The Production and Diffusion of Knowledge: Essays on Science, Technology, and Proximity

Doctoral Thesis
by
Giorgio Tripodi

A dissertation submitted in partial fulfillment of the requirements for the Doctoral Program in Data Science
XXXIII cycle

May 2022
Faculty of Sciences, Scuola Normale Superiore

Supervisors:

Prof. Francesca Chiaromonte

Sant'Anna School of Advanced Studies

Pennsylvania State University

Prof. Fabrizio Lillo

Scuola Normale Superiore

Bologna University

Abstract

The mainstream recipe for long-run economic prosperity rests upon generating novel ideas and turning them into marketplace innovations. Such innovations increase productivity, and an increase in productivity begets economic growth. Nonetheless, uncovering the roots, mechanisms, and consequences of the production and diffusion of ideas remains a crucial challenge. Ideas and knowledge are "public goods" that markets fail to deal with, causing distortions in the level and direction of research efforts. For instance, insufficient resources support fundamental research on crucial issues such as climate change. Fortunately, knowledge flows following patterns that can now be tracked by exploiting big data, combined with network and data science tools. Accordingly, this dissertation takes a data-driven perspective to provide new evidence on how social, technological, and geographical proximity affects the production and diffusion of knowledge.

The first essay provides evidence that centrality in the inter-sectoral knowledge space positively affects the competitiveness of industries, but only national knowledge flows have a significant impact. We shift the emphasis of the analysis from the bilateral knowledge flows that characterize industrial relationships to the position of an industry in the entire inter-sectoral knowledge space. We collect patent data on 14 manufacturing industries in 16 OECD countries over the period 1995–2009 to track down inter-sectoral flows, and then we investigate whether the relative position of an industry affects its international competitiveness. The analysis suggests that centrality in the inter-sectoral knowledge space positively affects industries' export market shares. Furthermore, national-level knowledge flows' impacts on international competitiveness are stronger than international ones. Industries that can intercept knowledge flows outperform their foreign counterparts. Interestingly, this is true as far as national flows are concerned: geographical boundaries still limit the transmission of tacit knowledge.

Next, the second essay shows that knowledge and social proximity drive scientists' research portfolio diversification, and social relationships become crucial when scientists move far from their core specialization. By looking at the research output of roughly 200k physicists and using bipartite networks, we derive a measure of topic similarity and a measure of social proximity to investigate to what extent knowledge and social proximity shape scientists' research portfolio diversification. We find that scientists' strategies are not random and significantly affected by both measures. However, social relatedness stands out in explaining research diversification, suggesting that science is an eminently social enterprise. In addition, a significant negative interaction between knowledge and social relatedness signals that the farther scientists move from their specialization, the more they rely on collaborations.

Finally, the third essay aims at quantifying knowledge spillovers stemming from research efforts in "Negative Emissions Technologies" (NETs), deemed as one of the leading potential solutions to tackle global warming. As of today, however, NETs hardly represent fully developed technologies to be deployed at scale. By looking at scientific articles, patents, and policy documents, we quantify the impact of NETs within and beyond the scientific realm. Our results suggest that knowledge spillovers are non-negligible for NETs research. Yet, the impact of different NETs varies greatly within science, and Direct Air Capture (DAC) is the option that generates more impact beyond the academic world (measured by citations coming from patents). Finally, we also apply network analysis to identify research hubs that can support future collaborations.

Science and technology policy will play a crucial role in shaping our response to crises and societal challenges, such as global health issues or climate change. The essays collected in this work contribute to the literature by offering novel insights into how scientific and technical knowledge flows across our economies and societies, including policy-relevant recommendations.

Keywords:
Science of Science; Networks; Innovation; Climate change; Knowledge Flows

Publications & works in progress

1. Lamperti, F., Malerba, F., Mavilia, R., and Tripodi, G. Does the position in the inter-sectoral knowledge space affect the international competitiveness of industries? *Economics of Innovation and New Technology*, 29(5):441-488, 2020. Journal link

2. Tripodi, G., Chiaromonte, F., and Lillo, F. Knowledge and social relatedness shape research portfolio diversification. *Scientific Reports*, 10(1), 2020. Journal link

3. Tripodi, G., Lamperti, F., Mavilia, R., Mina, A., Chiaromonte, F., and Lillo, F. Quantifying knowledge spillovers from advances in negative emissions research. Pre-print link.

4. Tripodi, G., Dehmamy, N., Wang, D. Specialized-birds: coherent exploration of the knowledge space is associated with greater scientific productivity and impact. *In preparation.*

5. Tripodi, G. et *al.* Localized knowledge spillovers in climate science and technology. *In progress.*

In economics journals authors are usually listed in alphabetical order, irrespective of contribution.

Conferences, talks, and summer schools

1. **NetSci** – Conference of the Network Science Society.
   Shanghai, China, 2022 (Forthcoming - virtual event)
2. **IAERE** – The Italian Association of Environmental and Resource Economists
   Cagliari, Italy, 2022
3. **Young Researchers Informal Meeting** – EmbeDS
   Pisa, Italy, 2022
4. **CSSI** – Kellogg School of Management
   Evanston, IL, 2022
5. **CCS** – Conference on Complex Systems.
   Lyon, France, 2021
6. **NetSci** – Conference of the Network Science Society.
   Rome, Italy, 2020 (virtual event)
7. **IC^{2}S^2** – The International Conference on Computational Social Science.
   Boston, MA, 2021
8. **Data Science Colloquium**
   Pisa, Italy, 2021 (virtual event)
9. **CCS** – Conference on Complex Systems.
   Singapore, Singapore, 2019
10. **Data Science Summer School**
    Pisa, Italy, 2019
11. **IC^{2}S^2** – The International Conference on Computational Social Science.
    Amsterdam, Netherlands, 2019
12. **Data Science Colloquium**
    Pisa, Italy, 2019
13. **CCS** – Conference on Complex Systems.
    Thessaloniki, Greece, 2018
14. **Lipari School on Computational Complex and Social Systems**
    Lipari, Italy, 2018
15. **Oxford Summer School on Economic Networks**
    Oxford, UK, 2018

“Not a dog. Not a wolf. All he knows is what he’s not.”

Acknowledgments

As most people mentioned here know far too well, I am a genuine pessimist. Getting through this incredible journey was beyond my expectations, and I could not make it on my own. Indeed, this thesis would not have been possible without the help and support of a considerable number of people.

First and foremost, I would like to thank my advisors: Francesca and Fabrizio. They have been an invaluable source of inspiration and an extraordinary scientific guide. Working with them has been a unique opportunity for me to learn how to navigate through cutting-edge interdisciplinary research. I have done my best to follow their lead.

Next, I would like to thank Dino, the architect and coordinator of the Ph.D. program in Data Science, for his constant support. I owe Dino a lot, both from a scientific and personal point of view.

In addition, I would like to thank Dashun, Nima, and all the CSSI group members at the Kellogg School of Management that hosted me during my visiting period. Despite the advent of Covid-19, the months that I have spent in the US have been a tremendous learning opportunity.

Likewise, I want to thank Franco, Roberto, and the ICRIOS affiliates at Bocconi Univerity, that ushered me into scientific research.

Further, I would like to thank Andrea, EmbeDS affiliates, and all people at the institute of economics at SSSA for their valuable scientific and personal support over the last two years.

Pursuing a Ph.D. means wandering through a winding road for years. It was tough indeed, but the company was so great. If I have survived, it is because I had the chance to share this journey with a group of friends (and colleagues) that turned all these years into a memorable experience.

First of all, let me mention all my Data Science pals: Luca, Cecilia, Gevorg, Jisu, Vassia, Tommaso, and Elisa. No Ph.D. cohort has ever been better!

However, my family at Scuola Normale was much larger, and I want to thank, among others: Gianmichele, Anna, Ale, Farzard, Marco, Vasco, Donato, Giulio, Aurelia, Stefano, Bibek, Giovanni, and Marcello.

After the first years, I had the fortune of being adopted by a new family at Scuola Sant'Anna. I am genuinely grateful to all the people who gave me shelter in Piccionaia. In particular, I would like to thank Gian, Matteo, Fede, Mario, and Marco.

I owe a massive debt to Francesco: a true friend, a scientific mentor, and an occasional therapist. If I am here today, it's also because of his sincere support.

It's too hard to pick the right words to thank Paola. I moved to Pisa to pursue a Ph.D. with great aspirations and some doubts. But then I met her and suddenly moving to Pisa started to look like one of the best choices of my life.

Last but not least, I want to thank all the people who bear with me for so many years. They know who they are; they have always been on my side – my parents on top of everyone else. I hope one day I can make them proud.

Contents

List of Figures

List of Tables

> "As long as scientists are free to pursue the truth wherever it may lead, there will be a flow of new scientific knowledge to those who can apply it to practical problems."
>
> Vannevar Bush (1945)

Chapter 1

Introduction

Innovation is the primary driver of long-run economic prosperity. Economists hardly agree on anything, but they all agree on the staple role of knowledge and innovation in fostering growth. The mechanism through which innovation affects growth is pretty straightforward in theory: novel ideas turned into practical innovations increase productivity, and a higher level of productivity fosters economic growth. In other words, as of today, the ingredients of long-run growth are known, but the exact recipe is still missing. The main reason behind this uncertainty is that knowledge is not a common good (Nelson, 1959; Arrow, 1962). In short, two features characterize goods: *rivalry* and *excludability*. Rivalry captures whether the value/availability of the good decreases as more people consume it. On the other hand, excludability elucidates whether the availability of the good can be restricted.

Knowledge is non-rival and, arguably, only partially excludable. Therefore, producing new knowledge generates spillovers, constituting a positive *externality*. Positive externalities occur when producing a specific good benefits third-party economic agents. The case of knowledge is quite intuitive: once a piece of valuable knowledge has been produced, anyone can reuse it with nearly no constraints. For instance, learning and applying the Pythagorean theorem does not reduce its value for anyone else. Simply put, we can define knowledge spillovers as flows of useful information that are not directly mediated by a market transaction. So, they might be considered a favorable occurrence, at least in principle. Unfortunately, however, they are the root cause of a daunting market failure: under-investment in science and innovation. If the benefits of working on novel ideas can not be entirely appropriated, private organizations might have less incentive to undertake risky research projects in the first place[1].

Furthermore, there are signals that research productivity is slowing down, and one possible explanation is that new ideas are getting harder and harder to find (Jones, 2009; Bloom et al., 2020). Markets are not efficient in fixing the rate and direction of innovation (Stiglitz and Greenwald, 2014). As a result, despite a sustained increase in university research, productivity growth has stagnated over the past years in many advanced economies. Too often, university and corporate research differ in nature and scope, hampering the transmission of scientific knowledge towards productivity-enhancing solutions (Arora et al., 2020). To put it simply, finding genuinely innovative ideas is increasingly costly, and corporate labs shift away from basic research. All told, globalization is crucially affecting the map of economic and scientific activities. Nevertheless, proximity – in multiple dimensions – still plays a leading role in our allegedly borderless economy, particularly for what concerns innovative and creative endeavors. (Breschi and Lissoni, 2001; Breschi and Malerba, 2005; Carlino and Kerr, 2015; Balland et al., 2020).

Consequently, the following questions arise rather naturally: how can we uncover the nature of knowledge production and diffusion? What are the reach and boundaries of knowledge flows, and how can we quantitatively measure them? To answer the questions above, we need an alternative – data-driven approach. In 1992, the Nobel laureate Paul Krugman argued that: "knowledge flows are invisible; they leave no paper trail by which they may be measured and tracked,..."(Krugman, 1992). Nowadays, this is not entirely true. Indeed, thanks to data and network science tools, coupled with the increasing availability of digital traces of innovation activities, we can now shed some light on the production and diffusion of knowledge (Fortunato et al., 2018b; Wang and Barabási, 2021). In a nutshell, this is the essential purpose of the dissertation. In line with this view, the following chapters provide new evidence on how social, technological and geographical proximity

[1]See Summers and Jones (2020) for a recent and rigorous treatment of the private vs. social returns of innovation investments.

shape the production and diffusion of knowledge, taking a computational social science perspective.

