit regression would be more appropriate given the bounded nature of the outcome. We did test these approaches during the modeling process and we decided that given the size of the data and the complexity of the specifications the computational cost is great and numerical stability becomes an issue. Additionally, our decision to use OLS estimation is also motivated by the need for a clear interpretation of the results.

For the entry benchmark (H1), we use a standardised version of the untransformed potential, $\tilde{p}_{r,t,y}$ (z-scored), as a predictor of next-period entry alongside standard relatedness density (z-scored).

For the spatial models (H5), we aggregate to region-year means (the geographic distribution of average potential across European regions is shown in Figure 4.1.):

$$P_{r,y} = \frac{1}{|\mathcal{J}|} \sum_{t \in \mathcal{J}} p_{r,t,y}$$

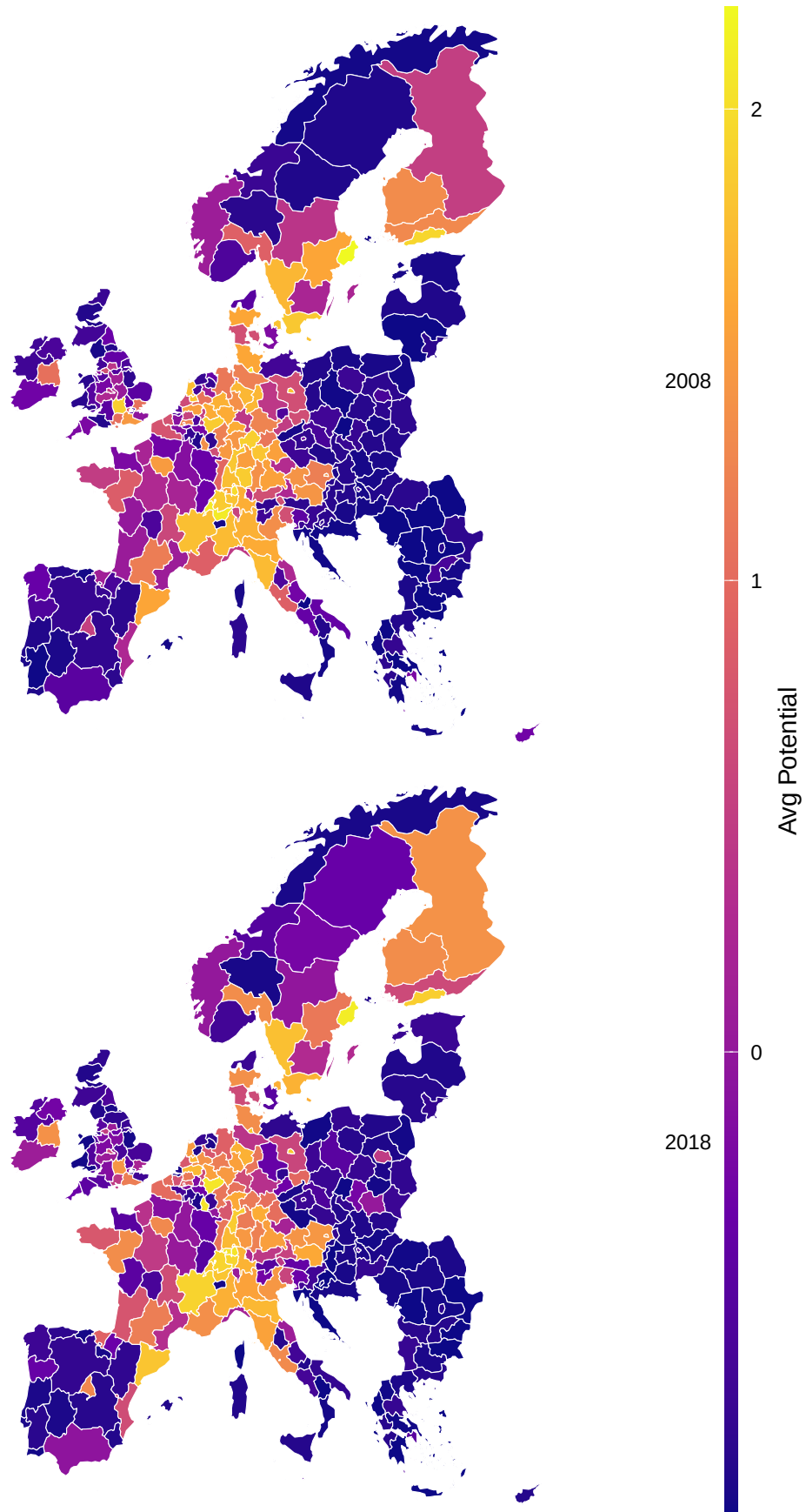


Figure 4.1: Average Regional Technological Potential (2008 vs 2018)

4.3.2 Technology–region level variables

Coherence ($\text{coherence}_{r,t,y}$) measures the alignment between the regional technology portfolio and a specific technology in the directed FITS network in terms of industrial embeddedness (see the previous chapter for more details). It captures whether the regional portfolio is structurally positioned across industries to support a given technology’s embeddedness in those industries. i.e., whether the region and the technology “sit” in compatible parts of the capability system. This aligns with the intuition in the literature that diversification feasibility depends on how well the target technology aligns with internal regional knowledge infrastructure base (Teece et al. 1994; Nesta and Saviotti 2005; Quatraro 2010; Pugliese, Napolitano, et al. 2019; Straccamore, Bruno, and Tacchella 2025). The geographic distribution of average coherence is shown in Figure 4.2.

Distance ($\text{rawdist}_{r,t,y}$) is the shortest path weighted length in the FITS network from the closest technology in the regional portfolio to a given technology regardless of the direction of the links. This distance measures topological accessibility in the network or how many steps separate the region’s current frontier from a given opportunity. This variable operationalises the product-space logic of Hidalgo et al. (2007) who show that the probability of entry decays with distance in the product space.

Patent count ($\text{patents}_{r,t,y}$) is the number of patent applications attributed to region r in technology t at year y . We use it to controls for pre-existing knowledge stock on a given technology in the region. This is to ensure that coherence and network position effects are do not turn into proxies for size as argued in Bathelt and Storper (2023). The geographic distribution of total patent activity is shown in Figure 4.6.

4.3.3 Technology level variables

Betweenness centrality¹ (betweenness_t) measures how frequently a technology lies on shortest directed paths between other technologies. High-betweenness technologies act as

¹the FITS network is directed, in earlier iterations explored HITS authority and hub scores separately but because HITS scores and betweenness are correlated we decided to use eigenvector centrality in the final specification.

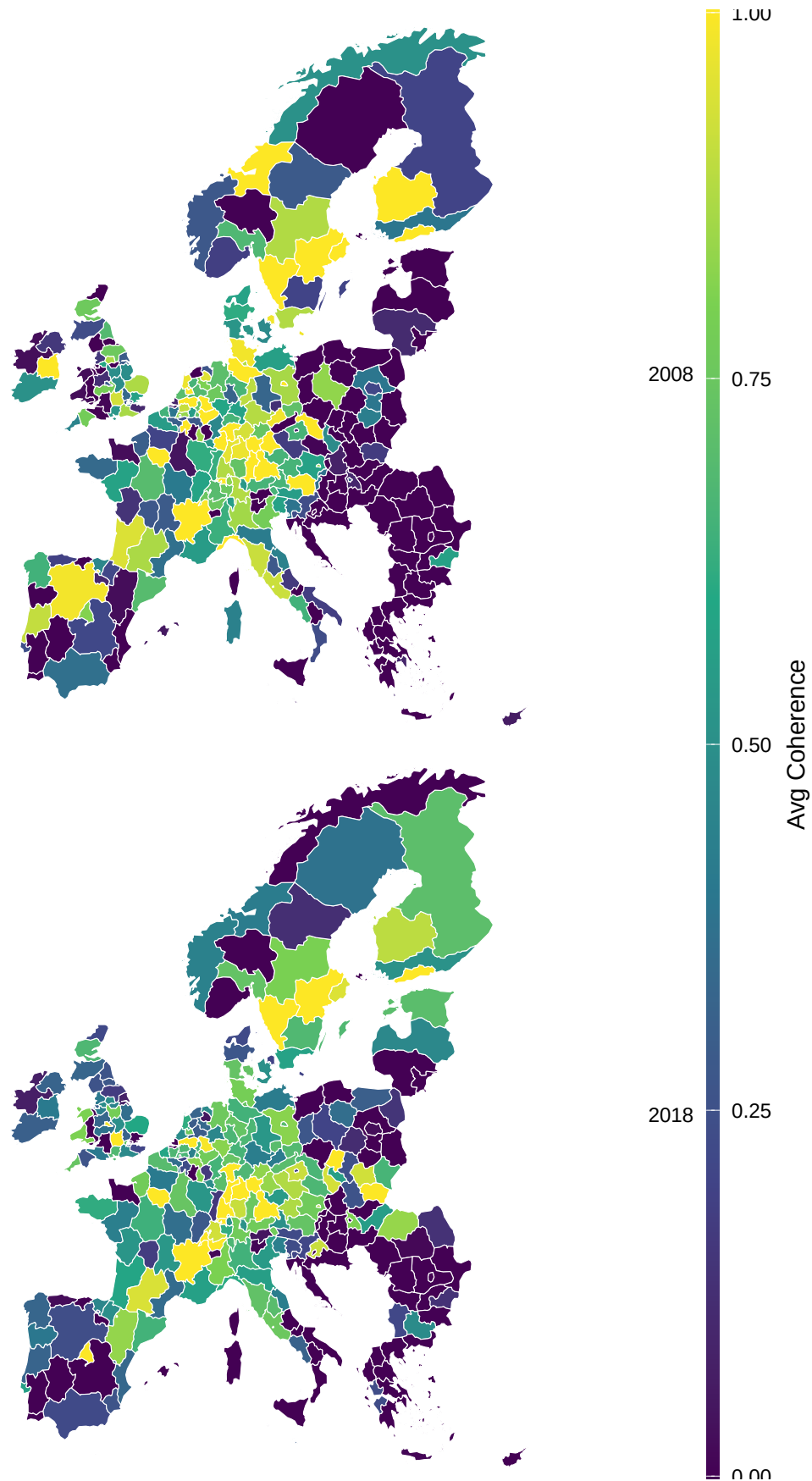


Figure 4.2: Average Coherence Across Regions (2008 vs 2018)

Table 4.2: Summary statistics of the Dyad level data

	Unique	NA	Mean	SD	Min	Median	Max
pred1	1672549	0	-0.00	1.00	-0.88	-0.43	3.95
patent_n	2810	0	-0.00	1.00	-0.10	-0.09	183.30
coherence	1985141	0	0.00	1.00	-2.54	0.08	2.01
betweenness	4132	0	0.00	1.00	-0.54	-0.46	10.79
eigenvector	7053	0	-0.00	1.00	-0.70	-0.41	6.63
rawdist	5624	4	0.00	1.00	-0.53	-0.18	12.22
relatedness_density	45315	4	0.00	1.00	-1.23	-0.20	11.44
reg_comp	3087	4	-0.00	1.00	-13.00	0.08	5.45
tech_comp	6833	4	0.00	1.00	-23.17	0.01	4.64
entry	3	18	0.06	0.23	0.00	0.00	1.00

bridges between otherwise separated domains of knowledge. In our interpretation, this is the structural analogue of GPT-like bridging roles where technologies that facilitate transitions across domains serve as an intermediary (Antonietti and Montresor 2021). The geographic distribution of average betweenness centrality is shown in Figure 4.4.

Eigenvector centrality (eigenvector_t) measures how strongly connected a technology is to other highly connected technologies. The idea is to capture authoritative positions in the capability system. In our interpretation, this corresponds to KET-like enabling roles capability-dense, strongly interdependent domains whose entry tends to require broader complementary competence rather than a single bridge. The corresponding distribution for eigenvector centrality is shown in Figure 4.5.

Regional complexity ($\text{reg_comp}_{r,y}$) and **technology complexity** (tech_comp_t) are included as controls following the economic complexity tradition in which diversity–ubiquity patterns proxy underlying capability endowments (Hidalgo and Hausmann 2009). This mirrors the rationale for including regional complexity controls in related European diversification work (Antonietti and Montresor 2021). The geographic distribution of regional complexity is shown in Figure 4.7.

Table 4.3: Correlation matrix of the Dyad level data

pred1	1	.	.	.	.	.	.	.	.	.
rawdist	-.07	1	.	.	.	.	.	.	.	.
patent_n	.24	-.02	1	.	.	.	.	.	.	.
coherence	.40	-.19	.11	1	.	.	.	.	.	.
betweenness	.23	-.10	.05	.40	1	.	.	.	.	.
eigenvector	-.15	.08	-.02	-.21	-.17	1	.	.	.	.
relatedness_density	.69	-.00	.22	.12	.02	-.02	1	.	.	.
reg_comp	.34	.00	.09	.03	-.01	.01	.48	1	.	.
tech_comp	-.20	.05	-.11	-.28	-.24	-.01	-.05	-.00	1	.
entry	.29	-.02	.26	.13	.07	-.05	.28	.09	-.08	1

4.3.4 National level variables: GEI pillars

To test H4, which assumes that national entrepreneurial ecosystem characteristics contribute to regional potential, we incorporate seven pillars from the Global Entrepreneurship Index (GEI). The GEI decomposes national ecosystem characteristics into fourteen that account for individual attitudes, systemic capabilities, and aspirational orientations (Acs, Autio, and Szerb 2014).

For the purposes of this work, we retain pillars x8–x14. The reason here is two folds, first pillars x1-x6 capture individual-level entrepreneurial attitudes² and mostly reflect social preferences of the entrepreneurial ecosystem and not the institutional landscape. Descriptive statistics for the GEI pillars are reported in Table 4.6

²opportunity perception, startup skills, risk acceptance, networking, cultural support, opportunity startup

Table 4.4: Correlation matrix of the GEI pillars

x1_opportunity_perception	1	.	.	.	.	.	.	.	.	.	.	.	.	.
x2_startup_skills	.33	1	.	.	.	.	.	.	.	.	.	.	.	.
x3_risk_acceptance	.68	.41	1	.	.	.	.	.	.	.	.	.	.	.
x4_networking	.71	.33	.57	1	.	.	.	.	.	.	.	.	.	.
x5_cultural_support	.88	.36	.73	.72	1	.	.	.	.	.	.	.	.	.
x6_opportunity_startup	.85	.41	.72	.70	.87	1	.	.	.	.	.	.	.	.
x7_technology_absorption	.58	.21	.56	.42	.54	.64	1	.	.	.	.	.	.	.
x8_human_capital	.46	.17	.37	.26	.43	.53	.40	1	.	.	.	.	.	.
x9_competition	.76	.21	.66	.59	.79	.82	.60	.62	1	.	.	.	.	.
x10_product_innovation	.59	.18	.45	.49	.54	.56	.51	.46	.61	1	.	.	.	.
x11_process_innovation	.60	.33	.63	.63	.60	.59	.57	.30	.63	.63	1	.	.	.
x12_high_growth	.51	.07	.33	.28	.46	.45	.37	.50	.50	.42	.32	1	.	.
x13_internationalization	.01	-.15	.06	-.10	-.02	.03	.14	.17	.21	.30	.25	.42	1	.
x14_risk_capital	.46	.04	.44	.42	.53	.55	.35	.44	.65	.40	.33	.20	.13	1

Table 4.5: GEI Pillars Retained in Analysis

Pillar	Variable name	Role in capability development
Human capital	x8_human_capital	Education and training infrastructure for technology development
Competition	x9_competition	Market pressure driving diversification and upgrading
Product innovation	x10_product_innovation	National capacity for new product development
Process innovation	x11_process_innovation	Technology absorption and adaptation capacity
High growth	x12_high_growth	Scaling orientation enabling specialisation
Internationalisation	x13_internationalization	Export orientation and global knowledge access
Risk capital	x14_risk_capital	Venture financing enabling exploration of new technologies

Second, pillars x1–x6 are heavily correlated ($r = .57$ – $.88$ across all pairwise combinations, as shown in Table 4.4). This is problematic in our case since it will create multicollinearity issues in the models and would compromise numerical stability, fit, and most likely bias the results.

Thus, we retained pillars x8–x14 that show higher heterogeneity and more moderate pairwise correlations. Regardless, the pillars we retain (see Table 4.5) are more informative for our case since they capture the institutional landscape of entrepreneurial ecosystem which is our main area of focus in this work.

Because GEI pillars are measured at the national level and vary primarily across countries rather than within countries over time, they cannot be identified in a two-way region-and-year fixed effects specification – national variation would be absorbed by region fixed effects. This is why H4 is tested in two complementary ways for robustness: (i) in fixed effects models, and (ii) in hierarchical random effects models where are regions nested within countries, where GEI variables operate at the country level as intended.

