THESIS

BENNOUR MOHAMED HSIN

2025-12-03

Table of contents

List of Figures

List of Tables

Preface

This is a Quarto book.

To learn more about Quarto books visit https://quarto.org/docs/books.

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 exiting 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 fabric is what facilitate 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 detrimental 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 detrimental 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. Albeit these simplifications provide valuable scope for analysis and interpretation, the cost from the loss of granular information 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 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 granular data and understand the contexts and contingencies that drives regional diversification.

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 about 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 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 detrimental 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 since path dependency of related variety is a direct result of how relatedness and relatedness density is calculated and interpreted. The constructs that enable these measures (co-occurrence and linear aggregation) are the core problem that we're highlighting here. The reason behind this specific focus relied on the implicit assumptions embedded in these methodological constructs.

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 frequency are misinterpreted, accrues 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.

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 the methodological gaps in REC, we propose a multilayered contextual framework that distinguishes between different contextual factors. Relatedness measures alone conflate these dimensions. By decomposing context into endogenous technology-region fit and exogenous environmental conditions, we can identify which combinations enable successful diversification across different regional contexts, following this design:
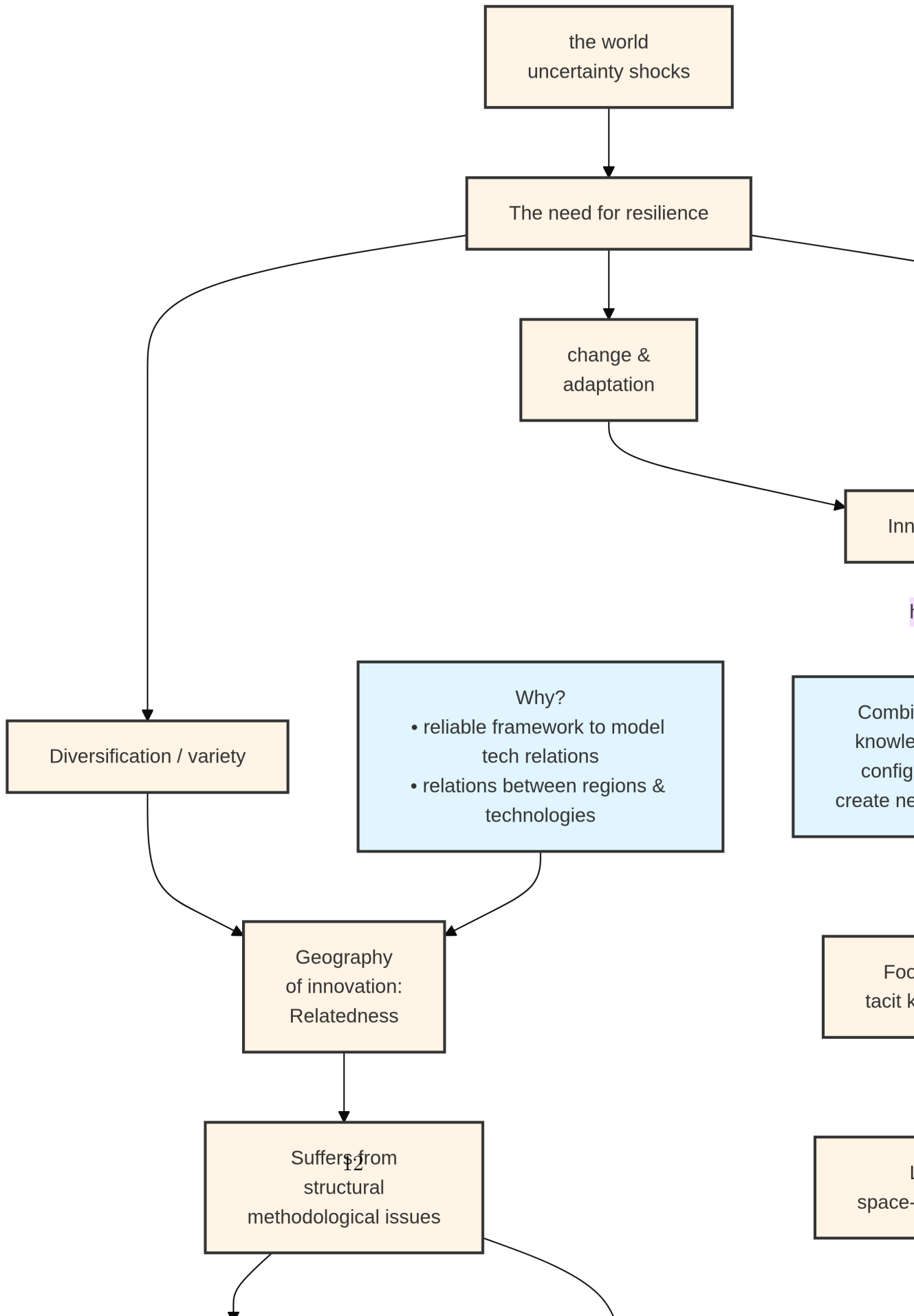
Endogenous factors (within-region characteristics):

- Technology-specific attributes (network position, centrality, embeddedness)
- Regional knowledge infrastructure (coherence between technology relationships and regional specialization patterns)
- Regional technological volume (number of patents in a specific technology)

Exogenous factors (beyond-region influences):

- National entrepreneurial ecosystem characteristics (institutions, financing, R&D culture)
- Spatial spillovers from neighbouring regions (outcome diffusion, knowledge flows)
- Geographic proximity to coherent knowledge clusters

The interaction between these layers is critical: endogenous capacity determines what a region can develop, while exogenous context shapes realisation possibilities. For instance, a technology may align with regional knowledge infrastructure (endogenous coherence), but without supportive national institutions or proximity to innovative neighbours (exogenous support), entry probability remains low.

the world
uncertainty shocks

The need for resilience

change &
adaptation

Inn

Why?
• reliable framework to model
tech relations
• relations between regions &
technologies

Combi
knowle
config
create ne

Diversification / variety

Geography
of innovation:
Relatedness

Foc
tacit k

Suffers from
structural
methodological issues

space-

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. With such an approach we end up accounting for more contextual layers than the mainstream approach.

1.4 Research questions and hypothesis

RQ1: How to further contextualise diversification strategy based on the relatedness framework?

- H1a: Asymmetric relatedness measures better predict technology entry than symmetric measures.

RQ2: Is successful diversification contingent on regional knowledge infrastructure?

- H2a: Technology-specific characteristics impact on diversification is contingent on the regional knowledge coherence.

- H2b: Technology-specific characteristics impact on diversification is contingent on the regional knowledge stock.

RQ3: Does the national ecosystem influence diversification?

- H3a: National Entrepreneurial Ecosystem characteristics positively affects regional entry into new technologies.

RQ4: What role does space have? Do neighbouring regions influence diversification?

- H4a: Spatial spillovers of outcome from neighbouring regions positively affect technology entry.

- H4b: Neighbouring regions' technological characteristics influence focal region diversification outcomes.
- H4c: Geographic proximity to regions with coherent knowledge infrastructure increases entry probability.

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 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 same chapter we also detail other elements that will be relevant to the empirical part specifically knowledge coherence. 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.

1.6 Relatedness(?)