1.1 Summary of main contributions

This section summarizes the main contributions of the thesis [2]. Chapter 2 (see Panel I for a graphical summary) investigates how inter-sectoral knowledge flows affect the international competitiveness of industries. The relationship between technology and the international competitiveness of industries has always been a central topic for both academic inquiry and economic policy. For simplicity, we can consider industries' international competitiveness as a function of cost and technological factors. Moreover, among technological factors, we can distinguish between innovative activity per se (e.g., knowledge production, patent stocks) and the diffusion of advanced knowledge (i.e., national and international knowledge spillovers). In our empirical investigation, we focus on the knowledge diffusion channel by shifting the emphasis of the analysis from the bilateral knowledge flows that characterize industrial relationships to the position of an industry in the entire inter-sectoral knowledge space. In more detail, we collect patent data on 14 manufacturing industries in 16 OECD countries over the period 1995–2009, and assess whether the relative position of an industry affects its international competitiveness. We derive national and international technological networks using co-occurrences in patent classes and citation data. Then, we test two hypotheses; namely: (i) that being central in the flow of information positively affects international competitiveness, and (ii) that the relevant scale for making effective use of information is the national level. To capture the position of industries in our knowledge space, we pick two relatively simple network measures: the strength (i.e., weighted degree centrality) and the local clustering coefficient. Given the persistency in the export market shares, in our primary empirical strategy, we use a GMM estimator in a dynamic panel setting to check whether centrality at the national and international level mat-

[2] I will refer to findings and analyses using plural first-person pronouns throughout the thesis. Science is increasingly a team endeavor, and the articles this dissertation is based upon have been co-authored.

ters for competitiveness. Our results suggest that (i) centrality and local clustering in the inter-sectoral knowledge space positively affect the export market shares of the industry; (ii) these two effects are somewhat redundant (i.e., being central in a knowledge space is far less relevant when the industry is already embedded within a cluster); and (iii) national-level knowledge flows affect international competitiveness much more than international ones. In short, our analysis suggests that the position in the inter-sectoral knowledge space is more relevant than standard technological indicators in explaining export market shares. Further, industries that can intercept knowledge flows outperform their foreign counterparts. However, this is true mostly as far as national flows are concerned: geographical boundaries still limit the transmission of tacit knowledge.

Chapter 3 (see Panel II for a visual summary of the results) investigates how scientists move and interact in knowledge space. Scientists, likewise innovators, may work on several topics during their careers, and their choices are driven by a variety of factors. Broadly speaking, there are at least two general mechanisms that may affect scientists' behavior: the first is the well-known trade-off between exploration and exploitation. The second is the so-called burden of knowledge - which forces scientists to seek a narrower specialization and rely on collaborations. Despite recent efforts to better clarify the mechanisms that affect scientists' diversification strategies, the empirical evidence is still puzzling. On the one hand, we know that most scientists change their research interests only slightly over time. Exploration characterizes their career, but mostly within their area of expertise. On the other hand, there are signals of an increasing trend for scientists to switch topics. The purpose of this chapter is to disentangle and quantify the role of different contributing factors, with a focus on knowledge and social relatedness. To do so, we reconstruct the careers of approximately two hundred thousand physicists relying on the hierarchical structure of PACS classification codes (Physics and Astronomy Classification Scheme). Then, using bipartite networks, we derive a measure of topic similarity and one of social proximity to investigate to what extent knowledge and social proximity shape scientists' research portfolio diversification. We model the probability that a scientist i, specialized in sub-field a, is also active in some sub-field b different from her own specialization as a function of the knowledge relatedness between the two fields, the social relatedness between the author and authors in the target sub-field, and including several control variables. In practice, to estimate the model, we need three ingredients: the core specialization of scientists (captured using the Revealed Comparative Advantage - RCA), a measure of knowledge relatedness among physics subjects (measured using the cosine similarity among PACS codes), and an indicator of social relatedness (measured as a dummy that captures whether a given scientist can reach a certain sub-fields through direct social interactions). We employ a logistic regression, and we find that scientists' strategies are not random and are significantly affected by both measures. Furthermore, a significant negative interaction between knowledge and social relatedness suggests that the farther scientists move from their specialization, the more they rely on collaborations. To quantify the relative role of various factors, we first run a group LASSO algorithm to track how different predictors are excluded or included in the model as the regularization penalty varies. Then, we compute the relative contribution to deviance explained - that is, the percentage of the logistic regression deviance that is captured by each predictor. Interestingly, social relatedness stands out in explaining research diversification, suggesting that science is an eminently social enterprise.

As mentioned before, too few resources are allocated to crucial issues such as climate change. Worse, some innovations hardly leave academia to spread across our societies. Chapter 4 (see Panel III for a minimalist storyboard) focuses on the multidimensional nature of knowledge spillovers generated by Negative Emissions Technologies (NETs), deemed one of the leading potential solutions to global warming. NETs pursue the removal of CO2 from the atmosphere through technical means; the latest (April 2022) report of the Intergovernmental Panel on Climate Change (IPCC) claims that without employing NETs, it will be impossible to stay within the 1.5 (or even 2) degrees temperature change threshold. Unfortunately, to date, NETs do not yet represent fully developed technologies ready to be deployed at scale; a lot of research and innovation are still needed to turn these technologies into practical options. Against this backdrop, we use an innovation network perspective or, more precisely, a science of science perspective, to quantify knowledge spillovers from NETs scientific advances within and beyond the academic realm. Because tackling climate change requires basic scientific research, practical innovation, and political support, we move beyond standard citation counts in our analysis. In addition to collecting all NET-related articles from the Web of Science (WoS) using keywords in titles and abstracts, we retrieve information from patents and policy documents. Focusing on eight major NET domains, we evaluate knowledge spillovers using article citation counts for the science dimensions, and citations in patents and policy documents for the technology and policy dimensions. Our statistical analyses employ Generalized Linear Models and we ensure stability constructing 30 suitable control groups with an exact matching procedure, and re-estimate our models with each. Our results suggest that NETs-related research generates significant, positive knowledge spillovers within science and from science to technology and policy. However, significant differences exist across carbon removal solutions. For instance, Direct Air Capture (DAC) is the only option linked to practical technological innovation, while Bio-energy with Carbon Capture and Storage (BECCS) appear to lag behind. Interestingly, policymakers tend to overlook advances in DAC, focusing on solutions such as BECCS and Blue Carbon (BC). Finally, we also study the spatial distribution of NET-related research with network analysis tools, identifying cities and countries that can serve as research hubs for supporting future collaborations.

Science and technology policy will play a crucial role in shaping our response to crises and societal challenges, such as global health issues or climate change. The essays collected in this thesis (Chapters 2,3,4) contribute novel insights into how scientific and technical knowledge flows across our economies and societies, including policy-relevant recommendations. We end with final remarks in Chapter 5.

1.2 Data and schematic outline

This Section briefly describes the main data sources used in the thesis, most of which are freely available patent and publication data. Figure 1.1 provides a schematic summary highlighting data sources and key concepts used in each chapter. This stylized guide can help the reader navigate the different parts of this dissertation by pointing out what kind of data, proximity measure, methodological approach, and level of aggregation characterizes each empirical analysis.

Data sources

PATSTAT: The European Patent Office (EPO) releases PATSTAT Global, which contains bibliographical data related to more than 100 million patent documents from leading industrialized and developing countries.

APS: The American Physical Society, upon specific request, provides data concerning scientific publications – for use in research about networks and the social aspects of science.

WoS: Web of Science is an extensive global citations database comprising information on millions of research articles. It is maintained by the private company Clarivate.

RoS: Reliance on Science is a publicly available database that includes citations from both patents and scientific articles (Marx and Fuegi, 2020).

Altmetric: Altmetric is a curated database that collects metrics complementary to standard citation-based data, such as mentions on a diverse set of outlets. Altmetric data is freely available for scientific purposes upon request.

STAN: The STructural ANalysis Database is a comprehensive OECD database including a broad range of information related to productivity growth, competitiveness and general structural change at the industry level.

Replicability

All data and code to replicate the main plots of this thesis are available upon request. Code and functions for regression and other statistical analyses/tests (e.g., for data cleaning and visualizations) are based on Python and R packages. More details are provided within each chapter.

Additional preliminary notes

Parts of this thesis (i.e., Chapters 2,3,4), consist of self-contained essays based on the following articles and working papers, with minor changes:

* Lamperti, F., Malerba, F., Mavilia, R., and Tripodi, G. Does the position in the inter-sectoral knowledge space affect the international competitiveness of industries? *Economics of Innovation and New Technology*, 29(5):441-488, 2020. Journal link

* Tripodi, G., Chiaromonte, F., and Lillo, F. Knowledge and social relatedness shape research portfolio diversification. *Scientific Reports*, 10(1), 2020. Journal link

* Tripodi, G., Lamperti, F., Mavilia, R., Mina, A., Chiaromonte, F., and Lillo, F. Quantifying knowledge spillovers from advances in negative emissions research. Pre-print link.

The contents and style of the works presented here reflect the interdisciplinary environment of the Ph.D. program in Data Science – a joint doctoral program that involved the Scuola Normale Superiore, the Scuola Superiore Sant'Anna, the Italian National Research Council (CNR), the University of Pisa, and the IMT School for Advanced Studies of Lucca. Further, these works greatly benefited from my research fellowship at Scuola Superiore Sant'Anna, my collaborations at Bocconi University, and my visiting research period at the Center for Science of Science and Innovation (Kellogg School of Management, Northwestern University).

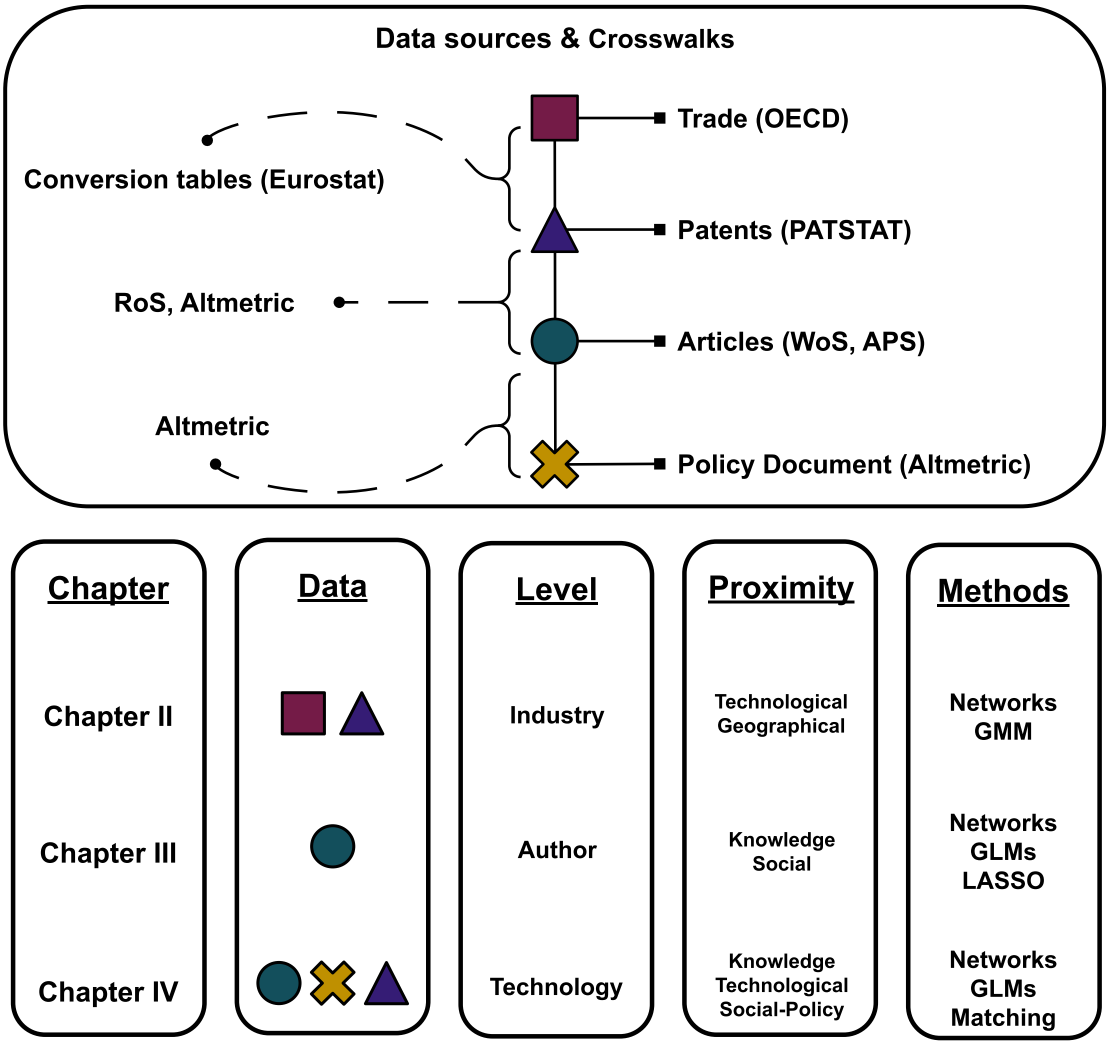


Figure 1.1: Schematic outline.