Table 4.6: Descriptive statistics of the GEI pillars

	Unique	NA	Mean	SD	Min	Median	Max
x1_opportunity_perception	216	0	0.00	0.98	-1.52	-0.22	1.86
x2_startup_skills	225	0	0.00	0.98	-2.43	0.01	2.08
x3_risk_acceptance	218	0	-0.00	0.98	-2.25	0.28	1.65
x4_networking	222	0	0.00	0.98	-1.85	-0.25	2.70
x5_cultural_support	213	0	-0.00	0.98	-1.94	-0.05	1.83
x6_opportunity_startup	210	0	-0.00	0.98	-2.06	-0.18	1.50
x7_technology_absorption	215	0	-0.00	0.98	-2.07	0.01	1.71
x8_human_capital	227	0	-0.00	0.98	-1.89	-0.19	2.49
x9_competition	218	0	-0.00	0.98	-1.53	-0.21	1.70
x10_product_innovation	222	0	0.00	0.98	-1.90	0.08	2.12
x11_process_innovation	226	0	0.00	0.98	-2.31	0.04	1.98
x12_high_growth	227	0	-0.00	0.98	-2.57	0.08	2.04
x13_internationalization	215	0	-0.00	0.98	-2.52	0.11	1.86
x14_risk_capital	222	0	0.00	0.98	-2.10	-0.00	2.75

Table 4.7: Descriptive statistics of the aggregated variables

	Unique	NA	Mean	SD	Min	Median	Max
reg_potential	1941	0	0.00	1.00	-0.99	-0.46	2.34
patent_count	1249	0	776.19	1985.54	0.00	186.00	27222.00
reg_comp	3097	0	0.00	1.00	-13.23	0.03	5.61
avg_coherence	3097	0	0.00	1.00	-10.13	0.26	2.10
avg_betweenness	3090	0	-0.00	1.00	-2.59	-0.03	10.63
avg_eigenvector	3097	0	0.00	1.00	-1.82	-0.06	10.61

4.3.5 Region level variables

For the spatial panel (H5), we aggregate the dyadic variables to region-year means. Specifically we retain average coherence, average betweenness, and average eigenvector of each region's existing portfolio. Additionally, we also aggregate patent counts to the total regional patenting activity across all technologies and keep regional complexity. Each of these variables reflect an aggregate characteristic of the regional technological portfolio. Further descriptive details on these variables are presented in Table 4.7 and Table 4.8.

Table 4.8: Correlation matrix of the aggregated variables

reg_potential	1	.	.	.	.	.	.
patent_count	.51	1	.	.	.	.	.
avg_coherence	.52	.30	1	.	.	.	.
avg_betweenness	-.05	-.09	.20	1	.	.	.
avg_eigenvector	.20	.17	.28	.08	1	.	.
entry_rate	.76	.39	.52	-.00	.21	1	.
reg_comp	.50	.27	.24	-.19	.01	.42	1

4.4 Model 0: Entry Benchmark

Before decomposing regional technological potential into its structural determinants, we establish that potential carries meaningful information about diversification feasibility beyond what traditional relatedness measures capture. This benchmark analysis validates the methodological investment in the machine learning framework by testing whether the added complexity–non-linear capability interactions, region-specific predictions–translates into improved empirical performance.

The outcome in this specification is *observed entry*: whether region r develops specialisation (RTA = 1) in technology t at year $y + 1$, conditional on not being specialised at year y . This is a binary outcome taking the value 1 if the region crosses the specialisation threshold in the subsequent period, and 0 otherwise. We estimate linear probability models (LPM) that include both standardised potential and standardised relatedness density as competing predictors, along with controls for regional and technology complexity.

The general specification is:

$$\text{Entry}_{r,t,y+1} = \beta_1 \text{Density}_{r,t,y} + \beta_2 \text{Potential}_{r,t,y} + \mathbf{Z}_{r,t,y} \gamma + \alpha_y + \varepsilon_{r,t,y}$$

where $\text{Density}_{r,t,y}$ is the standardised relatedness density (z-scored), $\text{Potential}_{r,t,y}$ is the standardised technological potential (z-scored), $\mathbf{Z}_{r,t,y}$ contains controls (regional complexity,

technology complexity, patent count), and α_y are year fixed effects. Standard errors are clustered by region and technology to account for the dyadic structure of the data.

The motivation for this specification is straightforward. Relatedness density has been established as a robust predictor of diversification outcomes across numerous empirical contexts (Hidalgo et al. 2007; Neffke, Henning, and Boschma 2011; Boschma 2017). If our ML-derived potential measure provides no improvement over this standard benchmark – or worse, if density remains the dominant predictor when both are included – then the added machinery of Random Forests, feature importance networks, and coherence calculations would be empirically unjustified. The test is therefore conservative by design: we assess whether potential subsumes, supplements, or is subsumed by the conventional measure. The geographic distribution of average relatedness density is shown in Figure 4.3

4.5 Model 1: Dyadic fixed effects models

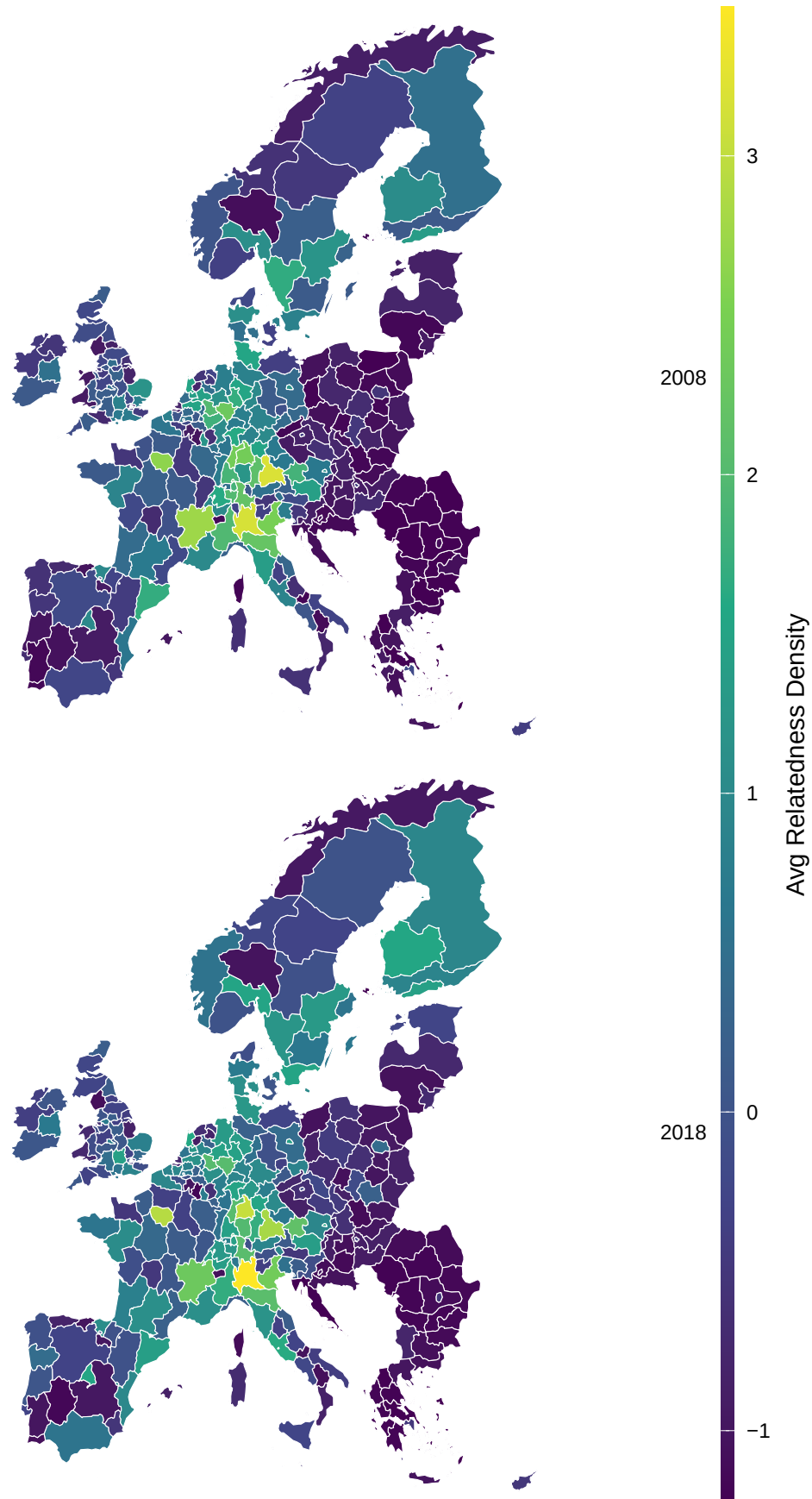
The first model in our work is specified to investigate the explanatory factors that drives regional technological potential. In here, as well as in the other dyadic models we estimate our linear models on the logit-transformed potential $p_{r,t,y}^*$.

Essentially we adopt a two-way region and year fixed effects as our preferred specification to assess the within region variations with standard errors clustered by region and technology. This choice accounts for the structural correlation of the observations in the same region or technology.

The general specification is:

$$p_{r,t,y}^* = \alpha_r + \alpha_y + \mathbf{X}_{r,t,y}\beta + \varepsilon_{r,t,y}$$

where α_r are region fixed effects, α_y are year fixed effects, and $\mathbf{X}_{r,t,y}$ is the covariates matrix.



(a) This map shows the average relatedness density for each region, computed as the mean density score across all non-specialized technologies—a measure of how ‘close’ the average opportunity is to the region’s existing portfolio using traditional symmetric co-occurrence-based relatedness. High density regions (bright colors) have diversified portfolios where most candidate technologies are related to what the region already does, indicating a dense

4 Empirical design

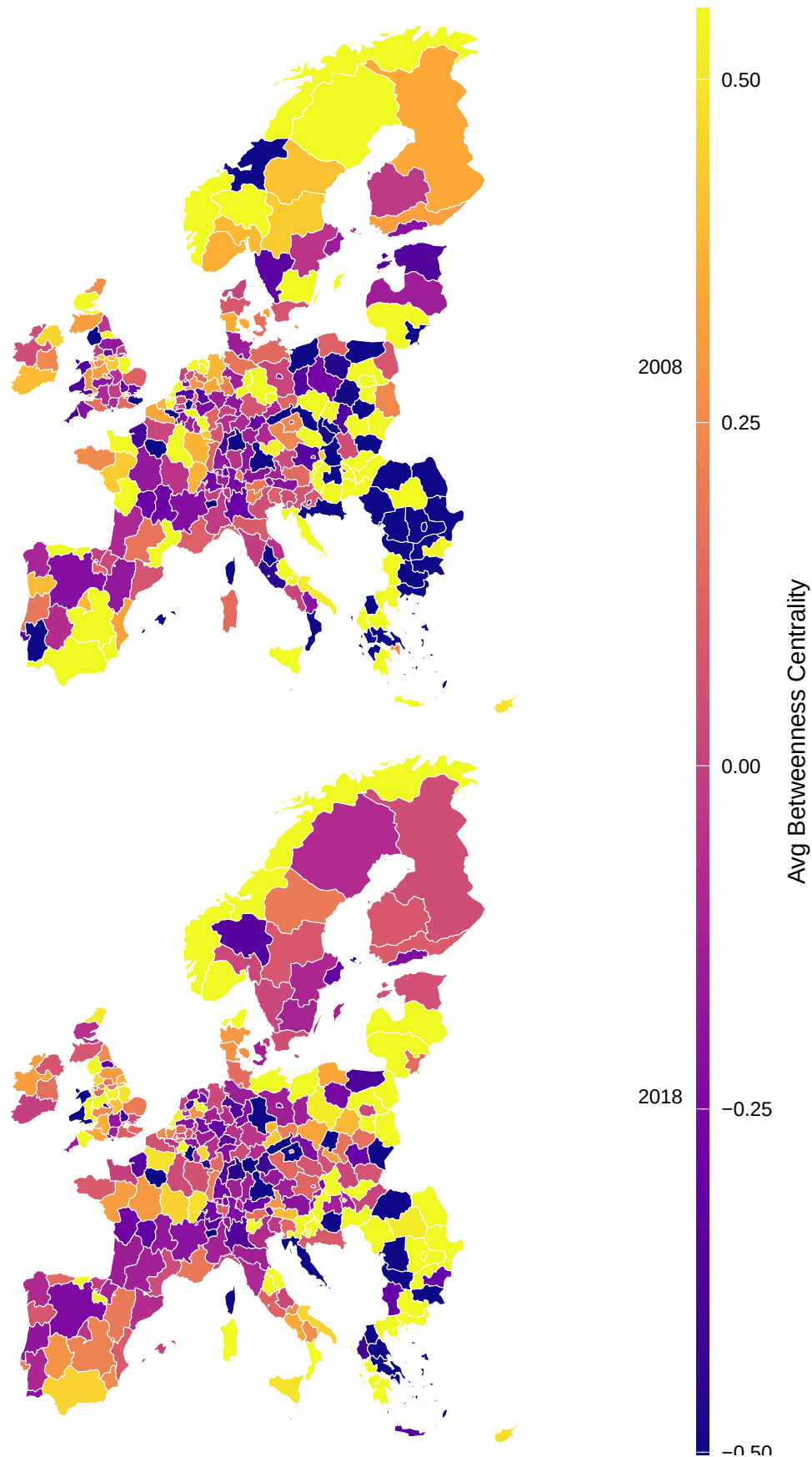
This specification is one part of our identification strategy where region fixed effects absorb all time-invariant regional characteristics that could confound the relationship between the regional portfolio characteristics and regional potential. Additionally, year fixed effects absorb all common temporal shocks. Thus, we identify within-group variations that change over time. Additionally, this specification should also address at least partially potential issues with endogeneity.

The idea here is to understand what factors explain regional technological potential when we account for the regional and temporal common variations. Does coherence moderate the effect of network position on potential? Is this moderation identical for KETs (key enabling technologies, proxied by eigenvector) and GPTs (general purpose;bridging technologies, proxied by betweenness)?

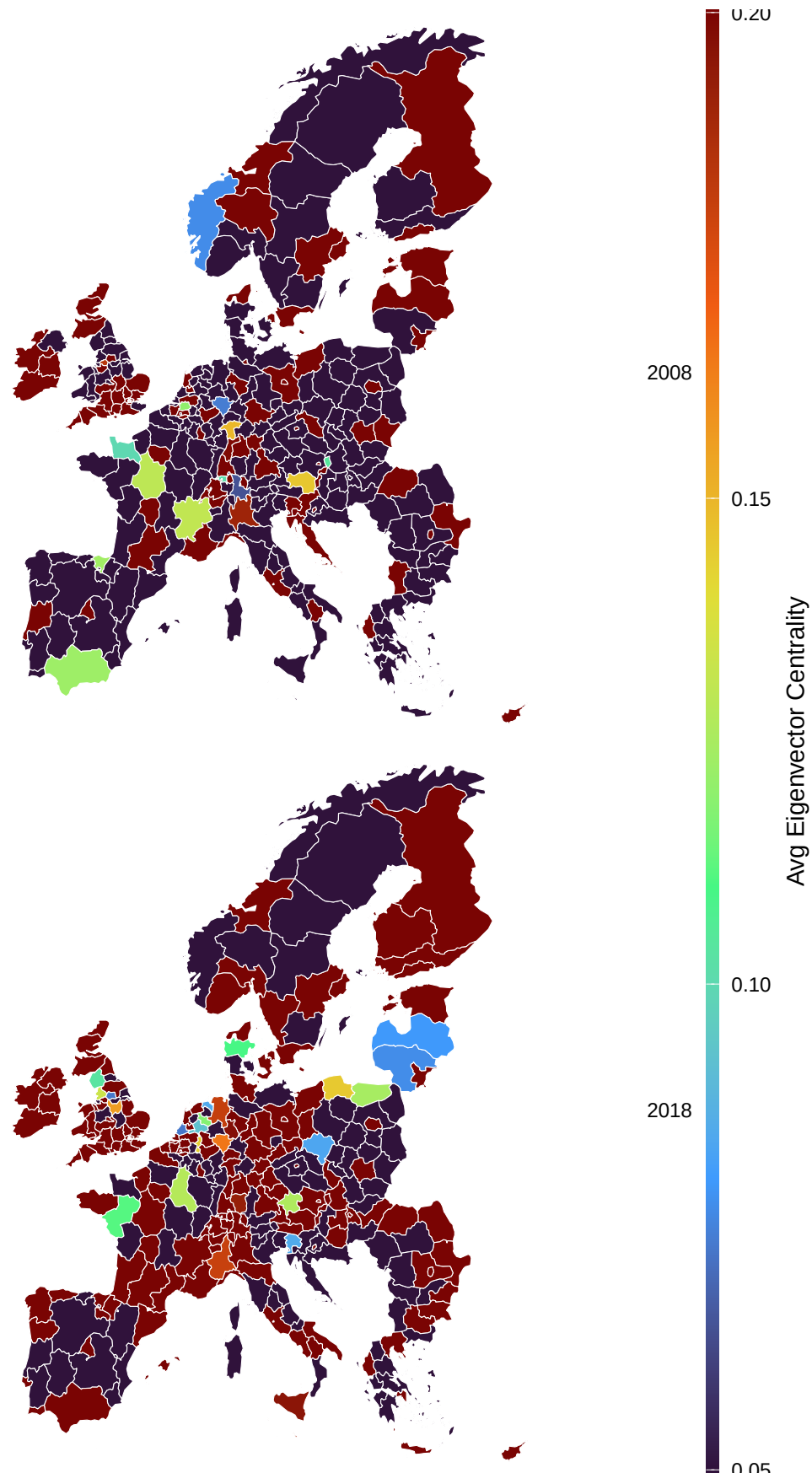
Regardless, we estimate a multitude of regression models and progressively tighten their complexity and specifications. The idea here is to compare the consistency of the contribution of the covariates across these specifications.

The baseline model includes coherence, distance, patent count, and complexity controls. The network model adds betweenness and eigenvector centrality testing whether network position consistently predicts potential. The interaction model add the coherence x betweenness and coherence x eigenvector terms. We also add a parallel specification for robustness where we replaces the coherence moderation with distance.

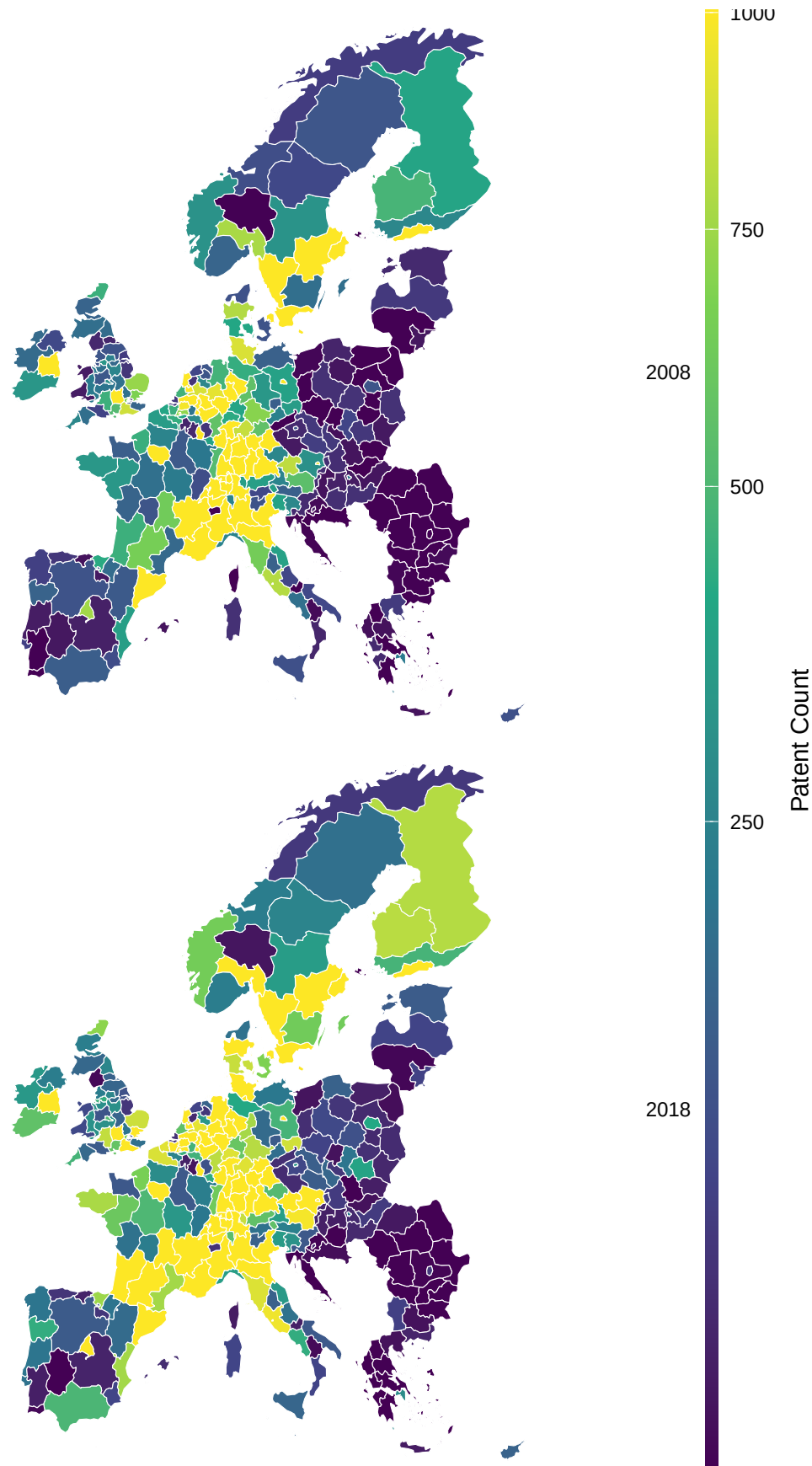
Of course with the inclusion of national ecosystem covariates, this identification loses some of its merits since we cannot account for cross-regional country level differences. This is because region fixed effect will absorb country level variations and because country fixed effect will leave out common regional variation. This motivates the adoption of a hierarchical model when we include the country level covariates. However, we also retain this (fixed effects) approach as robustness check for model 2.



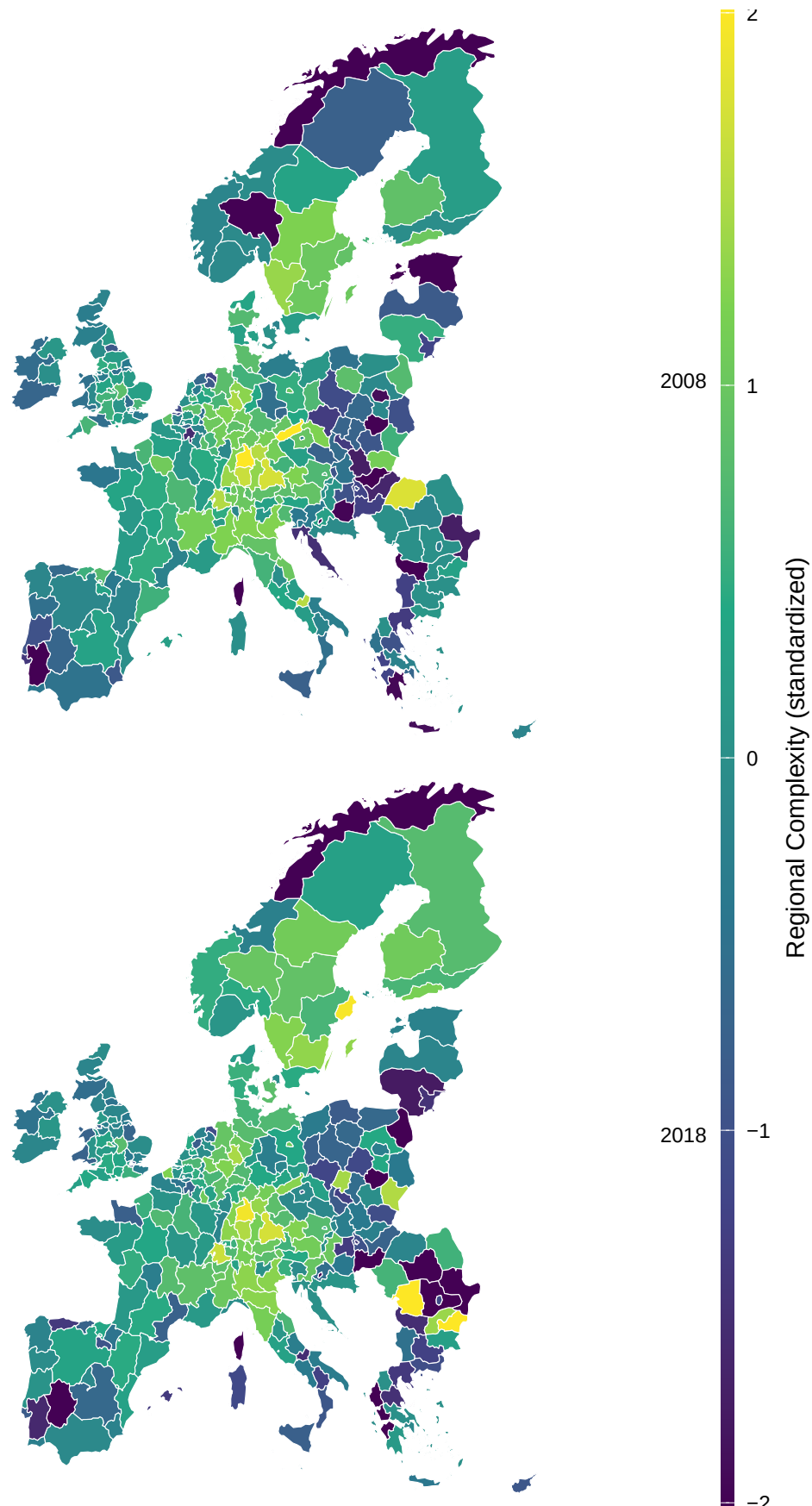
(a) This map displays the average betweenness centrality of regional technology portfolios, measuring the extent to which regions specialize in bridge technologies—those that sit on many shortest paths between other technologies in the FITS network and connect otherwise distant capability domains. High betweenness regions (bright yellow) have portfolios rich in general-purpose bridging technologies that mediate between specialized application areas, while low



(a) This map shows the average eigenvector centrality of technologies in which each region specializes, capturing the extent to which regional portfolios occupy authority-like positions in the FITS network—technologies that are well-connected to other well-connected technologies at the structural core of the capability space.⁸⁹ High eigenvector regions (red/pink) tend to be advanced innovation centers whose portfolios concentrate in central, highly interdependent



(a) This map visualizes total patent activity across European regions, representing the volume of technological production and innovation output. The square root transformation compresses the scale to make the distribution more interpretable, as raw patent counts exhibit extreme skew (a few metropolitan regions produce thousands of patents while many rural regions produce tens). High-activity regions (bright yellow) are concentrated in major innovation hubs—



- (a) This map displays the geographic distribution of regional complexity. This is a measure of the sophistication and diversity of regional technological portfolios-across 345 European NUTS2 regions. Values are standardized and capped at 2 standard deviations to highlight variation within the core distribution. The pattern reveals a clear core-periphery divide between European regions

4.6 Model 2: Hierarchical models

As we cross different literature domains in this work we observed that there's a clear and urgent gap that was not properly addressed that's the lack of time varying regional entrepreneurial ecosystem datasets. The Regional Entrepreneurship and Development Index (REDI) (Szerb et al. 2013; Argiles et al. 2014) and the work of Leendertse, Schrijvers, and Stam (2022) are a step in that direction but the challenge of time variation still persist. This means that to be consistent with the evolutionary scope with which we framed this work we are forced to accommodate the national covariates.

Although the fixed effect design we described in the previous section is relatively adequate for isolating the within-region mechanisms it still does not draw the full story as it discards the between-region variations. Yet between-region variation in our context is theoretically meaningful since we are interested in understanding why some regions are systematically characterised by higher potential compared to others. This interest then compliments our focus in the previous specification where the focal attention is on why potential fluctuates.

Thus to properly test H3, we need a specification that preserves national and regional variation. To this end we include random effects by modelling nested region within country intercepts that are drawn from a distribution rather than absorbing them as free parameters. For that we estimate linear mixed models with the following nesting structure:

$$p_{r,t,y}^* = \mathbf{X}_{r,t,y}\beta + \zeta_t + u_c + u_{r|c} + \varepsilon_{r,t,y}$$

where $u_c \sim \mathcal{N}(0, \sigma_c^2)$ is a country-level random intercept and $u_{r|c} \sim \mathcal{N}(0, \sigma_{r|c}^2)$ is a region-within-country random intercept. Year is included as a dummy (ζ_t) to absorb common temporal dynamics.

With these models we will be able to compare the contribution of the covariates to the contribution to the nested structure itself. This is to say that if the nested structure is highly predictive of the variation in regional technological potential then the outcome is primarily driven by the location rather than the covariates themselves. This justifies our

next layer of analysis: if geography matters then we need to assess the spatial dependence of the factors associated with regional potential. To that end we explore in the next section in more details how we will assess this dependence.

4.7 Model 3: Spatial panel models

Till now, we discussed two models that investigate variation in regional technological potential at different levels of grouping specification. The fixed effect models are designed to assess the within-region variation, and the hierarchical model is designed to simultaneously account for the national ecosystem factors while also serving as a robustness check for the network covariates and their interaction with coherence. As we justified earlier, if the nested structure of the data drives much of the explanatory power of the model, it could be an indication that spatial dependence and/or geographical proximity is what matters here. Additionally, the literature gives reason to expect real interdependence especially in the context of European regions where administrative borders are simply a construct and not a barrier for firms and entrepreneurs (Boschma and Iammarino 2009; Komlósi et al. 2025). The spatial models in this analysis test this formally.

To proceed with the spatial analysis, we construct spatial weights using k-nearest neighbours ($k = 5$) based on NUTS2 geographic centroids. We also perform row-standardisation for each focal region i and its candidate neighbour j , such that $\sum_j w_{ij} = 1$, where:

$$w_{ij} = \begin{cases} 1/k & \text{if } j \in \text{knn}_5(i) \\ 0 & \text{otherwise} \end{cases}$$

The preliminary diagnostics are summarised in Table 4.9 show that Moran's I statistics for our main outcome reveal a persistent positive spatial autocorrelation across all years in the panel (I around 0.44-0.51, $p < 0.001$). Furthermore, we apply the Baltagi-Song-Koh Lagrange Multiplier tests designed for panel data to discriminate between competing spatial mechanisms (Baltagi, Song, and Koh 2003). The robust LM-lag statistic is large and

Table 4.9: Spatial Autocorrelation Diagnostics

Test	Statistic	p-value	Interpretation
Moran's I	0.44-0.51	< 0.001	Strong positive spatial autocorrelation in regional potential
LM-lag (standard)	616.2	< 0.001	Spatial lag process significant before conditioning
LM-error (standard)	271.6	< 0.001	Spatial error process significant before conditioning
Robust LM-lag	350.4	< 0.001	Spatial lag dominates after conditioning on error
Robust LM-error	4.1	0.042	Marginal - spatial error secondary to lag process

Note: Moran I is tested for each year, the table shows tghe range of its statistic between 2008 and 2018. Rbust LM tests condition each null on the presence of the other type of dependence. The diagnostic pattern - highly significant robust LM-lag, marginally significant robust LM-error - favours the SAR specification.

highly significant (350.4, $p < 0.001$), while the robust LM-error statistic is only marginally significant (4.1, $p = 0.042$) after conditioning on the spatial lag. Thus, these diagnostic results indicate that spatial dependence operates primarily through neighbours' outcomes rather than through spatially correlated errors. In light of this, the SAR model is the adequate specification, while we use the SEM model as a robustness check.

4.7.1 Spatial Autoregressive (SAR) panel - primary specification

We estimate the Spatial Autoregressive (SAR) model with two-way (region and year) demeaning. In this model, the spatial autoregressive parameter ρ tests H4 where a positive and significant ρ indicates that regions tend to exhibit higher potential when neighbouring regions also display high potential.

$$P_{r,y} = \rho \mathbf{W} P_{r,y} + \mathbf{X}_{r,y} \beta + \alpha_r + \zeta_y + \varepsilon_{r,y}$$

The covariates matrix $\mathbf{X}_{r,y}$ includes regional patent count, regional complexity, average coherence, average betweenness, and average eigenvector, along with their interactions.

However, we need to note here that at the dyadic level, betweenness and eigenvector measures are technology level metrics. They capture the characteristics of the network position of a given technology whether it's a key enabling technology or general purpose technology. In this analysis however, average eigenvector and average betweenness measure how much a region is endowed with key enabling technology or general purpose technology in their portfolio and not the characteristics of the technologies themselves. That is to say in this level of analysis our focus is on the aggregate characteristics of the regional portfolio. For our work, this interpretation complements the specification of the dyadic models and reveal more context to the dynamics we're studying.

4.7.2 Spatial Error Model (SEM) - robustness

Alternatively, we estimate the Spatial Error Model (SEM) as a robustness check. In this model we assume that the unobservable determinant of regional potential are spatially autocorrelated. This is to say that the spatial dependence of regional potential is driven by unobserved spatial correlation rather than observed variations.

$$P_{r,y} = \mathbf{X}_{r,y}\beta + \alpha_r + u_{r,y}, \quad u_{r,y} = \delta \mathbf{W}u_{r,y} + \varepsilon_{r,y}$$

In here, δ is our benchmark such that its sign, magnitude and significance compared to ρ from the SAR model allows to differentiate the spatial dynamics that drives regional potential.

4.8 Summary

This chapter presents the empirical framework we adopt to investigate what shapes regional technological diversification feasibility. We build directly on the methodological contribution of the previous chapter and translate our research questions into a structured empirical design. We organise the analysis around two levels. The first is dyadic, meaning each

4 Empirical design

observation is a combination of a technology, and a region. At this level we ask how a technology’s position in the FITS network predicts regional technological potential, and whether this relationship changes depending on how well the region’s industrial portfolio is aligned with that technology. We then extend this by adding national entrepreneurial ecosystem characteristics to test whether institutional and market conditions at the national level contribute to regional potential beyond what the portfolio mechanisms already explain. The second level is spatial. Here we aggregate our variables to the region-year level and ask two questions. First, whether regions with high potential tend to neighbour other high-potential regions. Second, whether the overall composition of a region’s portfolio in terms of bridging and enabling technologies predicts its average potential, and whether this depends on the region’s overall coherence level. To address these questions we use three models. Model 1 is a fixed effects model that captures within-region variation. Model 2 is a hierarchical random effects model that accounts for regions being nested within countries, which allows us to properly identify the contribution of national ecosystem characteristics. Model 3 is a spatial autoregressive model that tests for and decomposes spatial interdependence. Before these three models we also run a benchmark comparison between regional technological potential and relatedness density as predictors of observed technology entry, to validate that our machine learning framework adds predictive value over the conventional approach.

5 Results

In the previous chapter we discussed thoroughly the different areas of interest to this work. We relied on a clear framing that justifies assessing the regional diversification potential in a more contextualised and granular approach. To do that we started from the conceptual constructs of relatedness and the subsequent technology space and introduced our machine learning approach as context-rich alternative. We also introduced our coherence metric that capitalises on our approach. With that we've drawn a clear conceptual framework on how all these ideas should interact and inform on our main outcome. Essentially we propose two levels of analysis each with a clear objective but with complementary interpretation. These two analysis are operationalised via three models which we described their rational, specifications, and how they meet our hypothesis in the previous chapter. In this chapter we present the results of these models and contrast their interpretation to the hypothesis and the questions this work aims to address.