Chapter II investigates how knowledge flows affect international competitiveness at the **industry** level. It relies on **patent** and **trade** data and employs two central notions of proximity: **technological** and **geographical**. From a methodological standpoint, the analysis is based on patents' **networks** in a dynamic panel setting, where the Generalized Method of Moments (**GMM**) estimator ensures consistent estimates.

Chapter III investigates how **social** and **knowledge** proximity shape individual **authors**' research portfolios. With a focus on academic publications in physics (**APS**), the primary empirical strategy is based on bipartite **networks** and generalized linear models (**GLMs**), with the addition of a feature selection algorithm (**LASSO**) to capture the relative contribution of each potential factor.

Chapter IV combines information about scientific **articles**, **patents** and **policy documents** to quantify **knowledge**, **technological** and **policy** spillovers generated by the emergent climate **technologies** defined as Negative Emissions Technologies (NETs). The empirical framework relies on citation **networks**, an exact **matching** procedure and **GLMs**. Network analysis is used to investigate the geographical dimensions of NETs research activities.

Giorgio Tripodi [3,4] Francesco Lamperti [1,3] Franco Malerba [2] Roberto Mavilia [2,3]

[1] Scuola Superiore Sant'Anna, [2]Bocconi University, [3]MEDAlics, [4]Scuola Normale Superiore

In One Sentence

Significance The relationship between technology and the competitiveness of industries is central to academic research as well as economic policy.

Network approach We shift the emphasis of the analysis from the flows of knowledge related to bilateral industrial relationships to the position of an industry in the entire inter-sectoral knowledge space.

Aim To investigate whether the relative position of industries in the transmission of technical knowledge affects international competitiveness, beyond the role played by sector's degree of innovativeness and controlling for cost-related factors.

Background

- International competitiveness Y is a function of technological T and cost variables C:

$$Y_{ij} = f(T_{ij},\ C_{ij}),\ \text{with}\ \begin{cases} i & \text{stands for } \mathit{Sector} \\ j & \text{stands for } \mathit{Country} \end{cases}$$

- Technological factors include:
  - Innovative activity
  - **Diffusion of advanced knowledge**
- Network of inter-sectoral flows of knowledge:
  - **National** flows $\longrightarrow$ co-occurrences of technological classes in patents
  - **International** flows $\longrightarrow$ patents citations

Our Propositions

Position: industries more central in the inter-sectoral knowledge space perform at the international level better than industries that are not central

Geographical boundaries: the position of an industry in the inter-sectoral knowledge space is more relevant at national level than at the international level

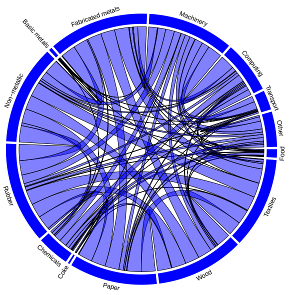
Data

- Patent data on 14 manufacturing industries in 16 OECD countries over the period 1995-2009 (PATSTAT)
- Competitiveness and key statistics at the industry level (STAN)

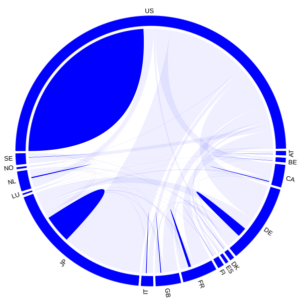
Table: Variables

Var. Name	Description	Data Source
XMS	Country's exports in the industry over the total industry's export	OECD-STAN
INV	Ratio between industry expenditures on gross fixed capital formation and value added (current prices)	OECD-STAN
WAGE	Labour cost per employee	OECD-STAN
POP	Total population	OECD-STAN
PATSH	Share of national industry patents applications over the sum of the industry's patents applications	CRIOS-PatStat
d.w	Degree centrality (technological class co-occurrence network)	CRIOS-PatStat
ev	Eigenvector centrality (technological class co-occurrence network)	CRIOS-PatStat
am	Local clustering (technological class co-occurrence network)	CRIOS-PatStat
d.w.cit	Degree centrality (citation network)	CRIOS-PatStat
ev.cit	Eigenvector centrality (citation network)	CRIOS-PatStat
am.cit	Local clustering (citation network)	CRIOS-PatStat

Methodology



Industry co-occurrence network - Italy 2009



Country citation network 2009

To account for persistency in industries' export performances, we rely on a GMM estimator in a dynamic panel setting:

$$\begin{aligned} XMS_{ijt} = {} & \alpha_0 + \gamma XMS_{ijt-1} + \alpha_1 PATSH_{ijt} + \\ & + \alpha_2 WAGE_{ijt} + \alpha_3 INV_{ij} + \\ & + \alpha_4 POP_{ijt} + \beta_1 d(w)_{ijt} + \beta_2 am_{ijt} + \\ & + \beta_3 d.cit(w)_{ijt} + \beta_4 am.cit_{ijt} + \\ & + \eta_{1i} + \eta_{2j} + \eta_{3t} + \epsilon_{ijt} \end{aligned}$$

Results

	Dependent variable: XMS			
	Baseline		Robustness	
	(1)	(2)	(I)	(II)
XMS_{-1}	0.956*** (0.027)	0.956*** (0.027)	0.953*** (0.032)	0.954*** (0.031)
XMS_{-2}	−0.028* (0.016)	−0.029* (0.016)	−0.047*** (0.015)	−0.046*** (0.015)
PATSH	−0.002 (0.004)	−0.003 (0.004)	−0.002 (0.005)	−0.004 (0.005)
WAGE	0.0004 (0.001)	0.00005 (0.001)	0.001 (0.001)	0.001 (0.001)
INV	0.001 (0.004)	0.001 (0.004)	0.002 (0.005)	0.001 (0.005)
POP	0.002*** (0.0003)	0.002*** (0.0003)	0.002*** (0.0004)	0.002*** (0.0004)
d.w	**0.005**** **(0.002)**	**0.004**** **(0.002)**		
am	**0.030***** **(0.009)**	**0.029***** **(0.009)**		
d.w.cit		0.00005 (0.0004)		−0.0004 (0.001)
am.cit		−0.033 (0.026)		−0.030 (0.029)
d.w:am	**−0.008***** **(0.003)**	**−0.007**** **(0.003)**		
d.w.cit.control			0.005** (0.002)	0.005** (0.002)
am.cit.control			0.032** (0.012)	0.029** (0.012)
d.w.cit.control:am.cit.control			−0.008** (0.003)	−0.007** (0.003)
Time Dummies	Yes	Yes	Yes	Yes
Observations	2811	2811	2811	2811
AR(order1)	−5.23***	−5.24***	−5.29***	−5.30***
AR(order2)	−1.54	−1.55	−1.01	−1.09
Wald Test (coef.)	7703.89***	8047.81***	6071.79***	6258.51***
Wald Test (int.)	477.00***	486.01***	422.93***	416.08***
Sargan/Hansen (χ^2)	55.02	55.01	12.89	12.94

Note: *p<0.1; **p<0.05; ***p<0.01.

Key Findings

Technological centrality matters, but geography still plays a role: for countries it is indeed important to promote inter-industry collaborations among firms.

- Centrality and local clustering in the inter-sectoral knowledge space positively affect industry competitiveness
- Being central in the inter-sectoral knowledge space is far less relevant when the industry is already embedded within a cluster
- National-level knowledge flows affect international competitiveness much more than international ones

Giorgio Tripodi [1,2] Francesca Chiaromonte [2,4] Fabrizio Lillo [1,3]

[1] Scuola Normale Superiore, [2]Scuola Superiore Sant'Anna, [3]Bologna University, [4]Penn State University

In One Sentence

Significance As the amount of knowledge necessary to reach the scientific frontier increases, for scientists to stay innovative and productive it is crucial to manage their research endeavours wisely.

Network approach We use bipartite networks - PACS-Articles, Authors-Articles and PACS-Authors - respectively to compute a measure of similarity among sub-fields, a measure of social proximity and identify patterns of exploration.

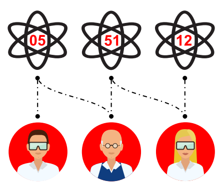
Aim To identify and quantify what are the possible drivers of exploration compared to key factors such as knowledge relatedness among research topics and social relationships among authors.

Background

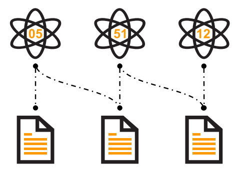
- Scientists and innovators' activities often span several topics and their choices of research endeavours are driven by an amalgam of factors:
  - The **"essential tension"** between exploration and exploitation
  - The **"burden of knowledge"** that increases over time

Data

- Publication history of $\sim$ 200k physicists (APS data)

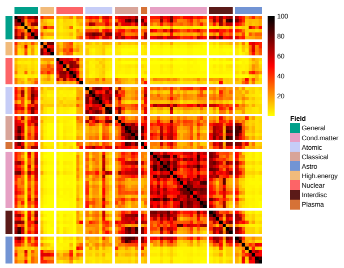


Authors-Articles/PACS

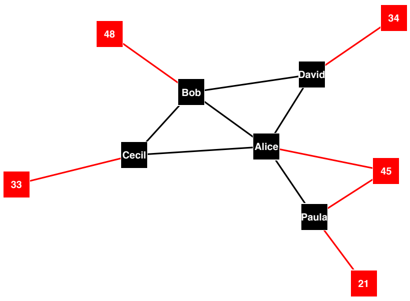


PACS-Articles

Methodology



Knowledge relatedness
(cosine similarity PACS)

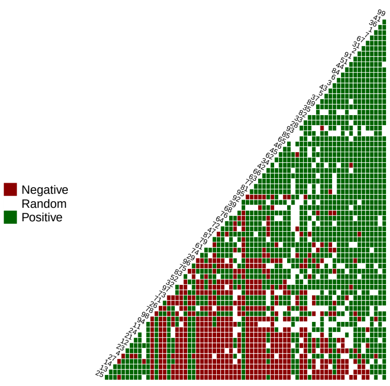


Social relatedness
(augmented co-authorship network)

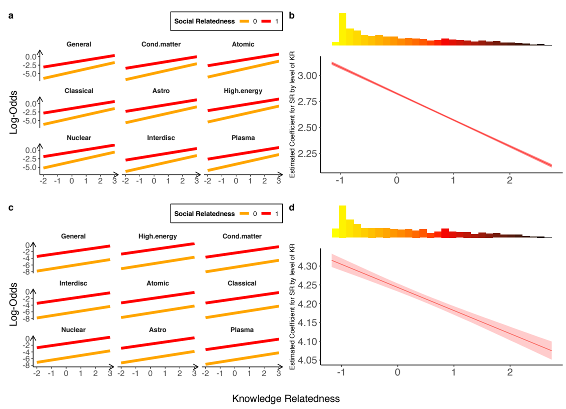
We model the probability that a scientist i, specialized in sub-field a, is also active in some sub-field b different from her own specialization as a function of the knowledge relatedness between the two fields, the social relatedness between the author and the target sub-field, plus a bunch of control variables that include individual factors, sub-fields popularity and competition.

$$Y = ln(\frac{p}{1-p}) = \alpha + \beta KR_{ab} + \gamma SR_{ib} + \zeta(KR_{ab} \times SR_{ib}) + \theta \cdot \mathbf{IF_i} + \eta \cdot \mathbf{SC_b} + \phi \cdot \mathbf{Cit_b}$$

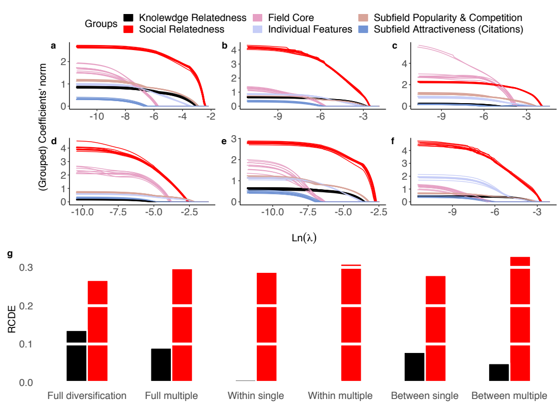
Results



Test of randomness



Scientists' research portfolio diversification



Relative importance of predictors

Key Findings

Science is more and more a social enterprise: the trade-off between exploration and the intellectual cost of diversification resolves trough collaborations.

- Diversification patterns are not random
- Both knowledge and social relatedness shape scientists' diversification
- Social relationships are crucial especially when scientists move far from their primary expertise

Giorgio Tripodi [1,2] Francesco Lamperti [1,2] Roberto Mavilia [5]
Andrea Mina [2] Francesca Chiaromonte [2,4] Fabrizio Lillo [1,3]

[1] Scuola Normale Superiore, [2]Scuola Superiore Sant'Anna, [3]Bologna University, [4]Penn State University, [5]Bocconi University, [6]RFF-CMCC

In One Sentence

Significance Negative Emission Technologies (NETs) represent a key element to reach the net-zero emissions target (or to effectively smooth an undoubtedly needed green transition).

Network approach Given the critical role played by climate-related technologies, we move beyond the standard citation networks to incorporate knowledge flows to practical innovations (i.e., patents) and the public discourse (i.e., policy documents).

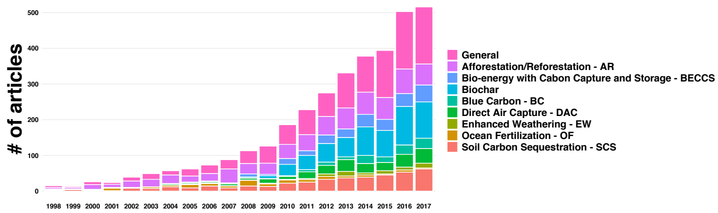
Aim To quantify the multidimensional nature of knowledge spillovers generated by 20 years of research in NETs, and identify cities and countries that can serve as research hubs for supporting future collaborations.

Background

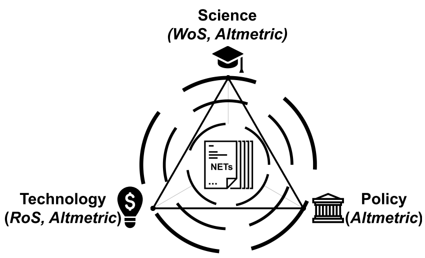
- By Negative Emissions we refer to "...the removal of CO2 from the atmosphere through technical means".
  - NETs are not all alike (e.g., measurement, verification, accounting, costs, and durability of carbon stored) $\longrightarrow$ **No universally superior option**
  - Despite the urgent need, NETs are hardly fully developed technologies ready for large-scale deployment $\longrightarrow$ **Policy support**

Data

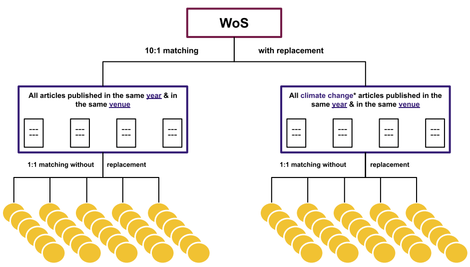
- Articles related to NETs from 1998 to 2017 (Web of Science)
- Citations from patents (Reliance on Science/Altmetric)
- Citations from policy documents (Altmetric)



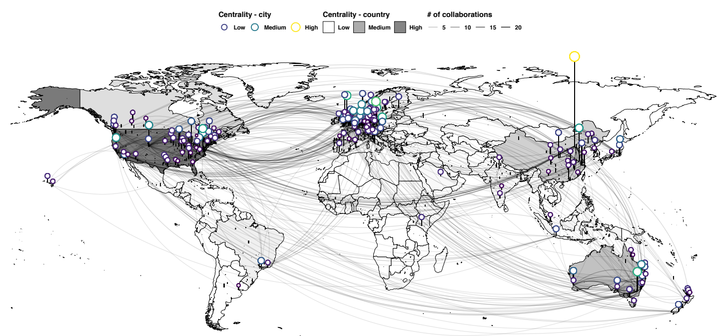
Methodology



Science-Tech-Policy links



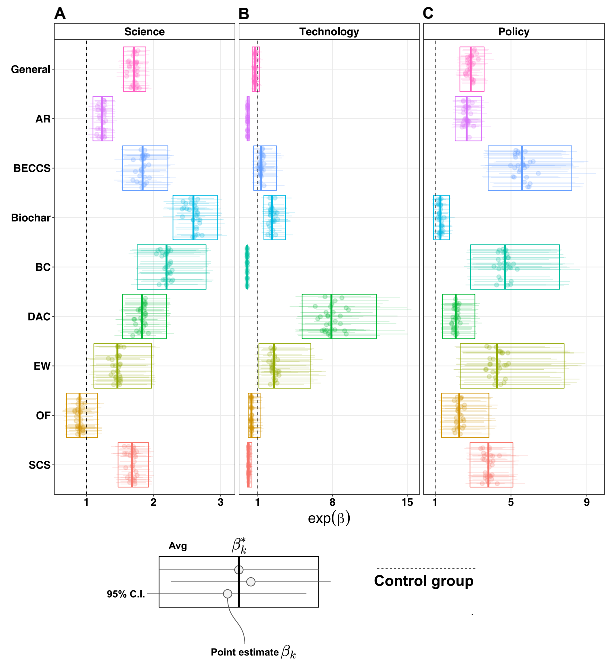
Matching strategy



We rely on econometric methods based on GLMs and exact matching to estimate the size of knowledge spillovers. Our preferred specifications employ negative binomial regressions for citations counts and logistic regressions for citation likelihoods.

$$g(E(S_{ikt}|NET_{ik}, T_{it}, \mathbf{X}_i)) = \alpha + \sum_k \beta_k NET_{ik} + \sum_t \gamma_t T_{it} + \delta \mathbf{X}_i \quad (1)$$

Results



NETs spillovers to science, technology and policy. Coefficients (exponentiated) of the regression models of Eq. (1). Results are obtained by fitting 30 negative binomial regressions (**A**) and 30 logistic regressions (**B–C**) on one-to-one matched samples with year dummies.

(A) Incident Rate Ratio (IRR) for each NET on the number of scientific citations.
(B) Odds Ratio (OR) for each NET on the probability of being cited by a patent.
(C) Odds Ratio (OR) for each NET on the probability of being cited by a policy document.

Key Findings

Knowledge spillovers play a key role, but NETs can only be complementary to more general emissions reduction pathways.

- Negative emissions research is highly heterogeneous and spread across different hubs
- DAC appears to be the most promising as far as practical technological innovations are concerned
- Science, technology, and policy need to be better coordinated to boost the efficacy of research endeavors

Chapter 2

Technology, geography, and trade

This chapter is largely based on the following published article: Lamperti et al. (2020)

2.1 Introduction

The relationship between technology and the international competitiveness of industries - defined as the ability of a given country or industry to compete with its foreign counterparts (Castellacci, 2008) - has been central to academic research as well as economic policy. A large body of both theoretical and empirical literature has investigated the role of technology and technological change in influencing international competitiveness at micro, meso and macro levels (see, for example, Fagerberg, 1988, Amendola et al., 1993, and more recently, Laursen and Meliciani, 2010 and Dosi et al., 2015). Taking an industry-level perspective, the present chapter proposes a complementary view to these contributions by isolating a novel and significant factor that explains the dynamics of competitiveness. We build on the notion of knowledge flows and rely on a network perspective to investigate whether the relative position of an industry within a relevant knowledge space affects the international competitiveness of that industry.

With respect to the emphasis on economies of scale and geography, relative productivity and wages that is typical in the new and new new theories of trade

(Krugman, 1980, 1992; Melitz, 2003), we aim at examining the nature and composition technical knowledge produced within industries. In the trade literature the investigation of technological factors in addition to cost-related ones dates back to the seminal work of Posner (1961), who posits that one of the main sources of (absolute) advantage of a country comes from its relative technological position against its competitors in any one activity. Since then and, particularly, since the second half of the 1980s, the literature has spurred. Following the evolutionary and disequilibrium perspective of Dosi et al. (1990), trade flows have been considered primarily driven by sector-specific absolute advantages, in turn stemming from widespread technological asymmetries between countries, due to differences in the capabilities to produce innovative products (i.e. which other countries are not yet capable of producing, irrespective of relative costs), to develop new process innovations or to use existing processes more efficiently or more rapidly. Along these lines, one may reasonably argue that the ultimate driver of sector specific advantages rests in the technical knowledge behind both product and process innovations (see also Dosi (1988) for a more general discussion). Indeed, following Fagerberg (1996), we can formally specify country-industry competitiveness as a function of both technological and cost factors.

Among technological factors, one may distinguish between innovative activity and the diffusion of advanced knowledge. Both factors have been widely examined in the literature. As far as innovation activity is considered, many have focused on the effects of knowledge production, patent stocks, R&D activities and national innovation systems on the competitiveness of industries and countries (Nelson and Winter, 1977; Freeman et al., 1982; Dosi, 1988; Dosi et al., 1990). With respect to the diffusion of advanced knowledge, while Grossman and Helpman (1991, 1995) have underlined the role of national and international knowledge spillovers, Laursen and Meliciani (2000, 2002) stressed the role of inter-sectoral linkages in affecting trade competitiveness. Our analysis builds on this last group of contributions.

In particular, the purpose of the present chapter is to empirically investigate how technology affects competitiveness not just directly, via the production of technical knowledge, but also indirectly, characterizing an industry's position in the network of inter-sectoral flows of knowledge - which we call the inter-sectoral knowledge space. The core idea, better detailed in the remaining of the chapter, is that the position of industry might allow both the acquisition and the diffusion of relevant pieces of knowledge. In addition, we allow for a dynamic specification tracking how industries change their position in the network of knowledge flows. Our approach considers both national and international relationships among industries and makes use of patent data to identify and quantify links among them. In that, our representation of knowledge flows differs from the stream of research on the role played by the position in product space (Hidalgo et al., 2007; Tacchella et al., 2012), as we directly map technological relationships - using patent data - and their effects on the competitiveness of industries (rather than countries)[1]. We follow Breschi et al. (2003) in the construction of a "national" knowledge network in terms of co-occurrences of all pairs of technological classes included in the patent stock of each country. In addition, to study international flows, we focus on patent citations (Jaffe and Trajtenberg, 2002). Results show that (i) centrality and local clustering in the inter-sectoral knowledge space positively affect the export market shares of an industry of a country, (ii) such two effects are rather redundant, i.e. being central in a knowledge space is far less relevant when the industry is highly connected within a cluster and, finally, (iii) national-level knowledge flows affect international competitiveness much more than international ones do. Actually, the latter are even not significant in boosting export performances.