In here we will decompose our finding for each of the models we specified. We will first explore the results of model 0 (Table 5.1), which treats the benchmark hypothesis and compare the empirical contribution of regional potential and relatedness to regional entry, following a large body of literature we already discussed in the literature review. Then we will present the results of model 1 (Table 5.2) which explores the hypothesis on the empirical contribution of technologies' roles in the network structure from the FITS network. This model also explores the interaction between these roles and coherence to investigate whether the relationship is universal or contingent on the alignment in industrial embeddedness between the region and the technology. Model 1 also extends this exploration by including the national ecosystem covariates (Table 5.3), this to test the stability of the network

metric and their interaction as predictive factors, and also as a robustness check for their contribution compared to model 2. Essentially model 1 fits our variables to regional technological potential and assess the within-group variation. But because the national ecosystem covariates are not at a regional level, we also want to assess their contribution when we account for both within-group and between-group variation since regions are administrative artefacts that share the same national ecosystem characteristics. This nested structure motivate our model 2 presented in Table 5.4. As we mentioned in the previous chapter, if the nested structure of model 2 drives most of the variation of the outcome, it could be an indication that this variation is mainly driven by spatial dependence. This is the case as we will explain later in this chapter, thus motivates our model 3. The idea here is to understand the spatial dynamics that drives regional outcomes, the test results presented in the previous chapter indicate a strong dependence on neighbouring region outcome. This justified the inclusion of the SAR model (Table 5.5), albeit we also present the SEM model as a robustness check (Table 5.8)

5.1 Entry as a benchmark

In Table 5.1 we benchmark how much conventional relatedness density and potential (pred11) explain regional entry to a new technology. In the pooled specifications, both measures strongly predict entry: relatedness density is positive and precisely estimated (0.055), while Potential is larger (0.081). This is consistent with the interpretation that both capture diversification opportunities, however, Potential here is closer to a feasibility metric that's learned from portfolio configurations than relatedness. Once we move to the fixed-effects designs and absorb region and year fixed effects, interpretation shifts to within-region changes over time, net of common year shocks. In this setting, relatedness density remains strongly positive (e.g., 0.089 in the density-only fixed effects model), and Potential also remains strongly positive (0.089 in the Potential-only fixed effects model). The coefficient of relatedness increase significantly more than potential (which remains relatively at a stable value) when we account for within-group variation in which case both

5 Results

relatedness and potential contribute similarly to the probability of entry. Albeit, potential seems to improve the overall and within-group fit more than relatedness (within- R^2 rises from 0.014 in the density-only fixed effects model to 0.062 in the Potential-only fixed effects model). Moreover, when we include both metrics together, the dominance of potential influence becomes clear, the coefficient of relatedness density decrease from 0.089 to 0.028, whereas the coefficient of potential remains stable at 0.086. In other words, Potential appears to be relatively the dominant predictor in this horse race, while density retains residual explanatory content that may reflect a simpler proximity component that survives beyond the non-linear feasibility signal embedded in Potential. Albeit, the evidence here is not conclusive since both potential and relatedness density display a similar influence on entry when treated separately, the dominance of potential when we include relatedness density indicates that it subsumes much of the linear relatedness signal since it captures a richer context and patterns by construction compared to the linear patterns captured in the relatedness metric.

This benchmark motivates the subsequent mechanism analysis: rather than treating relatedness as a sufficient summary statistic, we examine how portfolio alignment, network position, and accessibility structure feasibility. This move is consistent with the broader methodological critique that pairwise relatedness can miss higher-order portfolio information; for example, Tacchella et al. (2023) argues that shifting from “two-product correlations” to information that captures richer configurations can improve predictive performance. This benchmark motivates our subsequent mechanism analysis: having established that Potential is relatively more informative than relatedness density in predicting entry, we now decompose Potential itself to examine how portfolio alignment (coherence), network position (betweenness, eigenvector), and national ecosystem structure feasibility.

Table 5.1: Regional Potential vs Relatedness density – Entry(year + 1, LPM with fixed effects)

	baseline1	baseline2	full1	full2	full3
(Intercept)	0.069*** (0.002)	0.070*** (0.001)			
patent_n	0.008 (0.007)	-0.007** (0.002)	0.017* (0.008)	0.004 (0.003)	0.004 (0.003)
relatedness_density	0.055*** (0.002)		0.089*** (0.006)		0.028*** (0.002)
reg_comp	-0.003*** (0.001)	-0.003*** (0.001)	-0.001*** (0.000)	-0.000 (0.000)	-0.001** (0.000)
tech_comp	-0.015*** (0.002)	-0.005*** (0.001)	-0.013*** (0.002)	-0.004*** (0.001)	-0.003*** (0.001)
pred11		0.081*** (0.002)		0.089*** (0.002)	0.086*** (0.002)
Num.Obs.	1571364	1571364	1571364	1571364	1571364
R2	0.043	0.086	0.046	0.092	0.093
R2 Adj.	0.043	0.086	0.046	0.092	0.093
R2 Within			0.014	0.062	0.062
R2 Within Adj.			0.014	0.062	0.062
AIC	-36976.1	-109812.6	-41129.6	-119089.6	-120205.8
BIC	-36914.8	-109751.3	-37375.8	-115335.8	-116439.6
RMSE	0.24	0.23	0.24	0.23	0.23
Std.Errors	by: region & tech	by: region & tech	by: region & tech	by: region & tech	by: region & tech
FE: region			X	X	X
FE: year			X	X	X

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

5.2 Network effects and coherence moderation

As we mentioned in the beginning of this chapter, one of our main tests in our work is regarding the relationship between the technologies' network roles and regional technological potential and whether this relationship is universal or dependent on the alignment of industrial structure between regions' portfolios and a given technology. Table 5.2 portrays one side of the picture that we're testing here which focuses on the within-group variation.

The first observation here is that all the coefficients maintain the same sign across all specification. The coefficients of betweenness, eigenvector, coherence, and the interactions are the most notably stable in sign and significance with a slight fluctuation in size of the effect.

As a start, column 5 which represent our full specification shows that coherence—the alignment between a technology's FITS embeddedness and the region's portfolio—is the strongest predictor of potential (0.706 at mean betweenness and mean eigenvector)¹. Although regional complexity exhibits a similar predictive strength in columns 2 and 3, its effect is completely absorbed by the fixed effects.

¹Because coherence interacts with both betweenness and eigenvector in the full specification, the reported coefficient is its marginal effect when both interacted variables are at their sample means. The conditional marginal effect of coherence varies with these variables but remains positive and substantive throughout the empirical distribution, as we discuss further below.

Table 5.2: Network Interactions – Regional Potential (with fixed effects)

	baseline	network	interaction_a	interaction_a1	interaction_b	interaction_b1	interaction_c	interaction_c1
(Intercept)	-2.528*** (0.070)		-2.462*** (0.074)	-2.476*** (0.074)			-2.516*** (0.071)	
patent_n	0.211*** (0.054)	0.088*** (0.023)	0.213*** (0.053)	0.215*** (0.054)	0.088*** (0.023)	0.091*** (0.024)	0.212*** (0.053)	0.088*** (0.023)
coherence	1.026*** (0.038)	0.779*** (0.039)	0.898*** (0.039)	0.909*** (0.040)	0.706*** (0.037)	0.723*** (0.038)	0.944*** (0.039)	0.780*** (0.039)
rawdist	-0.028 (0.028)	-0.039 (0.028)	-0.011 (0.027)		-0.029 (0.027)		0.036 (0.046)	0.009 (0.045)
reg_comp	0.741*** (0.065)	0.013 (0.018)	0.746*** (0.066)	0.741*** (0.065)	0.010 (0.017)	0.010 (0.017)	0.745*** (0.065)	0.013 (0.018)
tech_comp	-0.213*** (0.035)	-0.256*** (0.040)	-0.212*** (0.035)	-0.198*** (0.034)	-0.254*** (0.039)	-0.238*** (0.038)	-0.212*** (0.035)	-0.255*** (0.040)
betweenness		0.131*** (0.023)	0.249*** (0.039)	0.266*** (0.040)	0.388*** (0.037)	0.407*** (0.038)	0.101*** (0.028)	0.147*** (0.029)
eigenvector		-0.237*** (0.030)	-0.206*** (0.028)	-0.201*** (0.029)	-0.221*** (0.027)	-0.216*** (0.028)	-0.209*** (0.030)	-0.233*** (0.030)
betweenness × coherence			-0.184*** (0.032)	-0.191*** (0.033)	-0.284*** (0.030)	-0.292*** (0.031)		
coherence × eigenvector			-0.015 (0.030)	-0.001 (0.031)	-0.001 (0.028)	0.013 (0.030)		
betweenness × rawdist							0.067 (0.081)	0.057 (0.078)
rawdist × eigenvector							-0.081*** (0.023)	-0.080*** (0.023)
Num.Obs.	1906137	1906137	1906137	1922371	1906137	1922371	1906137	1906137
R2	0.354	0.659	0.366	0.367	0.665	0.658	0.364	0.659
R2 Adj.	0.354	0.659	0.366	0.367	0.665	0.658	0.364	0.659
R2 Within		0.356			0.368	0.367		0.357
R2 Within Adj.		0.356			0.368	0.367		0.357
AIC	7682716.6	6466221.8	7645396.3	7738488.6	6429417.5	6556534.8	7650873.9	6461334.5
BIC	7682791.4	6470084.6	7645520.9	7738600.8	6433305.2	6560412.7	7650998.5	6465222.2
RMSE	1.82	1.32	1.80	1.81	1.31	1.33	1.80	1.32
Std.Errors	by: region & tech	by: region & tech	by: region & tech	by: region & tech	by: region & tech	by: region & tech	by: region & tech	by: region & tech
FE: year		X			X	X		X
FE: region		X			X	X		X

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

5 Results

Similarly, betweenness centrality, our network measure that captures bridging positions (as a proxy classification to general purpose technology role), exhibits consistent and substantive positive effects at mean coherence (0.388, column 5). The coherence $\times$ betweenness interaction is significantly negative (-0.284), but its interpretation requires evaluating the conditional marginal effect of betweenness across the coherence distribution rather than reading the interaction term in isolation². The marginal effect of betweenness when coherence is 0 is 0.388: at one standard deviation below mean coherence this yields $+0.672$, and at one standard deviation above mean coherence the marginal effect is $+0.104$. Therefore, bridging technologies associate with a higher likelihood of a region having a higher potential across the full coherence distribution – the interaction moderates the magnitude of that benefit, not its direction. The sign would only reverse at $c^* \approx +1.37$ SD, beyond the realistic empirical range. What the negative interaction establishes is a coherence-conditioned gradient: regions with weaker industrial alignment extract substantially larger returns from bridging technologies, while high-coherence regions still benefit but more modestly, consistent with their feasible opportunity set being more concentrated in already-aligned portions of the technology space³

On the other side of this logic, the eigenvector centrality pattern presents a another story. Technologies occupying authority/prestige positions (as a proxy to key enabling technology roles) consistently inhibit potential (marginal effect -0.221 at mean coherence, Column 5). Its interaction with coherence (coherence $\times$ eigenvector) shows consistent non-significance across all specifications, which means the conditional marginal effect of eigenvector remains

²The marginal effect of a variable on the outcome, defined as its partial derivative $ME(\tilde{x} | c) = \partial y / \partial \tilde{x} = \beta_{\tilde{x}} + \beta_{\tilde{x} \cdot \tilde{z}} \cdot c$, is not fixed when two z-standardised variables interact — it depends on the level c at which the conditioning variable $\tilde{z}$ is evaluated. Since all variables are z-standardised, $c = 0$ is the sample mean, $c = +1$ is one standard deviation above, and $c = -1$ one standard deviation below. The effect reverses sign only if the threshold $c^* = -\beta_{\tilde{x}} / \beta_{\tilde{x} \cdot \tilde{z}}$ falls within the empirical range of $\tilde{z}$. To illustrate: the marginal effect of betweenness conditioned on coherence at one standard deviation above the mean is $ME(\text{betweenness} | c = +1) = 0.388 + (-0.284) \cdot (+1) = +0.104$. The same logic yields all conditional values reported in this section.

³This is further supported by the coherence $\times$ eigenvector interaction, which remains consistently non-significant. Enabling technologies are a strong and negative predictor of potential (-0.221 , column 5), but this inhibiting effect does not vary with coherence – the marginal effect of eigenvector is approximately -0.221 regardless of the region’s industrial alignment. The null interaction therefore does not mean enabling technologies are irrelevant to potential; it means their structural difficulty is coherence-invariant. We will discuss this further as we progress in this chapter.

at -0.22 regardless of the region’s industrial alignment. These technologies are therefore systematically difficult to diversify into, and coherence does not condition that difficulty. C. Pinheiro (2025) (among others) argues in this context in favour of these results, considering that proximity alone is not a sufficient predictor of diversification. Antonietti and Montresor (2021) similarly argues that enabling technologies support more exploratory trajectories “only when” they are already embedded and used within the local productive system – a condition that candidate technologies, by definition, do not yet meet at the margin. We will return to this distinction between entry difficulty and portfolio-level enabling advantage as we progress.

Our null moderation for eigenvector centrality is consistent with the idea that adopting enabling technologies is challenging and that the conditions and context for this adoption may not be fully captured by our coherence measure. For that reason we added the interaction between the network measures and the distance covariate as an robustness check. The interaction between distance and eigenvector is negative and relatively robust (-0.080 in columns 7 and 8). This indicates that authority technologies become particularly difficult when they are also topologically remote from any regional foothold. This mechanism is close to the account in Zhu, Guo, and He (2021), the authors show that diversification is more feasible through stronger, closer links. Our results contextualise this further, as this distance seems more relevant for technologies with enabling roles rather than bridging roles. This contrast can be seen in the insignificant betweenness $\times$ rawdist interaction. This (limited evidence) aligns with the intuition that general purpose technologies by construct should be accessible by construction as they bridge different areas of the network.

5.3 National context and robustness

In Table 5.3 (and later on in Table 5.4) we extend our framework by incorporating covariates on the national entrepreneurial ecosystem characteristics from the Global Entrepreneurship Index. The idea here is to understand whether these covariates contribute to the variation in regional potential along side the network metrics (and their interactions). However,

5 Results

this specification for the national covariates serves as a robustness check for the models in Table 5.4 and `tbl-potential-fe-ols`, model 2 remains our preferred setting for these covariates.

First, we can see in Table 5.3 that across the different specification the inclusion of the entrepreneurial ecosystem covariates does not change the sign, significance nor the size of the effects of the other covariates we explored in Table 5.2. In fact we see a marginal improvements in the goodness of fit (within- R^2 rises from 0.368 to 0.371 and overall- R^2 from 0.665 to 0.668). Although small, this improvement shows that the inclusion of the national entrepreneurial ecosystem covariates is still contributive regardless. This stability supports a cautious conclusion that the coherence and network-position mechanisms are not simply proxies for national institutional variation.

The contribution of these covariates, however, doesn't show the same consistency across our specification with the exception of "x13_internationalization" and "x9_competition". This is especially notable when we include the fixed effects. We already briefly discussed in the previous chapter why this could be the case here, and we will return to this point in the discussion chapter.

Regardless, Competition is consistently positive, while internationalization is consistently negative (0.088 and -0.082 respectively in Column 4). Other GEI components show heterogeneous patterns across specifications.

This combination of stable dyadic mechanisms and non-negligible national effects is consistent with the caveat mentioned in C. Pinheiro (2025) (also in Boschma (2015)), which cautions against treating relatedness metrics as sufficient guides to diversification arguing that institutional and policy context can shape which opportunities are realized. This heterogeneity (regardless of the consistency of competition and internationalization effects), together with the stable core mechanisms that remain consistent, constitute one of our main empirical contribution in this work. Consequently and in alignment with these observations in the literature these results provides partial evidence for H4. To explore this further model 2 provides an alternative specification that's more adequate for the hierarchical structure of the data.

Table 5.3: Network Interactions and National Ecosystem – Regional Potential (with fixed effects)

	h0	h01	h1	h2	h3	h3
(Intercept)	-2.493*** (0.060)		-2.406*** (0.063)		-2.480*** (0.061)	
patent_n	0.181*** (0.044)	0.090*** (0.023)	0.181*** (0.045)	0.090*** (0.024)	0.181*** (0.044)	0.089*** (0.023)
coherence	0.884*** (0.039)	0.784*** (0.040)	0.826*** (0.038)	0.709*** (0.038)	0.885*** (0.039)	0.785*** (0.040)
rawdist	-0.024 (0.030)	-0.037 (0.030)	-0.015 (0.029)	-0.027 (0.029)	0.031 (0.046)	0.010 (0.045)
reg_comp	0.519*** (0.063)	0.015 (0.021)	0.518*** (0.063)	0.011 (0.020)	0.519*** (0.063)	0.015 (0.021)
tech_comp	-0.215*** (0.039)	-0.243*** (0.042)	-0.214*** (0.039)	-0.241*** (0.041)	-0.214*** (0.039)	-0.241*** (0.042)
betweenness	0.104*** (0.023)	0.133*** (0.024)	0.316*** (0.038)	0.398*** (0.038)	0.121*** (0.029)	0.147*** (0.030)
eigenvector	-0.224*** (0.031)	-0.240*** (0.031)	-0.211*** (0.028)	-0.222*** (0.028)	-0.220*** (0.031)	-0.235*** (0.031)
x8_human_capital	-0.132* (0.054)	0.044* (0.022)	-0.134* (0.054)	0.044* (0.022)	-0.132* (0.054)	0.044* (0.022)
x9_competition	0.568*** (0.082)	0.087* (0.034)	0.573*** (0.082)	0.088** (0.034)	0.568*** (0.082)	0.088* (0.034)
x10_product_innovation	0.198*** (0.056)	0.016 (0.018)	0.200*** (0.056)	0.017 (0.018)	0.198*** (0.056)	0.016 (0.018)
x11_process_innovation	0.235*** (0.063)	-0.020 (0.028)	0.236*** (0.063)	-0.022 (0.028)	0.235*** (0.063)	-0.020 (0.028)
x12_high_growth	-0.110+ (0.056)	0.008 (0.020)	-0.109+ (0.056)	0.009 (0.020)	-0.110* (0.056)	0.008 (0.020)
x13_internationalization	-0.088 (0.054)	-0.080** (0.029)	-0.090 (0.054)	-0.082** (0.029)	-0.088 (0.054)	-0.080** (0.029)
x14_risk_capital	0.109* (0.045)	-0.035 (0.023)	0.109* (0.046)	-0.035 (0.023)	0.109* (0.045)	-0.035 (0.023)
coherence × betweenness			-0.234*** (0.031)	-0.292*** (0.031)		
coherence × eigenvector			-0.006 (0.030)	0.002 (0.030)		
betweenness × rawdist					0.063 (0.079)	0.050 (0.078)
rawdist × eigenvector					-0.084*** (0.024)	-0.082*** (0.024)
Num.Obs.	1506300	1506300	1506300	1506300	1506300	1506300
R2	0.474	0.661	0.478	0.668	0.475	0.662
R2 Adj.	0.474	0.661	0.478	0.668	0.475	0.662
R2 Within		0.358		0.371		0.359
R2 Within Adj.		0.358		0.371		0.359
AIC	5748605.3	5088428.5	5735734.5	5057169.8	5745677.7	5084204.9
BIC	5748788.7	5092279.4	5735942.3	5061045.2	5745885.6	5088080.3
RMSE	1.63	1.31	1.62	1.30	1.63	1.31
Std.Errors	by: region & tech	by: region & tech	by: region & tech	by: region & tech	by: region & tech	by: region & tech
FE: region		X		X		X
FE: year		X		X		X

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

5.4 Hierarchical structures and spatial implications

Similar to the extended fixed effects specification in Table 5.3, model 2 accounts for the same covariates but considers the nested structure of the data. Table 5.4 employs random effects models where the regions' random intercepts are drawn from the same distribution for each country.