The chapter is organized as follows. Section 2.2 presents a critical overview of the literature, while Section 2.3 provides a discussion of mechanisms influencing

[1]Network methods have been employed to quantitatively measure the impact of relatendeness on diversification/specialization patterns of countries and regions. Recently, Alshamsi et al. (2018) and Petralia et al. (2017) provided evidence that the probability of diversification in terms of products, research areas and technologies increases with the number of related activities.

international competitiveness and derives two main propositions. Sections 2.4 offers a description of the data and the econometric strategy used in the analysis. Then Section 2.5 summarizes the results and Section 2.6 concludes the chapter.

2.2 The Relationship between Technology and Competitiveness

2.2.1 Technology, Costs and International Competitiveness

When examining international competitiveness, Schumpeterian insights have shifted the focus from cost-related variables towards technological factors. In this vein, following Dosi et al. (1990), a general formulation can be specified as a simple function of technological (T) and cost (C) variables:

$$Y_{ij} = f(T_{ij},\ C_{ij}),\ \text{with} \begin{cases} i & \text{stands for}\, Sector \\ j & \text{stands for}\, Country \end{cases} \tag{2.1}$$

where Y_{ij} is an indicator of international competitiveness such as export market share or trade balance.

The estimation of equations of type (1) generated a relevant stream of empirical literature pointing to the crucial role played by innovative activities and knowledge flows in explaining the international competitiveness of industries and countries. Due to data constraints, most of the empirical work within the "technological gap" framework has been carried out at country or industry-country level. In a pioneering empirical work, Soete (1981, 1987) provides some evidence of the relevance of technological factors as determinants of competitiveness. In a sample of OECD countries, across several sectors, results show a strong relationship between patent activities (as a proxy for technological performance) and export performance. At the country level, Fagerberg (1988) examines the effect of technological factors (patents, R&D)

and of investments over unit labor cost (as a proxy for competitiveness) in order to explain growth in export market shares. Results are consistent with the so-called "Kaldor paradox" (Kaldor, 1978, was among the first authors to show that export market shares and relative unit costs or prices move towards the same direction.). Greenhalgh (1990) supports as well the idea that innovations sustain export performances and also finds (focusing on UK) stark heterogeneity across industries, with relative prices negatively affecting export only in few sectors. As far as the time dimension is concerned, Amendola et al. (1993) report a positive and significant effect of technological variables (patents and investments) on export shares in the long-run. Unit labor cost plays a role only in the short-run.

Such results have been confirmed by analyses at level of country and industry. In particular, taking into account twenty countries and forty sectors, the cross-sectional analysis of Dosi et al. (1990) supports previous findings. Indeed, they clearly show that technological variables (investments and patent shares) positively affect several export measures, whereas cost-related factors (wages and unit labor cost) appear to have little or no effect.

Following a similar econometric approach, Magnier and Toujas-Bernate (1994) and Amable and Verspagen (1995) confirm the positive results for different innovation proxies (patents, investments and R&D). In addition, Wakelin (1998) uses bilateral trade flows and shows that R&D intensity and patents are crucial in high and low-technology sectors. Cost variables are instead significant only in medium and low knowledge-intensive sectors. Finally, Carlin et al. (2001) measure export market performance of OECD countries finding ambiguous results. Both costs and technology play a role in describing changes in export positions: however neither is sufficiently strong to fully explain such performances.

More recently, Guarascio and Pianta (2017) have analyzed the complexity of the so-called "virtuous circles" that link technological innovation, international competitiveness and profit dynamics. Building on previous work (Guarascio et al., 2015, 2016), they stress the relevance of *gains from technology* (vis-a-́vis cost factors) in boosting trade competitiveness, confirming results in Dosi et al. (2015).

2.2.2 The Role of Spillovers and Inter-sectoral Knowledge Flows

In general, technology affects competitiveness not just directly, but also indirectly through technological spillovers. Griliches (1979) distinguishes two types of spillovers: "rent-spillovers" and "pure knowledge spillovers". Such distinction arises from several different mechanisms through which knowledge and technology can spread. In particular, spillovers embodied in products represent the specific category of rent-spillovers. Thus rent-spillovers cannot be assumed as pure externalities since they are intrinsically dependent on the market structure of supplying and using industries. Conversely, pure knowledge spillovers are mainly related to the technology and may constitute true externalities.

Along these lines, Grossman and Helpman (1991, 1995) theoretically investigate how international trade in commodities may boost the exchange of intangible knowledge and ideas as well as how differences between international and national spillovers contribute to the formation of the knowledge base.

In parallel to international trade analysis, evolutionary scholars have focused on the effect of innovation on the dynamics of firms and industries (Nelson and Winter, 1982; Dosi, 1988; Dosi et al., 1990; Freeman et al., 1982; Malerba et al., 2016) and on the role played by institutions and national innovation systems in affecting the growth and competitiveness of countries (Nelson, 1993; Freeman, 1987). The evolutionary and Schumpeterian literature has associated spillovers to technology and knowledge and shifted the focus from automatic pure spillovers to flows of knowledge that may run across firms and countries in less automatic way, often related to the role of absorptive capabilities of the recipient firm and country (Cohen and Levinthal, 1989, 1990; Cimoli et al., 2009; Dosi et al., 2008).

One key aspect of knowledge flows refers to inter-sectoral flows. This is related to the importance that has been given to industries and sectors in the examination of the international performance of countries. Inter-sectoral knowledge flows has been intensively studied with input-output data and technology flows matrices based on patents (Scherer, 1982; Putnam and Evenson, 1994; Verspagen, 1997a,b; Laursen and Drejer, 1999).

As far as input-output links are concerned, Scherer (1982) and Putnam and Evenson (1994) follow an approach based on the relationships between supplier and user industries[2]. As input-output links, a certain innovation/product generated by an industry $\mathcal{A}$ can then be used by an industry $\mathcal{B}$. Clearly, this way of reasoning is consistent with the notion of what we defined rent-spillovers (Griliches, 1979).

As far as technology flows matrices are concerned, Verspagen (1997a,b) proposes three different approaches to analyze pure technological spillovers. The first matrix they use relies on data from EPO and it is constructed on the basis of main and supplementary IPC codes. Such step is employed for claimable knowledge. The second matrix is derived following the same principle, although it takes into consideration the supplementary codes for unclaimable knowledge[3]. In practice the main code identifies knowledge producing-sectors, whereas spillovers are eventually captured through the relationships with supplementary IPC codes. Finally, the third matrix is constructed using citations in the US patent database. It is argued, of course, that knowledge flows from the cited to the citing patent sector. An alternative approach is proposed by Jaffe (1986), who measures technological distance among US firms on the basis of the distribution of firms' patenting activities[4]. It must be noted that most of the aforementioned works have been carried out with the purpose of quantifying the impact of spillovers on productivity and innovative activities[5].

[2]The method is the backbone of the so-called "Yale-matrix" that relies on the Canadian Patent Office data.

[3]In the EPO data supplementary classes may contain invention information (claimable) and additional information (unclaimable).

[4]Formally, Jaffe (1986) employs the so-called *cosine index* to capture such distance.

[5]See Griliches (1998); Jaffe and Trajtenberg (2002) for a complete treatment of the topic.

In addition to the studies on spillovers, the "home market hypothesis" literature considers the effect of technological spillovers on international trade dynamics and specialization[6]. Particularly, it suggests that domestic inter-sectoral linkages are of paramount importance in explaining trade flows and specialization. The "home market hypothesis" has been empirically investigated by Fagerberg (1992, 1995)[7]. However, his empirical analysis only considers "backward linkages" and makes use of trade statistics and Revealed Comparative Advantage (RCA) to measure both competitiveness of the producers of technology and how advanced the domestic users are. Based on actual I-O data, Laursen and Drejer (1999) introduce upstream and downstream linkages as a possible technological source of export specialization. Such findings prove inter-sectoral linkages to be a determinant of specialization. However, the importance differs according to the type of sector (e.g. following the Pavitt taxonomy). Subsequently, Laursen and Meliciani (2000, 2002) find a positive effect of national R&D linkages on competitiveness. Interestingly, they find that only national spillovers have a clear impact on trade balance. Differently, Laursen and Meliciani (2010) investigate the role of ICT knowledge flows and conclude that in ICT industries both national and international linkages have a positive effect on export market shares.

2.3 Industries' position, knowledge space and international Competitiveness

2.3.1 Position

In this chapter, we propose a novel way to look at inter-sectoral flows of knowledge. We shift the emphasis from the flows of knowledge related to bilateral industrial

[6]We will discuss the *home market effect* in greater detail later in the chapter.

[7]Moreover, Fagerberg (1997) examined the effect of domestic and foreign R&D on export performance.

relationships to the position of an industry in the entire inter-sectoral knowledge space. The reason for such change is that the position of an industry in a technological space in terms of links with the other industries, both nationally and internationally, provides a more complete and articulated representation of all direct and indirect inter-sectoral knowledge flows that an industry has. For example, Antonelli et al. (2017) recently showed that the composition of local knowledge is a major determinant of innovative activities. Our approach also benefits from the literature concerning the measurement of technological relatedness and proximity in a broader sense. Indeed, the interaction among different dimensions of proximity results of paramount importance for learning and innovation (Breschi et al., 2003; Engelsman and van Raan, 1994; Boschma et al., 2014; Kogler et al., 2013). We advance the claim that an industry that is central in the flows of knowledge among sectors and that is highly connected with the other sectors, obtains major benefits in terms of competitiveness. We propose that the following three mechanisms may explain our claim.

Variety in knowledge and opportunities. A first mechanisms is that an industry that is central in the flows of knowledge across industries enlarges its opportunities to come across pieces of potentially useful knowledge and, hence, its chances to boost its market performances. This is consistent with the so-called "specialization-based" trade growth, which links trade performances to the ability to exploit above average technological opportunities arising in certain sectors (see Laursen, 1999, and references therein). In such a context, technological opportunities have been usually measured through growth rates in patenting activity (Cantwell and Andersen, 1996; Meliciani, 1998). However, Laursen (1999) shows that there is little empirical support for the hypothesis that being initially specialized in fast-growing industries yields a positive effect on trade performances. As extensively argued in Klevorick et al. (1995), technological opportunities in one industry can be enriched by technological advances that are achieved in others. Further, such an extra-industry source of technological opportunities positively and significantly correlates with both process and product innovation in the relevant industry. The relationship between opportunities and innovation has been investigated in various ways. Malerba and Orsenigo (1997) suggest that the the specific pattern of innovative activity of a sector can be explained by the structure of the underlying knowledge, which seize opportunities together with learning processes (see also Dosi, 1988). Empirically, Becker and Peters (2000) and Oltra and Flor (2003) confirm that technological opportunities from other industries sustain innovative performances in a sample of German and Spanish firms respectively. Cohen and Malerba (2001) point out that greater diversity in innovative activities results positively associated with faster technological change. Moreover, the existence of an inverted-U relationship between technological diversification and firms' technological performance (Leten et al., 2007; Garcia-Vega, 2006) suggests that the effect of broadening technological opportunities enhances performances, provided it does not become too high. Furthermore, the larger is the pool of opportunities and technological linkages, the lower are the chances that firms in a given industry remain locked in to inferior technologies. This effect comes from being exposed to a large learning basin and having the possibility to mold such flows into effective knowledge due to connections (Boschma, 2005; Balland et al., 2015).

Recombination. A second mechanism is that inventions and innovations develop more easily, and have a greater impact on the economic system (and therefore also on competitiveness), when firms combine knowledge across different technological domains, which in turn may belong to different sectors (Ferguson and Carnabuci, 2017; Fleming and Sorenson, 2001; Basalla, 1988). Scholars have found that a large part of technological advances comes to a good extent from multidisciplinary R&D (Kodama, 1986; Rosenberg et al., 1992). Moreover, both theoretical and empirical literature provide evidence that spanning knowledge domains might give inventors a wider vision of technological opportunities (Ferguson and Carnabuci, 2017; Hargadon and Sutton, 1997; Hargadon, 2002) while knowledge complexity substantially influences the diffusion dynamics (Sorenson et al., 2006). The idea that recombination might help creating something new and potentially useful goes back to Schumpeter (1934)[pag. 65]. Drawing on Galunic and Rodan (1998), we claim that recombination of resources - including knowledge - facilitate the creation of novel systems. Following these lines, being exposed to several different technological flows coming from different industries may reduce uncertainty and significantly increase the usefulness of innovation (Fleming, 2001). Hence, knowledge flows and technological linkages boost the possibility of recombining knowledge. Knowledge diffusion and the network structure of inter-sectoral relationships clearly affects the possibility to integrate different pieces of knowledge, especially for multidisciplinary innovation (Sorenson et al., 2006).

Improvement of absorptive capabilities. A third mechanism is that an industry exposed to knowledge coming from different other industries increases its absorptive capabilities of selecting, identifying and using various pieces of knowledge that can be relevant for its problem solving (Von Hippel, 1994; Owen-Smith and Powell, 2004) and innovative activities (Cohen and Levinthal, 1989, 1990; Lundvall and Johnson, 1994). In that, an increase in the absorptive capacities within an industry results from a successful process of learning and external knowledge management, which may be influenced by different - both geographically localised and not - factors (Boschma, 2005; Boschma and ter Wal, 2007; De Noni et al., 2017). Centrality in the knowledge flows increases the experience of firms in an industry in managing different types of knowledge. In addition, being exposed to knowledge coming from different industrial contexts increases the capability of understanding different application contexts (Christensen et al., 1998). If market success ultimately depends on the ability to channel R&D for attracting final demand rather spending in research activities per se (Iansiti, 1995), then being central in a network of knowledge flows from different industries increase the amount of information on fields in which a technology can be successfully exploited. Finally, Burt (2004) has witnessed how crucial network position (brokerage) and the development of organizational abilities are in influencing firms' innovative performance. To sum up, being exposed to knowledge flows may help industries develop technological as well as managerial capabilities to effectively master different technologies and eventually match them with the most appropriate context.[8]

We believe that a knowledge space approach offers new insights and fill the gap existing in the literature by merging together both social networks and absorptive capabilities lines of research. Recently, Duernecker and Vega-Redondo (2017) theoretically show that the social network is the main channel through which agents exploit new opportunities. In their empirical companion paper they found that centrality is a very significant variable in explaining differences in countries' growth performances (Duernecker et al., 2015). Operti and Carnabuci (2011) and Tortoriello (2015) provide additional empirical evidence consistent with the theoretical framework formulated here. A more structured modelization of knowledge space has been adopted by Tomasello et al. (2016) and Vaccario et al. (2017) in studying R&D alliances and knowledge exchange among firms. Finally, to these three factors related to knowledge, it is possible to add some remarks about the variety of channels and organizational forms through which knowledge crosses industry boundaries. While the channels that have been most widely studied refer to informal mechanisms (see for example Fagerberg et al. (2006)), personnel mobility between firms (for example Saxenian (1990); Almeida and Kogut (1999)), vertical integration (for example Helfat (2015)) and inter–organzational agreements (for example Hagedoorn (2002)), recently also channels related to new firms originated in the upstream or downstream industries that enter a focal industrial sector -i.e. vertical

[8]Along these lines, the interested reader may want to look also at the literature on the role of embeddedness in boosting performances at different levels (e.g. Ahuja, 2000; Andersen, 2013).

spinouts- have been studied (Adams et al., 2016, 2018). While this chapter does not aim to examine the informal, individual or organizational channels through which knowledge flows across industries, it must be emphasized here that a channel may affect how much and what type of knowledge is transmitted. For example, in the case of new firms spinning out from upstream or downstream industries and entering a focal industry, the knowledge transmitted that passes through industry boundaries is application knowledge for downstream spinouts and technological knowledge for upstream spinouts.

By combining the aforementioned arguments, we can advance the first proposition to be tested empirically:

> **Proposition I - Position in inter-sectoral knowledge space:** *Industries more central in the inter-sectoral knowledge space perform at the international level better than industries that are not central.*

Rethinking centrality in a knowledge space as a composite measure of innovativeness, we can appreciate the moderating effect of learning by being exposed to knowledge flows. As a matter of fact, the greater the amount of information passing through a certain node, the greater will be the capacity of that node to retain and process knowledge flows. Such learning channels may well be captured by degree centrality and local clustering. As we will discuss in depth in the methodological Section 2.4, degree centrality and local clustering measure how likely a node ends up being susceptible to all kind of information running through the network, giving us the possibility to measure its ”skills” as a recipient of technological flows. Summing up, the centrality of industries in the inter-sectoral knowldge space allows us to capture several possible mechanisms through which technology flows can boost international competitiveness. To some extent, either too much or to little proximity may result detrimental to innovativeness and effective learning (Boschma, 2005). On the one hand, a larger learning basin lead to a wider set of opportunities (superior technologies, innovative products, cost reductions, diffusion of best practices).

However, such advantages take place if and only if there are sufficient strong linkages to support knowledge transfer. All told, using network position as a proxy for a richer set of opportunities and capabilities, we can try to incorporate them into our model.

2.3.2 Geographical boundaries

In this chapter, we propose that not only the position in the the knowledge space, but also the geographical boundaries matter in the inter-sectoral flows of knowledge. We argue that the effects of inter-sectoral knowledge flows on international competitiveness are more relevant at the country level due to the geographical boundaries that affect knowledge flows. In a nutshell, the agglomeration literature posits that knowledge spillovers have clear geographical reach and they are subject to a significant spatial decay. The diffusion of tacit knowledge, to some extent, requires close and frequent interactions, i.e. geographical proximity (Lissoni and Miguelez, 2014). The geographical concentration of people and jobs enhances a rapid and effective spread of tacit knowledge, resulting in a boon for innovative activities. Although the specific mechanism behind such knowledge transfer is not completely disentangled, there is nowadays substantial empirical evidence confirming the localized nature of knowledge diffusion (Arzaghi and Henderson, 2008; Rosenthal and Strange, 2003; Audretsch and Feldman, 1996; Adams and Jaffe, 1996; Carlino and Kerr, 2015). Here we present two different possible explanations on why "local" knowledge flows are expected to be more effective in sustaining competitiveness.

Localized knowledge flows. First, effective mechanisms of knowledge exchange require close interactions, frequent meetings and development of trust among economic agents. Within this framework, spatial proximity boosts the flow of ideas by sharply reducing the cost of trading knowledge, enhancing skilled worker mobility and providing better conditions for cooperation among among firms and individuals

(Breschi and Lissoni, 2001). Hence, localized flows are relatively richer of easily exploitable ideas. Further, contrarily to codified knowledge, the diffusion of tacit knowledge might be seriously affected by proximity, which enhance shared routines, similar technology attitudes and trust (Bathelt et al., 2004). Since the first tests on the role of spatial proximity in fostering scientific collaborations (see e.g. Jaffe et al., 1993), the economic geography literature has largely extended the line of research concerning the geographical breath of knowledge flows and their features (Carlino and Kerr, 2015). More in detail, inventors are not very likely to relocate in space and their (bounded) mobility - as well as their co-invention networks - circumscribe the geographical diffusion of knowledge (Singh, 2005; Breschi and Lissoni, 2009; Sonmez, 2017). One of the emerging results suggests that national-scale interactions allow for a more effective transmission and exchange of tacit knowledge than on broader scale. For example, so-called Jacob externalities may result more effective at national or regional level, where the heterogeneity in the composition of the knowledge base can be managed more easily and flows integrated at a lower absorption cost (Antonelli et al., 2017). Indeed, in their review of the literature, Breschi and Lissoni (2001) underline that although there is variety of mechanisms behind the spread of ideas and expertise, such a diffusion remains, however, largely bounded in space even though exact co-location might not be essential (see also Gallaud and Torre, 2005; Torre, 2008). Moreover, inventor mobility and co-invention networks have been proved to account for a large francion of the spatial proxinty of knowledge diffusion

"National institutions and home market effect". Basically, the idea is that a country's domestic market may act as a supportive and protective environment for new products, then ready to be successfully exported to foreign markets. The product-life cycle model, introduced by Vernon (1966), supports the idea that geographical proximity is conducive to innovative activities due to the ease of communication and that, at least at the beginning of the product-life cycle, domestic market matters, providing easier, faster and more complete access to information and knowledge. Country's domestic markets can be thought as a space serving as "nurturing grounds" for new products (Linder, 1961; Hirschman, 1958; and more recently, Diodato et al., 2018 and Li et al., 2018). With respect to the home market's role emphasized in Krugman (1980), which largely depends of the size of the domestic market, we hypothesize that it is the domestically developed technical knowledge that positively affects competitiveness. Hence, if national-level institutions matter in the process by facilitating the flows of knowledge (see also Gittelman, 2006), it is reasonable to argue that they will be also more helpful to innovation and trade than international ones (Laursen and Drejer, 1999). In addition, the literature on national systems of innovation has frequently emphasized the pivotal role of within country knowledge flows and of national institutions as determinant of economic performances (Lundvall, 1988, 1992).

On the basis of these mechanisms we conjecture, in our second proposition, that a central position is still important but less relevant at the international level, where knowledge flows are more codified and available to all countries and competitors.

> **Proposition II - Geographical boundaries of inter-sectoral knowledge flows:** *The position of an industry in the inter-sectoral knowledge space is more relevant at national level than at the international level.*

To summarize, a variety of mechanisms point to the fact that knowledge flows suffer from geographical boundaries. Both knowledge production and diffusion entail a local dimension linked to the easier interaction of different actors. Further, part of the literature supports the idea that the national dimension matters, due to role of common institutions and a relatively more supportive market. Building on such premises, we conjecture that industries benefit more from their position in their national knowledge network rather than the one they have in the international space.

Table 2.1: List of Countries

Country	Code	Country	Code
Austria	AT	United Kingdom	GB
Belgium	BE	Italy	IT
Canada	CA	Japan	JP
Germany	DE	Luxembourg	LU
Denmark	DK	Netherlands	NL
Spain	ES	Norway	NO
Finland	FI	Sweden	SE
France	FR	United States	US

2.4 Data and methodology

2.4.1 Data

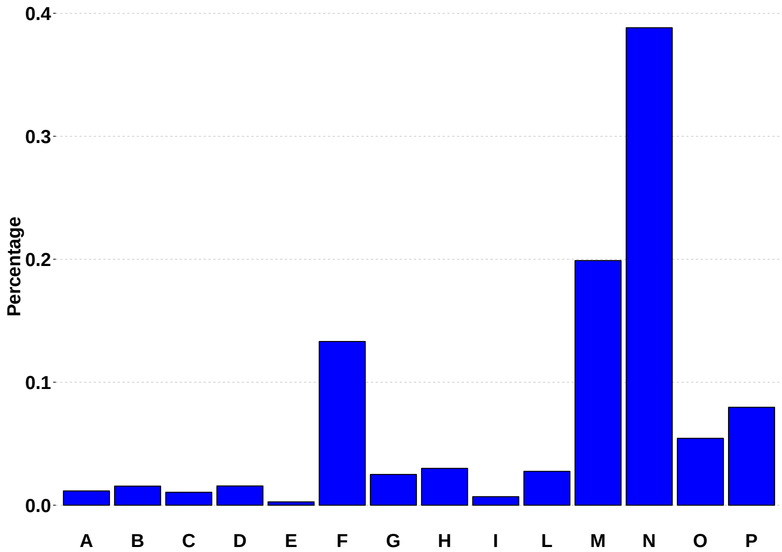
The empirical analysis of this chapter is based on two main sources of data: the ICRIOS-PatStat database and the STAN database (OECD). The STAN database for industry analysis provides comprehensive information to investigate industry performance across countries. The ICRIOS-PatStat contains the full set of bibliographic variables for patents applied at EPO and USPTO (Coffano and Tarasconi, 2014)[9].