Crucially, the core coefficients remain stable in the hierarchical models as well. In the random effects specification comparable to the fixed effects benchmark, coherence remains positive (around 0.709 at mean betweenness and eigenvector), and the coherence-conditioned gradient for betweenness is preserved: with a main effect of around 0.398 at mean coherence and an interaction of -0.292 , the marginal effect of betweenness ranges from approximately $+0.690$ at one standard deviation below mean coherence to $+0.106$ at one standard deviation above, remaining positive throughout the empirical coherence distribution. The consistency of these variables across fixed effects and random effects specifications confirms that the coherence condition is not an artifact of the within-group estimation strategy but a robust structural feature of the data. Eigenvector centrality remains negative (around -0.222) and its interaction with coherence also remains non-significant. This further supports our previous assessment, the enabling technology inhibiting effect holds across both within-group and cross-regional specifications regardless of the coherence level.

However and more importantly, in this model the national ecosystem covariates are uniformly stable and consistent in sign, significance and effect size. In this model we can see that variation in human capital, competition, product innovation, and high growth pillars are associated at 4%, 9%, 1.6% and 0.8% respectively with the outcome. This is in contrast to process innovation, internationalisation, and risk capital pillars whose variations are negatively associated with the outcome at 2%, 8%, and 3.5% respectively.

These results directly address H4, and relatively confirm the intuition we formalised in this work and the results in the previous model. National entrepreneurial ecosystem characteristics matters for regional potential and explain a significant proportion of the variation (our full specification in column 5 shows the highest marginal and conditional

R^2 , 0.201 and 0.668 respectively). In fact, this can also be observed in the high intra-class correlation ($ICC = 0.60$). This shows that 60% of variance in regional potential is between-group rather than within-group (stable across the different specifications). The variance components show that country-level heterogeneity (standard deviation around 1.12-1.13) exceeds region-within-country heterogeneity (approximately 1.02–1.03). These decompositions inform us that, while the fixed effects estimates document robust within-region mechanisms, a substantial portion of the feasibility landscape is structured by relatively persistent differences across countries and regions.

The results of this model also help justify the spatial models that follow. The large ICC and the variance decompositions indicates that regional potential is strongly structured across places. This confirms our intuition in the previous section as we introduced the rationale behind model 2. The nested structure of the dyadic context matters almost as much as the dyadic components themselves. This is to say that the geography of the region (where it is, in which country) is extremely informative of its technological potential. This could justify the idea that feasibility is not independent across space but may be shaped by neighbouring structures. Consequently, the distinction between within-region fixed effects patterns and between-place random effects patterns suggests that part of the remaining heterogeneity may be geographically patterned. This is our main motivation for an explicit test for spatial dependence which is the objective of model 3 that we present its results in the next section.

Table 5.4: Hierarchical Random Effects Model – Regional Potential

	inter_a	inter_a1	inter_b	inter_b1	h_a	h_b	h_b1
(Intercept)	-2.577*** (0.088)	-2.667*** (0.087)	-2.573*** (0.088)	-2.666*** (0.088)	-2.817*** (0.227)	-2.711*** (0.230)	-2.807*** (0.227)
patent_n	0.109*** (0.001)	0.108*** (0.001)	0.088*** (0.001)	0.088*** (0.001)	0.090*** (0.001)	0.090*** (0.001)	0.089*** (0.001)
rawdist	-0.032*** (0.001)	0.008*** (0.002)	-0.029*** (0.001)	0.009*** (0.002)	-0.037*** (0.001)	-0.027*** (0.001)	0.010*** (0.002)
betweenness	0.427*** (0.002)	0.200*** (0.001)	0.388*** (0.002)	0.147*** (0.001)	0.133*** (0.001)	0.398*** (0.002)	0.147*** (0.001)
coherence	0.729*** (0.001)	0.802*** (0.001)	0.706*** (0.001)	0.780*** (0.001)	0.784*** (0.001)	0.709*** (0.001)	0.785*** (0.001)
eigenvector	-0.212*** (0.001)	-0.223*** (0.001)	-0.221*** (0.001)	-0.233*** (0.001)	-0.240*** (0.001)	-0.222*** (0.001)	-0.235*** (0.001)
betweenness × coherence	-0.270*** (0.001)		-0.284*** (0.001)			0.002 (0.001)	
coherence × eigenvector	0.000 (0.001)		-0.001 (0.001)				
betweenness × rawdist		0.061*** (0.004)		0.057*** (0.004)			
rawdist × eigenvector		-0.080*** (0.001)		-0.080*** (0.001)			-0.082*** (0.001)
reg_comp			0.010*** (0.001)	0.014*** (0.001)	0.015*** (0.002)	0.011*** (0.002)	0.015*** (0.002)
tech_comp			-0.254*** (0.001)	-0.255*** (0.001)	-0.243*** (0.001)	-0.241*** (0.001)	-0.241*** (0.001)
x8_human_capital					0.044*** (0.003)	0.044*** (0.003)	0.044*** (0.003)
x9_competition					0.088*** (0.005)	0.088*** (0.005)	0.088*** (0.005)
x10_product_innovation					0.016*** (0.002)	0.017*** (0.002)	0.016*** (0.002)
x11_process_innovation					-0.020*** (0.004)	-0.022*** (0.004)	-0.020*** (0.004)
x12_high_growth					0.008** (0.003)	0.009** (0.003)	0.008** (0.003)
x13_internationalization					-0.080*** (0.003)	-0.082*** (0.003)	-0.080*** (0.003)
x14_risk_capital					-0.035*** (0.002)	-0.034*** (0.002)	-0.035*** (0.002)
coherence × betweenness						-0.292*** (0.002)	
rawdist × betweenness							0.050*** (0.004)
SD (Observations)	1.346	1.356	1.307	1.318	1.310	1.296	1.308
SD (Intercept regioncountry)					1.029	1.036	1.028
SD (Intercept country)					1.124	1.135	1.123
SD (Intercept region)	1.497	1.489	1.513	1.499			
Num.Obs.	1999432	1999432	1906137	1906137	1506300	1506300	1506300
R2 Marg.	0.190	0.186	0.202	0.198	0.197	0.201	0.198
R2 Cond.	0.638	0.631	0.659	0.651	0.659	0.668	0.659
AIC	6864439.5	6894186.6	6431985.8	6463890.4	5090780.3	5059555.4	5086578.2
BIC	6864689.6	6894436.8	6432259.9	6464164.5	5091110.4	5059909.9	5086932.7
ICC	0.6	0.5	0.6	0.6	0.6	0.6	0.6
RMSE	1.35	1.36	1.31	1.32	1.31	1.30	1.31

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

5.5 Role of space

Having established stable dyadic mechanisms at the technology–region level (Tables Table 5.2 and Table 5.3) and substantial between-place heterogeneity in the hierarchical models (Table Table 5.4), we test whether regional potential exhibits spatial dependence. We aggregate potential to region-year means and estimate spatial panel models with 5-nearest-neighbor weights matrices. In alignment with the robust Lagrange Multiplier tests presented in Table 4.9 in the previous chapter, we treat the SAR model as our main empirical assessment here and consider the SEM model for robustness. Essentially, this is also in alignment with the LISA map (Figure 5.1) which shows that most spatial autocorrelation is concentrated in the High-High and Low-Low quadrant of variations.

Tables Table 5.5 and Table 5.8 reveal substantial spatial dependence. In the SAR model, the spatial autoregressive parameter ρ is around 0.376 in the interaction specification. This indicates that (average) regional potential is positively associated with neighboring regions' outcomes. Albeit, the marginal effects of the covariates in this model should be observed in Table 5.7 which describes the details of the decomposition of effect type (indirect, direct, and total effects).

In the SEM model, the spatial error parameter δ is around 0.420 in the interaction specification, which indicates that spatial correlation could also be an unobserved components. These estimates are consistent with the interpretation that spatial context matters for regional feasibility landscapes, whether through spillovers or through correlated shocks and unobservables.

The spatial interaction terms in both models, especially in SAR, confirm the importance of space. These terms indicates that the regional feasibility landscape is spatially structured as we iterated previously when we discussed the results from the hierarchical model and the intuition behind it. However, these spatial parameters only capture correlation mechanisms and not causal structures.

The coefficients in Table 5.5 do not represent the marginal effects of the predictors, thus we refer to Table 5.7 for this. Our focus is the interaction specification because of the

Table 5.5: Spatial Autoregressive Model (SAR) - Regional Potential

	baseline	network	interaction
(Intercept)	0.116*** (0.003)	0.136*** (0.003)	0.132*** (0.003)
reg_comp	0.060*** (0.003)	0.055*** (0.003)	0.054*** (0.003)
patent_count	0.077*** (0.003)	0.062*** (0.003)	0.062*** (0.003)
rho	0.464*** (0.015)	0.375*** (0.016)	0.376*** (0.016)
avg_coherence		0.055*** (0.003)	0.065*** (0.003)
avg_betweenness		-0.009** (0.003)	0.009* (0.004)
avg_eigenvector		0.018*** (0.003)	0.022*** (0.003)
avg_coherence $\times$ avg_betweenness			0.013*** (0.002)
avg_coherence $\times$ avg_eigenvector			0.006*** (0.002)
Num.Obs.	3223	3223	3223
Log.Lik.	4350.51	4559.40	4578.55
AIC	-8693.02	-9104.81	-9139.09
BIC	-8668.71	-9062.26	-9084.39
Sigma	0.161	0.149	0.148
Res.Var	0.024	0.021	0.021
EDF Total	4	7	9

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Regional potential exhibits spatial dependence by aggregating to region-year observations and estimating spatial autoregressive (SAR) models where neighbors' potential directly affects own potential through the spatial lag parameter ρ . The large and highly significant spatial lag coefficient $\rho = 0.38$ in the interaction specification) indicates substantial positive spatial autocorrelation: regions tend to exhibit higher potential when neighboring regions also display high potential, even after controlling for regional portfolio characteristics (average coherence, betweenness, eigenvector) and complexity. In here regional aggregates describe the overall advantage of regional portfolios endowments and characteristics.

conditional interpretation it provides⁴.

The marginal effect of average portfolio betweenness centrality on regional potential is entirely conditioned by the level of average industrial coherence in the regional portfolio. From the interaction SAR specification (Table 5.7), the conditional total marginal effect of average betweenness (when average coherence = 0) is 0.0140. This association is further decomposed into a direct own-region effect of +0.0090 and an indirect spillover to neighbouring regions of +0.0050. This picture changes substantially as coherence departs from its mean. At one standard deviation above mean coherence the total betweenness effect rises to +0.0348 (more than double the mean-coherence estimate). At one standard deviation below mean coherence the total effect of average betweenness turns negative at -0.0068, with both the direct (-0.0043) and indirect (-0.0025) channels reversing sign as well. This means that regions characterised with high aggregate coherence are associated with higher average potential when their portfolio's average betweenness increase by one SD, whereas for low coherence regions, higher average betweenness negatively influence their average potential. This is to say, that regions already endowed with bridging (general purpose) technologies would benefit more from higher aggregate industrial alignment on their average potential. The sign flip occurs at approximately -0.67 SD: regions below this coherence threshold experience a negative portfolio-level betweenness effect, while regions above it experience a positive one. The interaction coefficient (+0.0208 total) is 142% of the betweenness main effect (+0.0140). In a non-causal sense, coherence is the governing condition that determines the direction and magnitude here. A portfolio with bridge like technologies without adequate industrial alignment is not sufficient for regions to increase their aggregate potential. To embed and activate those bridging connections into portfolio-level feasibility advantage a baseline of aggregate industrial alignment is required. This is further illustrated in Table 5.6, where we can see the association clearly

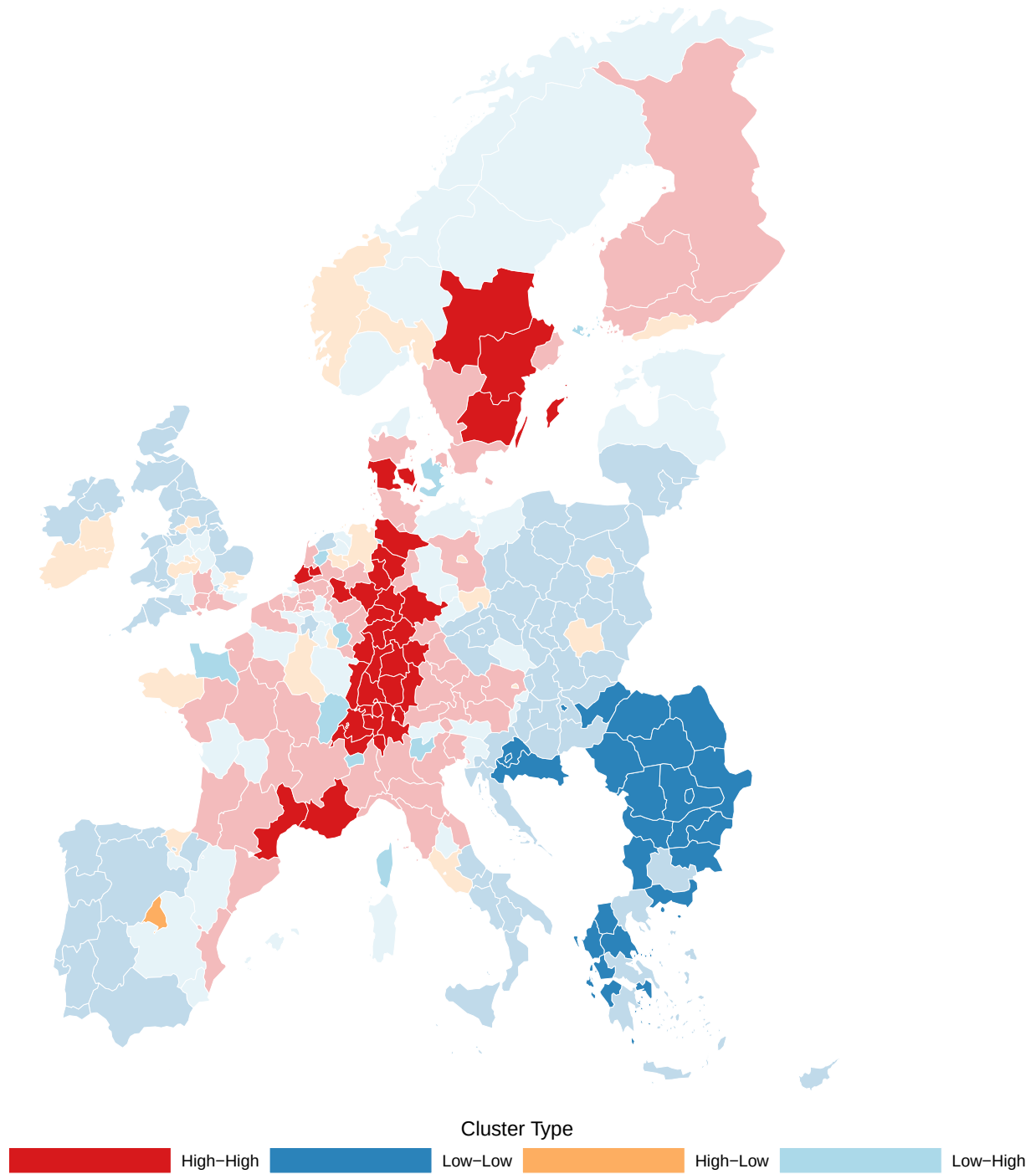
⁴In a SAR model the estimated coefficient vector $\hat{\beta}$ does not directly recover marginal effects due to the spatial multiplier $(I - \rho W)^{-1}$. The total parametric impacts reported in Table 5.7 – derived following LeSage and Pace (2009) – decompose each covariate's total effect into a direct component (own-region response, averaging the diagonal of the partial derivative matrix $\partial \mathbf{y} / \partial \mathbf{x}_k$) and an indirect component (average spillover to all other regions, averaging the off-diagonal entries). It is these decomposed total impacts, not the raw coefficient vector, that recover the correct marginal effects and are used throughout

Table 5.6: Heterogeneity of average betweenness at different average coherence deciles

	reg_potential	avg_coherence	avg_betweenness	decile pot	decile coh	decile bet
above coherence threshold						
ITC4	0.60	0.58	0.0038	9	9	3
DE92	0.60	0.58	0.0042	9	10	4
FRL0	0.58	0.57	0.0049	9	8	7
below coherence threshold						
CZ08	0.01	0.52	0.0040	2	2	3
NL34	0.02	0.51	0.0050	3	2	7
ES13	0.01	0.52	0.0059	2	2	9

for two groups of regions: those above the coherence threshold: Lombardia (ITC4, IT), Hannover (DE92, DE), and Provence-Alpes-Côte d’Azur (FRL0, FR); and those below it: Moravskoslezsko (CZ08, CZ), Zeeland (NL34, NL), and Cantabria (ES13, ES). In this table we can see that even with low average betweenness (third, forth, and seventh deciles), regions above the coherence threshold have higher average regional potential (ninth decile), whereas for the regions below the coherence threshold (having average coherence below the second decile) have lower average potential despite having higher average betweenness than the regions above the same threshold. This confirms and reflects known geographic dichotomies in Europe specifically the core-periphery divide.

The marginal effect of average portfolio eigenvector centrality complements the picture we discussed above. From the interaction SAR specification (Table 5.7), the conditional total marginal effect of average eigenvector at coherence level $\tilde{c}$ is $0.0354 + 0.0098 \cdot coherence$. At mean coherence ($coherence = 0$), a one-standard-deviation increase in average portfolio eigenvector is associated with a total effect of +0.0354 on average regional potential, this effect is further decomposed into a direct effect of +0.0227 and an indirect spillover of +0.0127. At one standard deviation above mean coherence, the total effect rises to +0.0452; at one standard deviation below, it falls to +0.0256. In contrast to the average betweenness coefficient, the average eigenvector effect never reverses sign across the coherence distribution. The sign-flip threshold would require coherence to be at -3.61 SD, well outside our current empirical range. Furthermore, the interaction coefficient



- (a) Spatial clustering of regional potential through a LISA map based on the 2018 observations, which identifies regions where high (or low) potential clusters geographically. Regions are classified into four quadrants: High-High (HH, high potential surrounded by high-potential neighbors, typically innovation cores in Western Europe), Low-Low (LL, low potential surrounded by low-potential neighbors, typically peripheral regions in Eastern Europe or rural areas), High-Low (HL, isolated high-potential regions in low-potential neighborhoods), and Low-High (LH, lagging regions surrounded by high-potential neighbors). The map reveals that most variation concentrates in the HH and LL quadrants, consistent with the positive spatial autocorrelation

Figure 5.1: Local Indicators of Spatial Association (LISA)

(+0.0098 total) is 28% of the eigenvector main effect (+0.0354 total). Thus, coherence in the case of enabling technologies is an amplifier. This is to say that regional portfolios endowed with enabling technologies have an advantage as they are generally associated with higher potential which increases as their aggregate industrial alignment also increases. But this mechanism does not determine whether that advantage exists. This stands in direct structural contrast to betweenness, where coherence conditions the advantage of betweenness. Regions whose portfolios are concentrated in high-eigenvector technologies command higher average potential regardless of their average coherence level. But the enabling advantage of holding enabling technologies is a static feature of the regional portfolio, not a conditional one. This is also supported by the positive indirect coefficient (+0.0127 at mean coherence, stable at approximately 36% of total across different coherence levels). This could mean that this enabling advantage is not locally contained as regions with authority-position portfolios generate feasibility spillovers to their neighbours, consistent with the argument in Antonietti and Montresor (2021) that enabling capacities matter when they are embedded and actively used within the local productive system rather than as isolated entry targets.

Table 5.7: SAR effect decomposition - Regional Potential

	dir			ind			tot		
	Estimate	Std. Error	Pr(> t)	Estimate	Std. Error	Pr(> t)	Estimate	Std. Error	Pr(> t)
baseline									
reg_comp	0.063	0.003	0.000	0.050	0.004	0.000	0.113	0.006	0.000
patent_count	0.080	0.003	0.000	0.063	0.004	0.000	0.143	0.006	0.000
network									
reg_comp	0.056	0.003	0.000	0.032	0.003	0.000	0.088	0.005	0.000
patent_count	0.064	0.003	0.000	0.036	0.003	0.000	0.100	0.005	0.000
avg_coherence	0.056	0.003	0.000	0.032	0.003	0.000	0.088	0.005	0.000
avg_betweenness	-0.009	0.003	0.002	-0.005	0.002	0.003	-0.014	0.005	0.002
avg_eigenvector	0.018	0.003	0.000	0.010	0.002	0.000	0.028	0.004	0.000
interaction									
reg_comp	0.056	0.003	0.000	0.031	0.003	0.000	0.087	0.005	0.000
patent_count	0.064	0.003	0.000	0.036	0.003	0.000	0.100	0.005	0.000
avg_coherence	0.067	0.004	0.000	0.038	0.003	0.000	0.104	0.006	0.000
avg_betweenness	0.009	0.004	0.030	0.005	0.002	0.033	0.014	0.006	0.030
avg_eigenvector	0.023	0.003	0.000	0.013	0.002	0.000	0.035	0.005	0.000
avg_coherence:avg_betweenness	0.013	0.003	0.000	0.007	0.002	0.000	0.021	0.004	0.000
avg_coherence:avg_eigenvector	0.006	0.002	0.000	0.004	0.001	0.001	0.010	0.003	0.000

This table reports the direct, indirect, and total parametric impacts of the SAR models for each specification. Because the spatial lag propagates any local shock through the neighbourhood network via the spatial multiplier, the raw coefficient vector does not recover marginal effects. The direct effect captures the own-region response, the indirect effect captures the average spillover to neighbouring regions, and the total effect is their sum.

These two conditional gradients (average betweenness and eigenvector) converge on coherence as the central organising condition of the regional feasibility landscape at the spatial level. From the interaction SAR specification (Table 5.7), average coherence carries a total effect of +0.1044 at mean betweenness and mean eigenvector – the largest main effect among all portfolio-level covariates – decomposing into a direct effect of +0.0669 and an indirect effect of +0.0375, with the indirect channel accounting for approximately 36% of the total. This baseline effect is itself modulated by both betweenness and eigenvector through their respective interactions, but what the conditional gradients in paragraphs above make precise is the asymmetry in how coherence operates across the two network dimensions. First, for betweenness, coherence is the condition that determines whether bridging technologies is an asset or a liability at the portfolio level. Second, for eigenvector, coherence amplifies an advantage that exists independently of alignment. In both cases, however, the direction of the coherence effect on regional potential is unambiguous and stable – higher industrial alignment is associated with higher average potential across the full distribution of betweenness and eigenvector values, with the indirect effects confirming that this alignment advantage propagates spatially to neighbouring regions. Coherence is therefore not only a dyadic feasibility correlate, as the earlier models established, but the structural condition that governs how portfolio-level network positions – bridging and enabling alike – translate into regional potential and its spatial diffusion.

5.6 Summary

In this chapter we presented the results of our empirical analysis which we introduced in the first section. The second section addressed our benchmark question of whether regional technological potential is a more informative predictor of entry than relatedness density. Our results show that potential subsumes much of the relatedness density signal and provides marginal improvements in model fit. Although potential does not clearly dominate relatedness density when both are included as co-predictors, the evidence is consistent with our conservative criterion – potential contributes more across specifications

Table 5.8: Spatial Error Model (SEM) - Regional Potential

	baseline	network	interaction
(Intercept)	0.217*** (0.006)	0.221*** (0.005)	0.217*** (0.005)
reg_comp	0.060*** (0.003)	0.059*** (0.003)	0.058*** (0.003)
patent_count	0.072*** (0.003)	0.059*** (0.003)	0.059*** (0.003)
delta	0.522*** (0.019)	0.421*** (0.023)	0.420*** (0.023)
avg_coherence		0.063*** (0.003)	0.073*** (0.004)
avg_betweenness		-0.008** (0.003)	0.009* (0.004)
avg_eigenvector		0.017*** (0.003)	0.021*** (0.003)
avg_coherence $\times$ avg_betweenness			0.012*** (0.003)
avg_coherence $\times$ avg_eigenvector			0.005** (0.002)
Num.Obs.	3223	3223	3223
Log.Lik.	4228.41	4443.00	4458.15
AIC	-8448.83	-8871.99	-8898.31
BIC	-8424.51	-8829.45	-8843.61
Sigma	0.181	0.162	0.161
Res.Var	0.025	0.023	0.022
EDF Total	4	7	9

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table estimates a spatial error model (SEM) as a robustness check to distinguish between substantive spatial interdependence (neighbors' outcomes affecting own outcome, as in SAR) and spatially correlated unobservables (common shocks or omitted factors that cluster geographically). The spatial error parameter lambda is positive and significant (0.42 in the interaction specification). After controlling for observable regional characteristics, residuals remain spatially autocorrelated. However, the robust Lagrange Multiplier diagnostics reported in the previous chapter favor the SAR specification. The regional covariate patterns in SEM closely mirror those in SAR.

5 Results

and there is no specification in which relatedness density dominates potential. We therefore accept H1, while acknowledging that a more robust comparison is warranted.

The third section presented the results of model 1 concerning the influence of technology characteristics and portfolio alignment on regional technological potential. Coherence is consistently the strongest predictor of potential across all specifications, confirming H3a – regions with more coherent portfolios face lower feasibility barriers across the technology space. Regarding technology characteristics, enabling technologies are consistently associated with lower potential regardless of the level of portfolio alignment, confirming H2b. This inhibiting effect is stable and does not vary with coherence, meaning H3c is not supported – the structural difficulty of entering foundational technologies appears to operate through mechanisms that portfolio alignment alone does not address. For bridging technologies, the picture is more nuanced. They are consistently associated with higher potential, confirming H2a, and this association is stronger for regions with lower coherence, confirming H3b. The advantage of bridging technologies diminishes as coherence increases but remains positive throughout the empirical distribution, meaning they are a reliable diversification pathway across the coherence spectrum, with the largest returns for regions where direct alignment is limited.

The forth section introduced national ecosystem covariates. In the fixed effects model the results were not fully consistent across specifications, which is expected given the identification constraints of that design. The hierarchical model addressed this directly and showed clear and consistent effects across all ecosystem dimensions, confirming H4. National institutional and market characteristics contribute independently to regional potential beyond what portfolio mechanisms explain, though the direction of effects varies across ecosystem dimensions.

The fifth section presented the spatial model results. The spatial lag parameter is positive and significant, confirming H5a – regional feasibility landscapes are spatially interdependent rather than locally contained. At the aggregate portfolio level, the role of coherence, bridging, and enabling characteristics shifts relative to the dyadic results. For portfolios concentrated in enabling technologies, the association with average potential is positive

5 Results

and amplified by coherence, consistent with H5c. This contrasts with the dyadic finding where enabling technologies inhibit potential – once they exist in the portfolio, their structural position in the capability network supports the feasibility of other technologies by construction, which is reflected in higher aggregate potential. For portfolios concentrated in bridging technologies, the association with average potential is conditional on coherence, partially confirming H5b. Below the coherence threshold the association is negative, and above it the association is positive. This means that bridging technologies as targeted diversification options are reliable at the dyadic level, but their contribution to aggregate portfolio potential requires a baseline of industrial alignment to materialise. Bridging characteristics alone are therefore not a sufficient portfolio-level instrument – their aggregate value depends on the broader industrial structure in which they are embedded.

6 Summary

6.1 Discussion

The covid 19 pandemic was a turning point in the global economy. Essentially it forced policy makers to reassess their decisions and their perception of resilience and sovereignty. Meanwhile uncertainties in global security and trade is only getting worse and vulnerable economies are the first victims. From here we posited that in the face of such shocks, understanding what makes an economy resilient is a sensible exercise. This was a clear research agenda in the literature which we presented and discussed in the review chapter. We framed adaptation as a theoretically observable consequence of resilience and that this adaptation, i.e. the capacity to adapt to constraints by recombining existing local capabilities in new configuration can be observed itself in the diversification patterns of these local economies. From there we justified the current study and framed it as a response to the need to a more contextualised assessment of regional technological diversification. Our methodological and empirical contribution both target the contextualisation dimension, following the research agenda proposed across several streams of literature. In this chapter we will conclude this dissertation and discuss different aspect of novelty, limitations and implication to our contributions.

6.1.1 On the methodology

Our contribution in this work spans both methodology and empirical application. The call in the literature for a more contextualised studies of regional capabilities were our main

theoretical motivation in this context. P.-A. Balland, Rigby, and Boschma (2015) call for a more focused consideration of the evolutionary accounts of regional change; technological relatedness, network structures, and institutions. However, C. Pinheiro (2025); Andreoni and Chang (2019) and Bathelt and Storper (2023) provided another optic regarding this agenda specifically on the limitation of the current conventional instruments of measuring proximity (relatedness) and thus the technology space (network) and their conceptual flaws that obscure granular context. This means that even if this research agenda is operationalised through relatedness measurements, the loss in information compound regardless. To address this specific issue, Fessina et al. (2024), Albora et al. (2023), and most notably Tacchella et al. (2023) propose a machine learning conceptualisation to the relatedness measurements.

Their work propose a shift from co-occurrence and frequency based measurements to a predictive framework that serves three objectives. First, a falsifiable and testable measure of feasibility, second, leveraging already granular data for a richer context, and third accommodate the non-linearity of knowledge structure and their interactions. We adopted this framework as a direct response to the gaps in the literature which constitutes our main methodological contribution. However, this methodology is not without limitation itself.

In fact, C. Pinheiro (2025) acknowledged the merits of this framework with a clear factual scepticism regarding the performance of these machine learning models. The problem she mentions is related mainly to false negative predictions, which are the result of the unbalanced data itself (few regions with observed specialisation, the majority of the outcome is 0). Although this is a well merited criticism, it still can be addressed in future iterations with different training strategies (using synthetic minority oversampling technique or SMOTE) or with innovative machine learning modeling algorithms (causal random forest as an potential candidate). Moreover as we've shown in the comparative example in chapter 3, FITS produce a much more stable prediction than conventional technology space, the merits of that is our main motivation for this adoption.

However, the most interesting criticism here from C. Pinheiro (2025)—which transcends our approach and target the literature in general—is the use of the Revealed Comparative

6 Summary

Advantage index as a stepping stone to compare regional patterns of specialisation. The author argue that this index is not necessarily adequate for application in (sparse) patents data because then its interpretation is flawed as it assumes bi-directional flows. We agree with this assessment and acknowledge that further attention is needed.

There are two critical cases in our opinion regarding our methodology, the first is the training validation and the second is the resulting FITS network. With machine learning algorithms in general, a validation process is needed to avoid over fitting. This is usually conducted with specific tuning and hyper parametrisation protocols. In this regard, our validation approach was less rigorous than a full grid search and cross-validation protocol, since we applied a fixed set of hyperparameters derived from a stratified sample of four technologies rather than tuning each model independently, however, we justify it via two arguments. First, the complete process with grid search, tuning, and cross validation was not omitted on purpose but because of the computational burden. For that, we needed to sacrifice rigorous specification for feasibility. Second, Albora et al. (2023) benchmark their training strategy with and without tuning and they reported that even with default parameters, the random forest algorithms is a better predictor of future specialisation/diversification patterns than relatedness, raw RCA measure, etc. Their results in this context show that with cross validation the performance of the models does increase but even without it, random forest's performance remains superior to the benchmark.

The second problem here that FITS provides an asymmetric adjacency matrix, in our case we normalise once over the outgoing edges (column standardisation) because these are the technologies that contribute or point to a focal technology so their importance metrics need to be comparable, when the network is undirected as in the conventional case of technology space, the direction of normalisation is not problematic because the matrix is symmetric, but in our case this can be flawed as the incoming direction is not similarly normalised. This was not an issue addressed in Fessina et al. (2024) where the focus was instead to have a falsifiable framework for the technology space. This limitation in our work calls for extra attention to find methods in the literature that could address this but was beyond the main scope of this dissertation.

6 Summary

On another note, in previous related work, Bennour (2026) discuss other areas of limitation to the FITS network approach. One of the issues here is that the time horizon δ is relatively arbitrary. Although Fessina et al. (2024), Albora et al. (2023), and Tacchella et al. (2023), choose it empirically by testing the most adequate value, it remains a limitation. The reason for this is that with this training strategy, every technology is modeled independently, which means that the choice of δ should change with each model. This is to say, even when Albora et al. (2023) showed that $\delta = 5$ is optimal when they assess the aggregate performance of their models, they, and consequently this work, still leave much more dynamic information on the table.

The idea here is that as we operationalise the argument in the literature that past capacities should predict future ones (Napolitano et al. 2018; Andreoni and Chang 2019; Martin and Sunley 2007), we do not know how much it takes for the past capabilities to transform into new ones. This prediction window should not logically be constant across technologies or activities that are heterogeneous in their complexity. For instance, the past capacities of a region that predict textile related technology, should not have the same time horizon as semiconductor technology or other highly complex technologies.

This is one of the areas that were discussed in Hidalgo (2023): the 4Ws that drives policy implication of the REC literature (where, what, When and Who?). The question of When diversification occur was not properly addressed in the literature with the exception of Alshamsi, Pinheiro, and Hidalgo (2018). In their work, the authors formalise a dynamic graph model that essentially predict that on average, diversification is feasible when 10% to 15% of related capabilities exist in a location. Although this aligns relatively with known graph theory ideas, their assessment depends heavily on the symmetric activity space and its structure and orientation. Regardless, Alshamsi, Pinheiro, and Hidalgo (2018)’s approach could be beneficial to address the time horizon limitation, and could result in a richer more granular contextualisation of the FITS network and regional technological potential.

The empirical design in our work also has its own share of limitations. Our assessment posits that the regional technological potential is a proxy for diversification feasibility.

6 Summary

However, there's no clear indication in the literature that such framing is adequate since we are the first to adopt it. This is to say that potential might need to be assessed with a different interpretation than the one we adopted. Moreover, using network metrics of the technologies could have been better addressed with random graph models or other alternatives designed for this topic, but in our assessment, this was beyond the scope of the dissertation and thus we settled on a simpler approach.