More in detail, for patents we consider all the applications with priority date in the time interval 1995-2009. By merging and elaborating the aforementioned inputs, we obtain a dataset that includes information about 14 manufacturing industries in 16 OECD countries for 15 years[10]. A similar approach has been used in order to collect citation data. The following tables (2.1 and 2.2) and figures (2.1 and 2.2) provide a more quantitative and exhaustive description of our data and our industry classification based on ISIC3 codes[11].

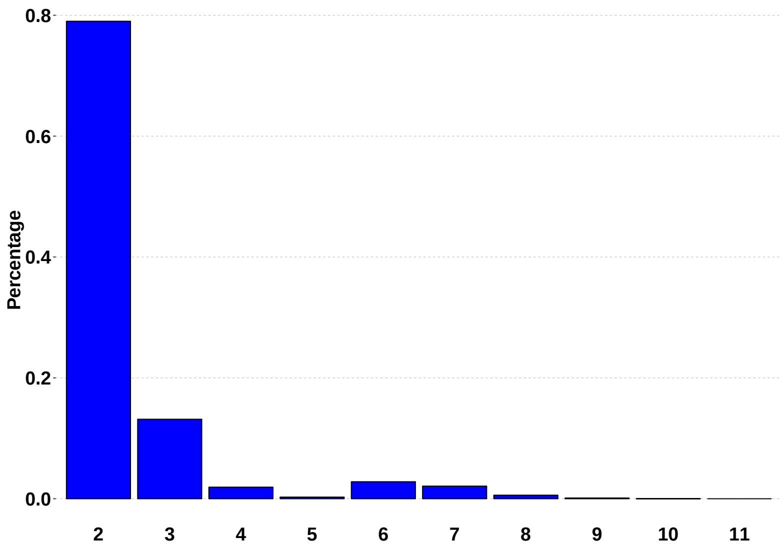
[9] PatStat (i.e. EPO Worldwide PATent STATistical Database) is a single patent statistics raw database, held by the European Patent Office (EPO) and developed in cooperation with the World Intellectual Property Organisation (WIPO), the OECD and Eurostat.

[10] The timespan for which we collected and analyzed the data stops in 2009. Such choice is driven by the occurrence of the Great Recession, that severely affected all the OECD countries in our dataset.

[11] For compatibility reasons our classification is based on ISIC3 codes. The initial NACE2 classification has been converted into ISIC3 by means of standard conversion tables.

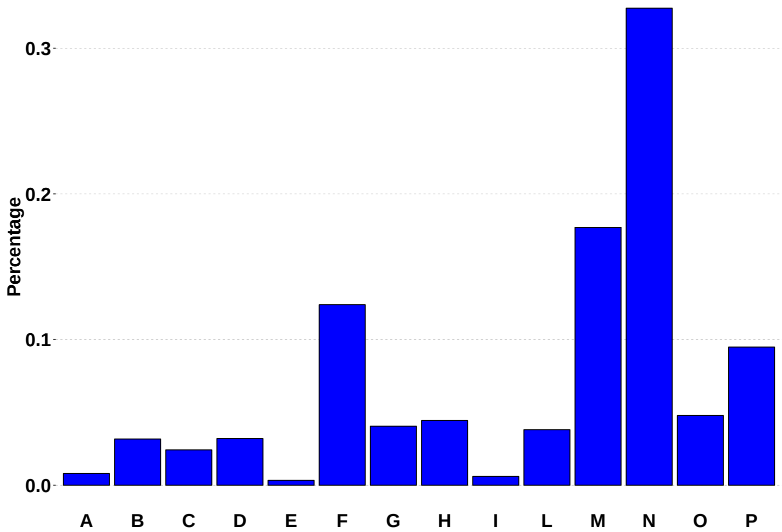


(a) Distribution of Patent Applications by Industry

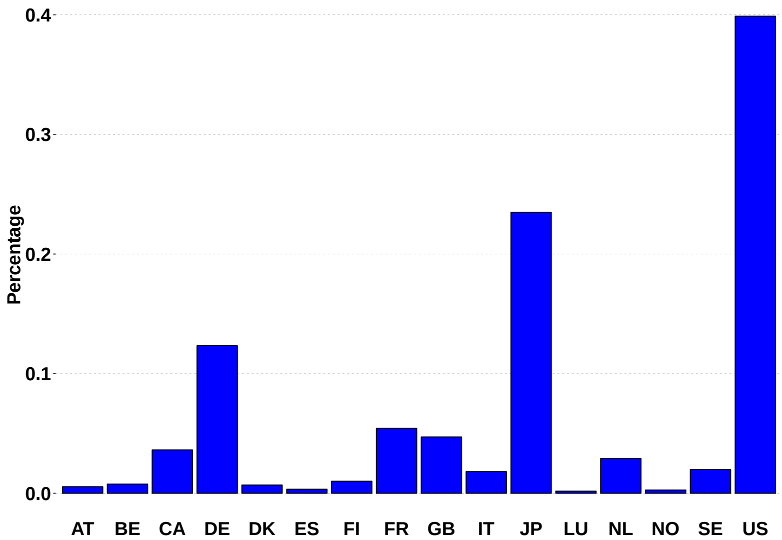


(b) Distribution of Patent Applications by Number of Industries

Figure 2.1: Patent Applications



(a) Distribution of Aggregate Citations by Industry



(b) Distribution of Aggregate Citations by Country

Figure 2.2: Aggregate Citations

Table 2.2: Industry Classification

Industry	ISIC3	CODE	APPLICATIONS		CITATIONS	
			%	#	%	#
Food, beverages and tobacco	15-16	A	1,17%	44091	0.80%	437616
Textiles, wearing, leather	17-19	B	1,55%	58900	3,18%	1723416
Wood	20	C	1,05%	39944	2,44%	1318938
Paper and printing	21-22	D	1,57%	59407	3,20%	1737386
Coke	23	E	0.28%	10402	0,33%	181814
Chemicals	24	F	13,31%	504805	12,40%	6720031
Rubber and plastic	25	G	2,50%	94891	4,05%	2200584
Non-metallic (mineral products)	26	H	2,99%	113633	4,45%	2409767
Basic metals	27	I	0.70%	26383	0,60%	328165
Fabricated metals (products)	28	L	2,75%	104441	3,81%	2068636
Machinery	29	M	19,90%	754265	17,70%	9601586
Computing and electrical (machinery)	30-33	N	38,82%	1472210	32,75%	17761391
Transport	34-35	O	5,44%	206271	4,79%	2595884
Other manufacturing	36-37	P	7,97%	301956	9,50%	5147308
Total			100%	3791599	100%	54232522

Notes: If an application (citation) has been assigned to two (or more) industries according to original IPC codes then it is counted twice (or more). The total number of applications (citations) has been derived accordingly.

2.4.2 Knowledge flows and the network of industries

The approach used in this chapter basically follows a two steps procedure. The first step consists in mapping technology flows among the 14 industries included in our dataset. Taking into consideration the empirical evidence in Laursen and Meliciani (2002, 2010), we consider both national and international knowledge flows. Consequently, we distinguish between the national and the international dimension of the flows. In order to do so, we obtain two sets of symmetrical matrices that will constitute the adjacency matrices for our networks. This methodology represents the framework to construct a national and an international technology space in the form of a network. Such networks provide a representation of inter-sectoral relationships and a characterization of industries' position in our space of knowledge flows. Moreover, this framework allows us to eventually capture the relative centrality of industries. The network representation of a knowledge space has been adopted by Kogler et al. (2013) and Boschma et al. (2014) in order to link technological sectors according to their relatedness. Yet, the goal of our analysis is to capture flows.

The main source of information is given by patent classification codes. As we explain later in this Section, relying on classification codes has a number of advantages with respect to patent citations (Joo and Kim, 2009). However, some methodological issues arise in capturing international flows. We aim to overcome technical difficulties by approximating such relationships through a patent citation network (Verspagen, 1997b; Jaffe and Trajtenberg, 2002).

Following the methodology employed in Engelsman and van Raan (1994) and Breschi et al. (2003), we can perform a co-classification analysis based on co-occurrences according to our classification of industrial sectors[12] As pointed out by Breschi et al. (2003), Hinze et al. (1997) and several other WIPO documents, main and supplementary IPC codes cannot be used to disentangle knowledge-producing and knowledge-incorporating sectors. Hence, contrary to Verspagen (1997a,b) we do not infer anything about the direction of the flows. Our purpose is simply to map technological relationships among industrial sectors regardless of formal spillover effects.

Our choice of using co-occurrences based on patent classification codes (with respect to patent citations) derives from several methodological considerations. Patent citations provide a great source of information, although it has been shown that they present several drawbacks in certain applications. For instance, citations are a fully reliable measure in scientific academic settings. Indeed, Joo and Kim (2009) clearly state that the channels through which classification and citations are generated may lead to substantial differences. Alcacer and Gittelman (2006) show how citations added by patent examiners generate noise in the data resulting in a relevant measurement error. Conversely, IPC codes are carefully assigned by patent examiners of the issuing office in accordance to strict WIPO requirements. Leydesdorff (2008); Cockburn et al. (2002) and Criscuolo and Verspagen (2008) argue that citations are subject to authors and examiners choices and that may be the result of legal and strategic factors (Meyer, 2000). Finally, Breschi et al. (2003) show that citations do

[12] From patent data we match technology classes (IPC) with industry classes (ISIC3). In particular, we rely on the information on the NACE code associated to patents from the PatStat database (see Van Looy et al., 2015 for the conversion table IPC-NACE2) and then use the EUROSTAT RAMON coversion tables to move from NACE to the desired ISIC classification employed by the OECD STAN database.

not add any relevant information to track simple technology flows.

Unfortunately, co-classification is not feasible for examining international technology flows since available information does not allow us to fully disentangle industry classes for couples of countries and industries. As a result, we need to rely on patent citations[13]. Notwithstanding all the shortcomings outlined above, patent citations provide a good approximation for a measure of knowledge flows among industries of different countries (Jaffe and Trajtenberg, 2002, 1999; Verspagen, 1997b). EPO and USPTO data which are, indeed, sufficiently complete to have a good coverage of innovative activities for all countries that we take into consideration (Joo and Kim, 2009). For all these reasons, we believe that our approach to map knowledge flows across industries is the most suitable in this specific application[14].

Summing up, the first step of our methodological approach is essentially driven by two factors: the superiority of co-classification analysis in mapping technology flows across sectors and the impossibility to replicate the exact procedure for international relationships. However, for completeness we perform a robustness check using citation data for both national and international flows. The results can be found in Table 2.6 and prove that citations can eventually represent a good approximation of international knowledge flows.

In what follows, we formally describe our procedure to build a national knowledge space. We apply an almost identical methodology in order to construct a citation network to control for international relationships.

Let $\mathcal{A}$ be the set of all patent applications. Then, $\mathcal{A}_{ct} \subset \mathcal{A}$ is the set of all patent applications for a given country c at a certain point in time t[15]. Each $a_{ct} \in \mathcal{A}_{ct}$ has

[13] Investigations of citation patterns in our dataset show a clear tendency of a country-specific dimension. See Figure A.2 in the appendix.

[14] The distinction between national and international measures is not a matter of differences among countries/industries, it rather concerns the nature of co-occurrence and citation data. Using co-occurrences, we are not able to disentangle, and thus to count in a meaningful way, every IPC-country link. We overcome such difficulty by relying on citations, which include, in a way, an additional layer of information to map within-country industry relationships as well as across-country industry linkages.