This work's main limitation is that we don't provide evidence of causality due to the endogeneity problem. Specifically mechanical circularity: coherence is built from the same lagged portfolio information that the Random Forest uses to generate potential. Within a given region-year, a technology with higher coherence will tend to resemble the portfolio patterns the RF associates with entry, which can inflate the coherence-potential link. This is not reverse causality in the usual sense, but shared data construction. The Two-way fixed effects mitigate this by absorbing time-invariant regional alignment, so identification relies on within-region changes over time. Furthermore, the results are similar in fixed- and random-effects models, and the key network-position variables are technology-specific rather than portfolio-derived. We therefore interpret coefficients as conditional associations describing the feasibility of diversification opportunities, not causal effects. This is a limitation that we will address in future iterations as it remains the bottleneck for actual informative policy recommendations.

Another issue here is that including the national ecosystem variables as predictors could have been more informative had we included an interaction term, but then the complexity of the interpretation forced us to not go in that direction for now, although with a country level analysis this can be addressed properly but we decided to leave that out for the purposes of this work.

Building on that, there's another area that's extremely important to address here. This is the question of the data itself that we used for the network metrics and their interaction with coherence. Our main network structure identifiers are eigenvector and betweenness centrality. We assume that the first captures prestigious position and we consider the metric a proxy to the Key enabling technology role while for the latter we assume it

to proxy General purpose technology roles. This is in accordance with the framing and classification in Antonietti and Montresor (2021).

In our context, betweenness centrality is theoretically a justifiable proxy, technologies with high betweenness centrality in our FITS network are nodes where many of the shortest paths between other nodes pass through in both directions. These technologies links different technological areas and thus can be considered as general purpose. However, we can't confidently say the same for our interpretation of eigenvector centrality. Because technologies with high eigenvector centrality, are those that are well connected to other technologies that are also well connected. A more appropriate interpretation here would have been foundational role. The distinction and interpretation of these roles is not as clear as we intended it to be, and it's a major flaw of our framing which we acknowledge and will address in future work. This however doesn't systematically affect the core agenda we subscribe to nor the general ideas we've portrayed evidence for, network structures—regardless of the interpretation—have a sizeable contribution to region diversification.

6.1.2 On the results

Our results presented in the previous chapter, showcase an interesting story derived from a research agenda framed almost a decade ago. The entire idea from this work is to operationalise the idea that “context matters”. Our design focuses on explaining “How context matters? and for Who?”, we achieve this by including three contextual layers to assess diversification potential of regions in Europe, relatedness, network structures, and institutions.

We started by arguing that in order to obtain granular context, we first need to address the conceptual issues in relatedness and the derived networks via our regional potential and its derived FITS network. We extended that with defining another measure of industrial embeddedness which we use as a device that measures the industrial alignment between a region's network, and a technology. And finally we adopted a set of pillars from the

GEI data to proxy institutional structure. With that we addressed the concerns in the literature on contextualisation.

The three models we specified each target a specific idea, combined they are each other robustness checks. The benchmark model 0 shows that regional potential and entry are both associated with a positive change in the probability of that a region will enter a new technology. But because both retain significant explanatory power when introduced together, we cannot confidently say that our measure of potential is a better predictor, despite the fact that across all specification the contribution of potential to the models is higher than that of relatedness density. This result is more modest than anticipated, as we expected to see a much more significant difference. Regardless, the overall picture when we take the results of all specifications together tell us one thing, there’s no evidence that relatedness density predict more variation than regional potential.

However, the marginal benchmark advantage shown in the results we’ve seen might not reflect the entirety of the picture. we argue this is the case for two reasons, first, unlike most of the studies who use similar specification, we are limited to the time window of the estimated regional potential. Most studies here include up to 30 years of data, while we’re limited only to 10 (1 year is lagged), albeit we have an extensive panel of observation (1.57 million observation), our findings might be affected by this issue. Second, we didn’t pay more attention to other covariates who might contribute more to the predictive power of the model. Essentially this is to say that although the benchmark is not strongly conclusive, a thorough comparison should be in the agenda of future work.

When we address the influence of network structure on regional potential, our results show consistency and robustness in the dyadic models 1 and 2. Throughout, technologies with high betweenness centrality—assuming a bridging role in the network—are shown to be the most feasible. Whereas this is the complete opposite for technologies with high eigenvector centrality—assuming an enabling role in the network—who are shown to be an inhibitor of diversification.

To contextualise these results further, we tested whether the alignment in industrial embeddedness between a region and a given technology influence the contribution of bridging

6 Summary

or enabling technologies to regional potential. Our results show that this actually is the case specifically for bridging technologies. Essentially, the evidence from model 1 and 2 tell us that bridging technologies are more more influential on regional potential for regions with low coherence in those technologies. This influence is then dampened, but remains positive and significant for regions with higher coherence. Albeit, for enabling technologies, we've seen no conclusive results on the moderating effect of coherence, this is to say that an enabling technology remains a highly challenging diversification option regardless of the industrial alignment of the region with that technology.

Furthermore, we added another layer of context as we included the national ecosystem covariates. We started by assessing their influence in model 1 and we've seen a relative stability for few of the pillars we introduced. Of course the apparent inconsistencies were not empirical, but rather structural since the fixed effects models do not handle the nested structure of the data properly. This was confirmed in model 2 where we adopted a hierarchical model that accounts for that structure. Model 2 shows a clear consistency in the results and provide evidence for the case on the importance of institution in regional diversification as the variables were all significantly associated with the outcome. Our results show that the most predictive pillar of the entrepreneurial ecosystem of regional technological potential is competition (positively associated at 8.8%) and internationalisation (negatively associated at 8%).

These results tell us an interesting story that's bounded to the reality of European regions. First, the internationalisation pillar as defined in Szerb et al. (2013) and subsequent updates are informative of export intensity, and export complexity of the country. This means that ecosystems that are heavily export oriented inhibit their regional technological potential. This was relatively mirrored in the literature of trade and global value chains, and more specifically in Sebestyén, Braun, and Elekes (2024). The idea here is that some export intensive countries might tend to lock themselves out into activities that have low added value in complexity and innovation intensity, or into activities where the value added is entirely exported. However we acknowledge here that further research is needed given the heterogeneity between European regions. Second, our results on the role of competition

6 Summary

confirms the intuition in Komlósi et al. (2025) who employs a necessary condition analysis to assess the structure of the ecosystem that shape regional economic performance. In their study they found that in the context of their sample of EU regions, competition was almost always necessary for their outcome.

Model 3 completes our final layer of contextualisation and confirms our initial intuition relatively supported by the variance decomposition in model 2. Space matters, which country a region belongs to and who are their neighbours are a structural baseline predictors that close the logical chain we started in this work.

When we average our main variables of interest to a regional panel (eliminate the individual technology signal) we have a different level of analysis. Unlike model 1 and 2 who ask what are the (joint) characteristics of regions and technologies that predict regional potential in each technology, model 3 asks on aggregate, does the characteristics of the technologies endowed in the region predict it aggregate (portfolio level) potential? I.e. now we can see if the stock of bridging and enabling technologies predict the region's average feasible opportunities, and whether or not the spatial component is confirmed. To this end, the results provide a contrasting story to the dyadic models. First, the average potential of a region is spatially dependent on the average potential of it neighbouring regions. Second, regions endowed with high level of bridging roles technologies are associated with a decrease in their potential, whereas for enabling roles technologies, they are positively associated. Lastly, average coherence conditions the marginal effect of bridging portfolios on average regional potential but not enabling portfolios. This seems in contradiction with the results of models 1 and 2, but a deeper assessment is warranted. At the aggregate regional level, model 3 informs us that regions with low industrial alignment do not benefit (in terms of their average potential) from an increase in their bridging technologies endowment, there's a threshold of industrial alignment that condition whether or not an increase in bridging technologies improve regional potential. This complements our results in model 1 and 2, bridging technologies are a facilitator of diversification for regions with low coherence, and the opposite for regions with high coherence. This means that bridging technologies matter only for catch-up, at some inflection point its effect reverses. Similarly,

regions highly endowed with enabling technologies are associated with larger diversification possibilities. This reflects the merit of the naming itself, because enabling technologies enable (positively associated with) higher regional potential on average regardless of the industrial alignment. This is one of our most important conclusions here, because this means that bridging technologies at the portfolio level are not universally beneficial when considered as policy direction. However, they can be considered as a targeted catalyst to improve industrial alignment itself because at the dyadic level they are associated with higher regional potential. This is to say that bridging technologies should be considered as a secondary, targeted, and precise diversification option, that supports other primary targets aimed at improving the overall industrial alignment.

6.1.3 Policy implication

Our work here is not merely an empirical application and contains many area of important policy implications. We will divide these implication into two parts, the first addresses the the asymmetric FITS network and regional interdependence, while the second is regarding the characteristics of the technologies.

First, as we've mentioned in earlier section the FITS network reveals that technologies are not related by homogeneous/undirected links. There's a clear, consistent and logical unbalance on how much a technology is related to another. For regions this means that with their endowment of any technological portfolio, they also inherit this asymmetric dependence as well (Wanzenböck et al. 2025). This dependence is then spatially dependent also because a region's technological portfolio doesn't exist in a vacuum or only in the limited context of that region but in other regions as well (Stuck, Broekel, and Revilla Diez 2016). A region who specialises in precision machinery, would use technologies from regions who specialise in robotic systems as an example, but the reverse may not be true or at least weaker. Leveraging this asymmetry aligns with the literature on regional linkages and could be useful is in the framework of European cohesion.

The granular and context-rich information FITS provide can therefore enable an assessment

of technological linkages at a regional level, with that for each technology we can know which regions are innovate together, which regions compete together, and which regions complement the activities of other regions. Such mapping would be extremely helpful to increase European technological integration, strengthen existing regional capacities, manage competition, and increase the efficiency of the innovation landscape in general (Migueluez and Moreno 2018).

Second, the characteristics of the technologies in the FITS network as we studied them inform us that they condition the feasible diversification options of a region. Our findings, can enable policy makers to strategise depending on the initial situation in the region: their portfolio size, diversity, and stock characteristics in term of bridging or enabling technologies. There's no one size fit all solution, and our results explicitly show that. Given the heterogeneity in the technological characteristics of European regions, policy makers can prioritise incentives and investment in specific the specific technologies, for regions on the technological periphery this means a focus on technologies with bridging roles, those that are general purpose and can be used in different other technologies or domains. Other regions with higher level of coherence, could incentivise investments in the ecosystem in general to support the establishment and survival of high growth firms, product innovation, harbour competition and human capital. This however is not a single layer, combining these policy directions with other policy directives derived from the inter-regional technological linkages we discussed in the first point would directly support European cohesion directives and branch out other policy paradigms.

6.2 Future research agenda

This dissertation started with a specific framing issued from clear and relevant policy and research gap in the literature. The need to further contextualise the study of regional capabilities through more detailed context. We responded to this agenda with a methodological and empirical framing. Albeit, we are convinced of the general usefulness of our approach much more work remains to fully conceptualise an ontological framework. As we discussed

6 Summary

earlier, the machine learning approaches are a shift from a descriptive scope of relatedness and the technology space to a predictive one. The result of such a shift is a context-rich asymmetric network that contains much more information on the dependencies between technologies than the conventional relatedness approaches. Similarly, the Entrepreneurial Ecosystem indicators are an appropriate summary of the national institutional context, that are reliable and used extensively in the literature. These three elements constitute a baseline for future work.

First, the asymmetric FITS network is great start in this direction to accommodate different ideas in the literature mainly regional linkage. To that end, a more robust approach is needed. This means instead of filtering the links of the initial RF estimation, we can adopt a sampling based falsifiable approach as was the case in Fessina et al. (2024), with that we can keep only statistically significant links. Other approaches here like causal RF could be an interesting direction also, such an approach would allow us not only to see relevant important links in the network, but also the causal links as well. This will allow us, in future iteration to assess the causal implications of adding a technology or removing from the region and could prove a powerful tool for policy makers. Adding to that, we can also flip the dimensions of the estimation process in our training strategy, instead of predicting the outcome based on the information of the technology portfolio in a region, we can predict the outcome based on the portfolio of other regions. I.e. we can predict a region's own outcome from the contribution of all the others regions, which would allow us to create an inter-regional dependence matrix (Feature Importance Regional Space) that enable us to investigate regional linkages in a more depth and certitude.

Second, the patents data that we use contains an important dimension that we ignored in this work, which is the collaboration patterns. Many studies in the literature investigate the role of collaboration in regional technological adoption, albeit the results in general are not conclusive the general consensus is that collaboration is a catalyst for technological adoption because it mirrors knowledge import which can be used to recombine existing knowledge stock into new knowledge. Our instinct here, is that these patterns are extremely informative and accounting for them would further strengthen the ideas we argued for in

this dissertation and provide a clearer path for policy. The challenge in this context is how to accommodate collaboration patterns in a machine learning-based estimations, and whether this accommodation is even necessary. What is clear however, is that collaboration patterns—which are considered as a catalyst for innovation—could have a circular feedback effect on the entrepreneurial ecosystems characteristics, whether these characteristics influence more collaboration, or collaboration influence these characteristics is question that deserve attention.

6.3 Conclusion

In conclusion, this dissertation provides a methodological and an empirical answer to explicit gaps and research agenda in the literature of Evolutionary Economic Geography. The frequent calls for a contextualised assessment of regional diversification opportunities led us to integrate different streams of literature in one body of work which was an extremely challenging task.

We first, addressed a methodological issue in the literature of relatedness and economic complexity specifically targeting the conventional measures of relatedness/proximity that obscure much of the granular information in the patenting data. To this end, we adopted a machine learning training strategy that addresses this issue, then we used the results of our approach to investigate how the characteristics of different technologies, the regions' own network structures, their national entrepreneurial ecosystem characteristics, and their spatial position associate with their technological potential (or the feasible opportunities). This methodological contribution and empirical investigation respond to the research agenda and the gaps in the literature.

Our results, validate many of the ideas in the literature and highlight the importance of the national context and the heterogeneity of the influence of different technological characteristics on regions' technological potential. Our main findings suggest that bridging technologies can be used a secondary targeted policy direction as they are relatively feasible. These targets should aim at increasing the aggregate industrial alignment of the region as

6 *Summary*

this alignment is the key to unlock the benefit of having this type of technologies. Our results also show that the entrepreneurial ecosystem characteristics at the national level have a significant predictive power and that it should be considered when designing policy instrument specifically targeting human capital, competition, and supporting high growth firms.

References

- Acs, Zoltan J, Erkkö Autio, and László Szerb. 2014. “National Systems of Entrepreneurship: Measurement Issues and Policy Implications.” *Research Policy* 43 (3): 476–94.
- Acs, Zoltan J, Saul Estrin, Tomasz Mickiewicz, and László Szerb. 2018. “Entrepreneurship, Institutional Economics, and Economic Growth: An Ecosystem Perspective.” *Small Business Economics* 51 (2): 501–14.
- Albora, Giambattista, Luciano Pietronero, Andrea Tacchella, and Andrea Zaccaria. 2023. “Product Progression: A Machine Learning Approach to Forecasting Industrial Upgrading.” *Scientific Reports* 13 (1): 1481.
- Alshamsi, Aamena, Flávio L Pinheiro, and Cesar A Hidalgo. 2018. “Optimal Diversification Strategies in the Networks of Related Products and of Related Research Areas.” *Nature Communications* 9 (1): 1328.
- Amsden, Alice Hoffenberg. 1989. *Asia’s Next Giant: South Korea and Late Industrialization*. Oxford University Press.
- Andreoni, Antonio, and Ha-Joon Chang. 2019. “The Political Economy of Industrial Policy: Structural Interdependencies, Policy Alignment and Conflict Management.” *Structural Change and Economic Dynamics* 48: 136–50.
- Antonietti, Roberto, and Sandro Montresor. 2021. “Going Beyond Relatedness: Regional Diversification Trajectories and Key Enabling Technologies (KETs) in Italian Regions.” *Economic Geography* 97 (2): 187–207.
- Argiles, Raquel Ortega, Zoltan J Acs, Laszlo Szerb, Erkkö Autio, and Eva Komlosi. 2014. “REDI: The Regional Entrepreneurship and Development Index—Measuring Regional Entrepreneurship Final Report.”

References

- Audretsch, David B, and Maksim Belitski. 2017. “Entrepreneurial Ecosystems in Cities: Establishing the Framework Conditions.” *The Journal of Technology Transfer* 42 (5): 1030–51.
- Audretsch, David B, and Maryann P Feldman. 1996. “R&d Spillovers and the Geography of Innovation and Production.” *The American Economic Review* 86 (3): 630–40.
- Balassa, Bela. 1965. “Trade Liberalisation and ‘Revealed’ Comparative Advantage 1.” *The Manchester School* 33 (2): 99–123.
- Balland, PA, and R Boschma. 2019. “Smart Specialization: Beyond Patents.” *European Commission: Brussels, Belgium*.
- Balland, Pierre-Alexandre, and David Rigby. 2017. “The Geography of Complex Knowledge.” *Economic Geography* 93 (1): 1–23.
- Balland, Pierre-Alexandre, David Rigby, and Ron Boschma. 2015. “The Technological Resilience of US Cities.” *Cambridge Journal of Regions, Economy and Society* 8 (2): 167–84.
- Baltagi, Badi H, Seuck Heun Song, and Won Koh. 2003. “Testing Panel Data Regression Models with Spatial Error Correlation.” *Journal of Econometrics* 117 (1): 123–50.
- Bathelt, Harald, and Michael Storper. 2023. “Related Variety and Regional Development: A Critique.” *Economic Geography* 99 (5): 441–70.
- Bennour, Mohamed Hsin. 2026. “Regional Potential and Friction for Technological Diversification: A Machine Learning Frontier Approach.” *Journal of Physics: Complexity* 7 (1): 015005.
- Boschma, Ron. 2004. “Competitiveness of Regions from an Evolutionary Perspective.” *Regional Studies* 38 (9): 1001–14.
- . 2005. “Proximity and Innovation: A Critical Assessment.” *Regional Studies* 39 (1): 61–74.
- . 2015. “Towards an Evolutionary Perspective on Regional Resilience.” *Regional Studies* 49 (5): 733–51.
- . 2017. “Relatedness as Driver of Regional Diversification: A Research Agenda.” *Regional Studies* 51 (3): 351–64.
- Boschma, Ron, Pierre-Alexandre Balland, and Dieter Franz Kogler. 2015. “Relatedness

References

- and Technological Change in Cities: The Rise and Fall of Technological Knowledge in US Metropolitan Areas from 1981 to 2010.” *Industrial and Corporate Change* 24 (1): 223–50.
- Boschma, Ron, and Simona Iammarino. 2009. “Related Variety, Trade Linkages, and Regional Growth in Italy.” *Economic Geography* 85 (3): 289–311.
- Boschma, Ron, Ernest Miguelez, Rosina Moreno, and Diego B Ocampo-Corrales. 2023. “The Role of Relatedness and Unrelatedness for the Geography of Technological Breakthroughs in Europe.” *Economic Geography* 99 (2): 117–39.
- Bresnahan, Timothy F. 2003. “The Mechanisms of Information Technology’s Contribution to Economic Growth.” In *Institutions, Innovation and Growth*. Edward Elgar Publishing.
- Brown, Ross, and Colin Mason. 2017. “Looking Inside the Spiky Bits: A Critical Review and Conceptualisation of Entrepreneurial Ecosystems.” *Small Business Economics* 49 (1): 11–30.
- Bun, Joël, Jean-Philippe Bouchaud, and Marc Potters. 2017. “Cleaning Large Correlation Matrices: Tools from Random Matrix Theory.” *Physics Reports* 666: 1–109.
- Bustos, Sebastián, Charles Gomez, Ricardo Hausmann, and César A Hidalgo. 2012. “The Dynamics of Nestedness Predicts the Evolution of Industrial Ecosystems.” *PloS One* 7 (11): e49393.
- Carayannis, Elias, and Evangelos Grigoroudis. 2016. “Quadruple Innovation Helix and Smart Specialization: Knowledge Production and National Competitiveness.” 10 (1 (eng)): 31–42.
- Christopherson, Susan, Jonathan Michie, and Peter Tyler. 2010. “Regional Resilience: Theoretical and Empirical Perspectives.” *Cambridge Journal of Regions, Economy and Society* 3 (1): 3–10.
- Cohen, Wesley M, Daniel A Levinthal, et al. 1990. “Absorptive Capacity: A New Perspective on Learning and Innovation.” *Administrative Science Quarterly* 35 (1): 128–52.
- Coniglio, Nicola D, Davide Vurchio, Nicola Cantore, and Michele Clara. 2021. “On the Evolution of Comparative Advantage: Path-Dependent Versus Path-Defying Changes.”

References

- Journal of International Economics* 133: 103522.
- Cooke, Philip, and Kevin Morgan. 1993. "The Network Paradigm: New Departures in Corporate and Regional Development." *Environment and Planning D: Society and Space* 11 (5): 543–64.
- Diodato, Dario, Lorenzo Napolitano, Emanuele Pugliese, and Andrea Tacchella. 2023. "Economic Complexity for Regional Industrial Strategies." Joint Research Centre.
- E, Pugliese, and Tacchella A. 2021. "Economic Complexity Analytics: Country Factsheets." <https://doi.org/10.2760/368138>.
- Fessina, Massimiliano, Giambattista Albora, Andrea Tacchella, and Andrea Zaccaria. 2024. "Identifying Key Products to Trigger New Exports: An Explainable Machine Learning Approach." *Journal of Physics: Complexity* 5 (2): 025003.
- Fingleton, Bernard, Harry Garretsen, and Ron Martin. 2012. "Recessionary Shocks and Regional Employment: Evidence on the Resilience of UK Regions." *Journal of Regional Science* 52 (1): 109–33.
- Frenken, Koen, and Ron A Boschma. 2007. "A Theoretical Framework for Evolutionary Economic Geography: Industrial Dynamics and Urban Growth as a Branching Process." *Journal of Economic Geography* 7 (5): 635–49.
- Gertler, Meric S, Päivi Oinas, Michael Storper, and Philip Scranton. 1995. "Discussion of Regional Advantage: Culture and Competition in Silicon Valley and Route 128 by AnnaLee Saxenian." *Economic Geography* 71 (2): 199–207.
- Hartmann, Dominik, Miguel R Guevara, Cristian Jara-Figueroa, Manuel Aristarán, and César A Hidalgo. 2017. "Linking Economic Complexity, Institutions, and Income Inequality." *World Development* 93: 75–93.
- Hausmann, Ricardo, and César A Hidalgo. 2011. "The Network Structure of Economic Output." *Journal of Economic Growth* 16: 309–42.
- Hekkert, Marko P, Roald AA Suurs, Simona O Negro, Stefan Kuhlmann, and Ruud EHM Smits. 2007. "Functions of Innovation Systems: A New Approach for Analysing Technological Change." *Technological Forecasting and Social Change* 74 (4): 413–32.
- Helpman, Elhanan. 1998. *General Purpose Technologies and Economic Growth*. MIT press.

References

- Hidalgo, César A. 2021. “Economic Complexity Theory and Applications.” *Nature Reviews Physics* 3 (2): 92–113.
- . 2023. “The Policy Implications of Economic Complexity.” *Research Policy* 52 (9): 104863.
- Hidalgo, César A, Pierre-Alexandre Balland, Ron Boschma, Mercedes Delgado, Maryann Feldman, Koen Frenken, Edward Glaeser, et al. 2018. “The Principle of Relatedness.” In *Unifying Themes in Complex Systems IX: Proceedings of the Ninth International Conference on Complex Systems 9*, 451–57. Springer.
- Hidalgo, César A, and Ricardo Hausmann. 2009. “The Building Blocks of Economic Complexity.” *Proceedings of the National Academy of Sciences* 106 (26): 10570–75.
- Hidalgo, César A, Bailey Klinger, A-L Barabási, and Ricardo Hausmann. 2007. “The Product Space Conditions the Development of Nations.” *Science* 317 (5837): 482–87.
- Isaksen, Arne, and Michaela Trippel. 2017. “Exogenously Led and Policy-Supported New Path Development in Peripheral Regions: Analytical and Synthetic Routes.” *Economic Geography* 93 (5): 436–57.
- Komlósi, Éva, Marcus Dejardin, László Szerb, and Balázs Páger. 2025. “Is a Balanced Entrepreneurial Ecosystem Essential for Success? A Configurational Analysis of European Regional Entrepreneurial Ecosystems.” *The Journal of Technology Transfer* 50 (4): 1669–1708.
- Landau, William Michael. 2021. “The Targets r Package: A Dynamic Make-Like Function-Oriented Pipeline Toolkit for Reproducibility and High-Performance Computing.” *Journal of Open Source Software* 6 (57): 2959.
- Lawson, Clive. 1999. “Towards a Competence Theory of the Region.” *Cambridge Journal of Economics* 23 (2): 151–66.
- Lee, Changyong, Ohjin Kwon, Myeongjung Kim, and Daeil Kwon. 2018. “Early Identification of Emerging Technologies: A Machine Learning Approach Using Multiple Patent Indicators.” *Technological Forecasting and Social Change* 127: 291–303.
- Lee, Keun, and Franco Malerba. 2017. “Catch-up Cycles and Changes in Industrial Leadership: Windows of Opportunity and Responses of Firms and Countries in the Evolution of Sectoral Systems.” *Research Policy* 46 (2): 338–51.

References

- Leendertse, Jip, Mirella Schrijvers, and Erik Stam. 2022. “Measure Twice, Cut Once: Entrepreneurial Ecosystem Metrics.” *Research Policy* 51 (9): 104336.
- LeSage, James, and Robert Kelley Pace. 2009. *Introduction to Spatial Econometrics*. Chapman; Hall/CRC.
- Li, Yang, and Frank MH Neffke. 2024. “Evaluating the Principle of Relatedness: Estimation, Drivers and Implications for Policy.” *Research Policy* 53 (3): 104952.
- Mariani, Manuel Sebastian, Zhuo-Ming Ren, Jordi Bascompte, and Claudio Juan Tessone. 2019. “Nestedness in Complex Networks: Observation, Emergence, and Implications.” *Physics Reports* 813: 1–90.
- Martin, Ron. 2012. “Regional Economic Resilience, Hysteresis and Recessionary Shocks.” *Journal of Economic Geography* 12 (1): 1–32.
- Martin, Ron, and Peter Sunley. 2007. “Complexity Thinking and Evolutionary Economic Geography.” *Journal of Economic Geography* 7 (5): 573–601.
- Mealy, Penny, and Diane Coyle. 2022. “To Them That Hath: Economic Complexity and Local Industrial Strategy in the UK.” *International Tax and Public Finance* 29 (2): 358–77.
- Migueluez, Ernest, and Rosina Moreno. 2018. “Relatedness, External Linkages and Regional Innovation in Europe.” *Regional Studies* 52 (5): 688–701.
- Montresor, Sandro, and Francesco Quatraro. 2017. “Regional Branching and Key Enabling Technologies: Evidence from European Patent Data.” *Economic Geography* 93 (4): 367–96.
- Napolitano, Lorenzo, Evangelos Evangelou, Emanuele Pugliese, Paolo Zeppini, and Graham Room. 2018. “Technology Networks: The Autocatalytic Origins of Innovation.” *Royal Society Open Science* 5 (6): 172445.
- Neffke, Frank, Matté Hartog, Ron Boschma, and Martin Henning. 2018. “Agents of Structural Change: The Role of Firms and Entrepreneurs in Regional Diversification.” *Economic Geography* 94 (1): 23–48.
- Neffke, Frank, Martin Henning, and Ron Boschma. 2011. “How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions.” *Economic Geography* 87 (3): 237–65.

References

- Nelson, Richard R, and Sidney G Winter. 1985. *An Evolutionary Theory of Economic Change*. harvard university press.
- Nesta, Lionel, and Pier Paolo Saviotti. 2005. “Coherence of the Knowledge Base and the Firm’s Innovative Performance: Evidence from the US Pharmaceutical Industry.” *The Journal of Industrial Economics* 53 (1): 123–42.
- Pendall, Rolf, Kathryn A Foster, and Margaret Cowell. 2010. “Resilience and Regions: Building Understanding of the Metaphor.” *Cambridge Journal of Regions, Economy and Society* 3 (1): 71–84.
- Penrose, Edith Tilton. 2009. *The Theory of the Growth of the Firm*. Oxford university press.
- Perez, Carlota, and Luc Soete. 1988. “2 1 Catching up in Technology: Entry Barriers and Windows of Opportunity.” Umweltbundesamt Berlin.
- Pike, Andy, Stuart Dawley, and John Tomaney. 2010. “Resilience, Adaptation and Adaptability.” *Cambridge Journal of Regions, Economy and Society* 3 (1): 59–70.
- Pinheiro, Cristina. 2025. “Relatedness and Economic Complexity as Tools for Industrial Policy: Insights and Limitations.” *Structural Change and Economic Dynamics* 72: 1–10.
- Pinheiro, Flavio L, Pierre-Alexandre Balland, Ron Boschma, and Dominik Hartmann. 2025. “The Dark Side of the Geography of Innovation: Relatedness, Complexity and Regional Inequality in Europe.” *Regional Studies* 59 (1): 2106362.
- Pinheiro, Flávio L, Dominik Hartmann, Ron Boschma, and César A Hidalgo. 2022. “The Time and Frequency of Unrelated Diversification.” *Research Policy* 51 (8): 104323.
- Pugliese, Emanuele, Giulio Cimini, Aurelio Patelli, Andrea Zaccaria, Luciano Pietronero, and Andrea Gabrielli. 2019. “Unfolding the Innovation System for the Development of Countries: Coevolution of Science, Technology and Production.” *Scientific Reports* 9 (1): 16440.
- Pugliese, Emanuele, Lorenzo Napolitano, Andrea Zaccaria, and Luciano Pietronero. 2019. “Coherent Diversification in Corporate Technological Portfolios.” *PloS One* 14 (10): e0223403.
- Qian, Haifeng, Zoltan J Acs, and Roger R Stough. 2013. “Regional Systems of Entrepreneurship: The Nexus of Human Capital, Knowledge and New Firm Formation.”

References

- Journal of Economic Geography* 13 (4): 559–87.
- Quatraro, Francesco. 2010. “Knowledge Coherence, Variety and Economic Growth: Manufacturing Evidence from Italian Regions.” *Research Policy* 39 (10): 1289–1302.
- Rocchetta, Silvia, Andrea Mina, Changjun Lee, and Dieter F Kogler. 2022. “Technological Knowledge Spaces and the Resilience of European Regions.” *Journal of Economic Geography* 22 (1): 27–51.
- Sabatier, Paul A. 1998. “The Advocacy Coalition Framework: Revisions and Relevance for Europe.” *Journal of European Public Policy* 5 (1): 98–130.
- Saracco, Fabio, Riccardo Di Clemente, Andrea Gabrielli, and Tiziano Squartini. 2015. “Randomizing Bipartite Networks: The Case of the World Trade Web.” *Scientific Reports* 5 (1): 10595.
- Saracco, Fabio, Mika J Straka, Riccardo Di Clemente, Andrea Gabrielli, Guido Caldarelli, and Tiziano Squartini. 2017. “Inferring Monopartite Projections of Bipartite Networks: An Entropy-Based Approach.” *New Journal of Physics* 19 (5): 053022.
- Saxenian, AnnaLee. 2008. “The New Argonauts: Regional Advantage in a Global Economy.” *Economic Geography* 84 (1): 105–8.
- Sebestyén, Tamás, Erik Braun, and Zoltán Elekes. 2024. *Resolving the Complexity Puzzle: Economic Complexity and Positions in Global Value Chains Jointly Explain Economic Development*. Utrecht University, Human Geography; Planning.
- Simmie, James, and Ron Martin. 2010. “The Economic Resilience of Regions: Towards an Evolutionary Approach.” *Cambridge Journal of Regions, Economy and Society* 3 (1): 27–43.
- Stam, Erik, and Friederike Welter. 2020. “Geographical Contexts of Entrepreneurship: Spaces, Places and Entrepreneurial Agency.” In *The Psychology of Entrepreneurship*, 263–81. Routledge.
- Storper, Michael. 1995. “The Resurgence of Regional Economies, Ten Years Later: The Region as a Nexus of Untraded Interdependencies.” *European Urban and Regional Studies* 2 (3): 191–221.
- Straccamore, Matteo, Matteo Bruno, and Andrea Tacchella. 2025. “Comparative Analysis of Technological Fitness and Coherence at Different Geographical Scales.” *PLoS One*

- 20 (8): e0329746.
- Stuck, Jérôme, Tom Broekel, and Javier Revilla Diez. 2016. “Network Structures in Regional Innovation Systems.” *European Planning Studies* 24 (3): 423–42.
- Szerb, Laszlo, Zoltán Ács, Erkkó Autio, Raquel Ortega-Argiles, and Eva Komlosi. 2013. “The Regional Entrepreneurship and Development Index—Measuring Regional Entrepreneurship.”
- Tacchella, Andrea, Andrea Zaccaria, Marco Miccheli, and Luciano Pietronero. 2023. “Relatedness in the Era of Machine Learning.” *Chaos, Solitons & Fractals* 176: 114071.
- Teece, David J, Richard Rumelt, Giovanni Dosi, and Sidney Winter. 1994. “Understanding Corporate Coherence: Theory and Evidence.” *Journal of Economic Behavior & Organization* 23 (1): 1–30.
- Tóth, Gergő, Zoltán Elekes, Adam Whittle, Changjun Lee, and Dieter F Kogler. 2022. “Technology Network Structure Conditions the Economic Resilience of Regions.” *Economic Geography* 98 (4): 355–78.
- Varga, Attila, Tamás Sebestyén, Norbert Szabó, and László Szerb. 2020. “Estimating the Economic Impacts of Knowledge Network and Entrepreneurship Development in Smart Specialization Policy.” *Regional Studies*.
- Visser, Evert-Jan, and Ron Boschma. 2004. “Learning in Districts: Novelty and Lock-in in a Regional Context.” *European Planning Studies* 12 (6): 793–808.
- Wade, Robert. 2018. “Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization.”
- Wanzenböck, Iris, Silvia Rocchetta, Keungoui Kim, and Dieter F Kogler. 2025. “The Emergence of New Regional Technological Specialisations: Exploring the Role of Organisations and Their Technological Network Structure.” *Industry and Innovation* 32 (5): 513–39.
- Whittle, Adam, and Dieter F Kogler. 2020. “Related to What? Reviewing the Literature on Technological Relatedness: Where We Are Now and Where Can We Go?” *Papers in Regional Science* 99 (1): 97–114.
- Wright, Marvin N, and Andreas Ziegler. 2017. “Ranger: A Fast Implementation of Random Forests for High Dimensional Data in c++ and r.” *Journal of Statistical Software* 77:

References

1–17.

- Xiao, Jing, Ron Boschma, and Martin Andersson. 2018. “Resilience in the European Union: The Effect of the 2008 Crisis on the Ability of Regions in Europe to Develop New Industrial Specializations.” *Industrial and Corporate Change* 27 (1): 15–47.
- Zaccaria, Andrea, Saurabh Mishra, Masud Z Cader, and Luciano Pietronero. 2018. “Integrating Services in the Economic Fitness Approach.” *World Bank Policy Research Working Paper*, no. 8485.
- Zhu, Shengjun, Qi Guo, and Canfei He. 2021. “Strong Links and Weak Links: How Do Unrelated Industries Survive in an Unfriendly Environment?” *Economic Geography* 97 (1): 66–88.