[15] $c \in \{AT, \ldots, US\} \equiv \mathcal{C}$ and $t \in \{1995, \ldots, 2009\} \equiv \mathcal{T}$

been assigned to one or more industry class. Let $P_{ia_{ct}}$be a function such that

$$P_{ia_{ct}} = \begin{cases} 1 & \text{if } a_{ct} \text{ has been assigned to industry } i \\ 0 & \text{otherwise} \end{cases}$$

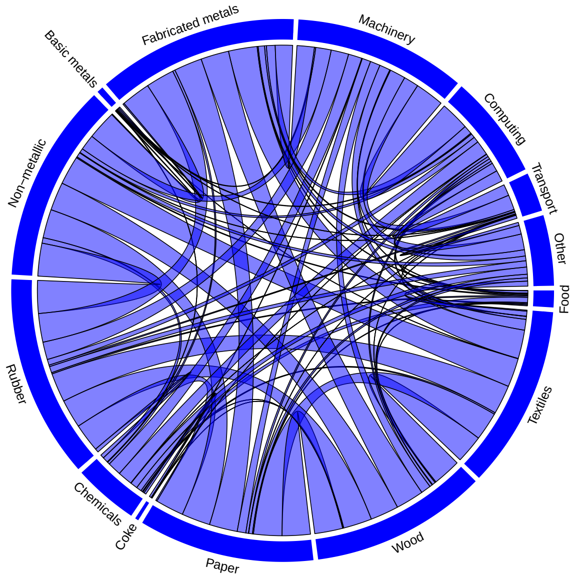
with $i \in \{A, \ldots, P\} \equiv \mathcal{I}$. Thus, for each country c at time t, the total number of patent applications that has been assigned to code $i \in \mathcal{I}$ can be written as $T_{ict} = \sum_{a_{ct} \in \mathcal{A}_{ct}} P_{ia_{ct}}$; while the total number of patent applications classified in both industrial sectors i and j is simply given by $C_{ijct} = \sum_{a_{ct} \in \mathcal{A}_{ct}} P_{ia_{ct}} P_{ja_{ct}}$. By repeating the count for every pairs of possible industry codes, we obtain a symmetric co-occurrences matrix $\mathbf{C}_{ct}$, of dimension (14×14), for every country at each point in time.

We consider such matrices as adjacency matrices of our networks. That is, $\mathbf{C}_{ct}$ formally defines a network of inter-sectoral relationships among industries for country c at time t. We use the notation $\Gamma_{ct} = (\mathcal{I}, \mathcal{L})$ where $\mathcal{I} = \{A, \ldots, P\}$ is the set of nodes and $\mathcal{L} \subseteq \mathcal{I} \times \mathcal{I}$ is the set of links.

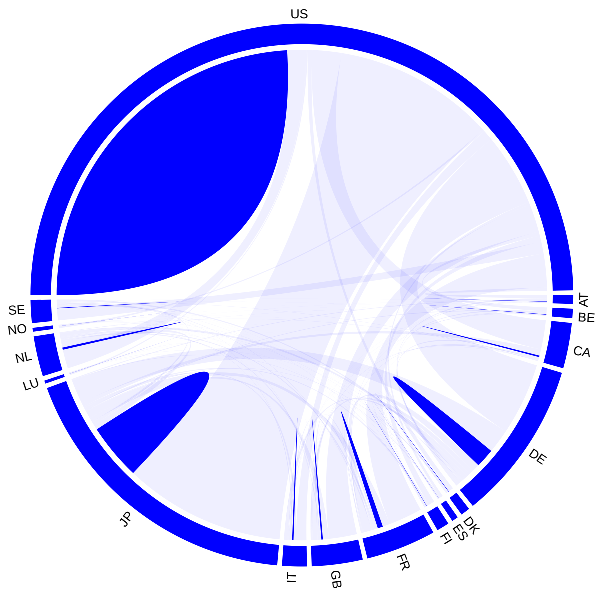
As long as we consider a weighted network, the matrix representation takes the following form:

$$\mathbf{c}_{ct} = \begin{cases} C_{ijct} & \text{if } (i,\, j) \in \mathcal{L} \\ 0 & \text{Otherwise} \end{cases}$$

with $C_{ijct} \in \mathbb{N}_{+}$. Figure 2.3 is an illustrative example of networks derived through the above mentioned procedure and describes the national (Italian) and international knowledge space in 2009. More in detail, in (a) nodes' sizes are set according to the degree centrality and links' widths are proportional to weights C_{ijct}. In (b) the entire space - aggregated by country - is mapped to visualize international connections, within-country relationships are empathized. In the next Sections, we derive more insightful network measures and we rely on such measures to construct the econometric strategy of our analysis.



(a) 2009 - ITALY



(b) 2009 - By Country

Figure 2.3: Knowledge Networks based on Patent Data

2.4.3 Variables & descriptive statistics

The variables used in order to study the relationship between the competitiveness of industries and their position in our knowledge space are summarized in Table 2.3.

As a general measure of international competitiveness we consider export market shares (XMS). Such measure is derived by taking into account country's exports in a given industry (current dollars) over the total industry's exports from all countries included in our dataset. The choice of some regressors follows Dosi et al. (2015). More in detail, tech-related variables are represented by Patent-share (PATSH) and Investments (INV). Patent-share captures the share of national industry patent applications (USPTO and EPO) over the total industry's patent applications of all countries in the dataset. Investments is defined as the ratio between industry expenditures on gross fixed capital formation and value added (current prices). Moreover, we include in the analysis a price-related variable: Labor-cost-per-employee (WAGE). Total population (POP) controls for possible size effects.

The impact of industries' position in our knowledge space captures the inter-sectoral diffusion of advanced knowledge. Several measures of centrality have been developed in order to capture different features of the network structure and identify key players. Here it is necessary to briefly review the most important ones. Freeman (1978) formalizes three different basic measures of centrality: degree, closeness and betweenness. The most direct measure of popularity is the degree centrality, which is defined as the number of links a node has in the network. It can be interpreted in terms of the immediate risk of a given node for catching whatever is flowing through the network. Instead, closeness centrality is defined as the inverse sum of shortest paths to all other nodes from a given node in the network and it measures whether a node is in the position of reaching information quickly. Betweenness centrality is defined as the geodesic path that passes through a given node and it captures the property of controlling information flows within a given graph. Therefore, it can be used to identify who plays the role of a broker or a gatekeeper. As Burt (2004) points out, such bridging position can represent power and can be associated with consistent advantages since knowledge and information must pass through such nodes. Finally, Bonacich (1987) develops a more sophisticated measure to evaluate the most influential nodes which is called Eigenvector centrality. Such measure assigns different weights to links according to the relative influence of a node and it has been widely applied in the literature to assess power, the structure of inter-organizational networks and the role of an individual or an entity in a general social network.

The local clustering coefficient of a node in a network is used to quantify how connected its neighbors are and whether they form a clique (complete graph) or not. Watts and Strogatz (1998) in their seminal paper constructed a model that accounts for both local clustering and small-world property of networks. Despite most of such measures have been initially developed for binary networks, they can easily be generalized for weighted networks (Opsahl et al., 2010; Barrat et al., 2004). For our purpose, we choose two simple network measures for both networks (co-occurrences and citations): the weighted degree centrality and the local clustering coefficient[16].

Formally, we can define weighted degree centrality for a network $\Gamma = (\mathcal{I}, \mathcal{L})$ as follows:

$$d.w_i = \sum_{j \in \mathcal{I}} C_{ij} \tag{2.2}$$

Such simple measure captures network centrality in a direct and immediate fashion (Borgatti, 2005). Indeed, weighted degree can be interpreted as the opportunity to influence as well as be influenced directly. As a result, central actors are more likely to be exposed to what is flowing through the network, in this specific case knowledge.

[16]For completeness, in the appendix we include the unweighted degree centrality and the eigenvector centrality.

For what concerns the local clustering coefficient, we use the generalization for weighted networks proposed by Barrat et al. (2004). The analytical expression in which we removed the dependence from time to ease notation, reads as follows:

$$am_i = \frac{1}{d.w_i(k_i - 1)} \sum_{j,h} \frac{C_{ij} + C_ih}{2} \xi_{ij}\xi_{ih}\xi_{jh}, \tag{2.3}$$

where k_i is the number of industries linked to i and ξ_{ij} is an indicator function that takes value 1 if industry i is linked to j and 0 otherwise. This coefficient is a measure of the local cohesiveness that takes into account the importance of the clustered structure on the basis of the amount of interaction intensity actually found on the local triplets. Indeed, am_i counts, for each triplet formed in the neighborhood of the vertex i, the weight of the two participating edges of i. Using this measure we are considering not just the number of closed triplets in the neighborhood of a vertex but also their total relative weight with respect to the strength of the vertex. The normalization factor $d.w_i(k_i - 1)$ accounts for the weight of each edge times the maximum possible number of triplets in which it may participate, and it ensures that the local clustering coefficient always falls between 0 and 1.

Within this setting the neighborhood of a node can play a crucial role. Even if an industry would result not particularly central according to the weighted degree, it might belong to a clique and such embedness in the network can guarantee a competitive advantage anyway. Therefore, the weighted local clustering coefficient helps us to eventually capture the impact of such connectiveness.

By looking at Table 2.4 below we can observe how the two network measures correlate with our baseline variables through a cross-correlation matrix. For instance, at national level we can notice how weighted degree ($d.w$) is positively associated with both export market share and patent share. Such positive relationship holds for local clustering (am) as well. We will focus on the two network measures described above for all the aforementioned reasons, although some alternative specification are summarized in the appendix (Table A.1).

Table 2.3: Variables

Var. Name	Description	Data Source
XMS	Country's exports in the industry over the total industry's export	OECD-STAN
INV	Ratio between industry expenditures on gross fixed capital formation and value added (current prices)	OECD-STAN
WAGE	Labour cost per employee	OECD-STAN
POP	Total population	OECD-STAN
PATSH	Share of national industry patents applications over the sum of the industry's patents applications	CRIOS-PatStat
d.w	Degree centrality (technological class co-occurrence network)	CRIOS-PatStat
ev	Eigenvector centrality (technological class co-occurrence network)	CRIOS-PatStat
am	Local clustering (technological class co-occurrence network)	CRIOS-PatStat
d.w.cit	Degree centrality (citation network)	CRIOS-PatStat
ev.cit	Eigenvector centrality (citation network)	CRIOS-PatStat
am.cit	Local clustering (citation network)	CRIOS-PatStat

Table 2.4: Cross-correlation Matrix

	XMS	PATSH	WAGE	INV	POP	d.w	am	d.w.cit	am.cit
XMS		0.13	0.19	-0.01	0.16	0.13	0.22	0.1	-0.06
PATSH			0.2	-0.16	0.95	0.79	0.22	0.68	-0.28
WAGE				0.09	0.16	0.13	0	0.19	-0.17
INV					-0.14	-0.13	0.01	-0.07	0.03
POP						0.78	0.24	0.67	-0.27
d.w							0.16	0.91	-0.22
am								0.12	0.12
d.w.cit									-0.2
am.cit									

Note: The correlation coefficients between export market shares (XMS) and its first and second lag are respectively: 0.84 and 0.73.

Moreover, to describe the relative position of industries in a national knowledge space, Figure 2.4 compares how industries in Italy in 2009 are ranked according to our network measures. It captures the possible heterogeneity in terms of centrality and local clustering among industries. The closer two triangles are in the plot, the bigger is the "difference" in terms of degree and clustering for a given industry. For instance, sector A (i.e., Food, Beverage and Tobacco) has a relative low degree but its embedded into a well connected cluster. Finally, Figure 2.5 describes the evolution over time of our network measures in Italy.

In the next Sections, we will investigate whether the network structure and the relative position of industries - as indicated by centrality and local clustering - in the space of knowledge flows is positively associated to export performances. Of course, beyond the national network of industries we also take into account international relationships (Laursen and Meliciani, 2002, 2010). As mentioned above, given the impossibility of using co-occurrence information, we characterize the latter dimension relying on patent citations.

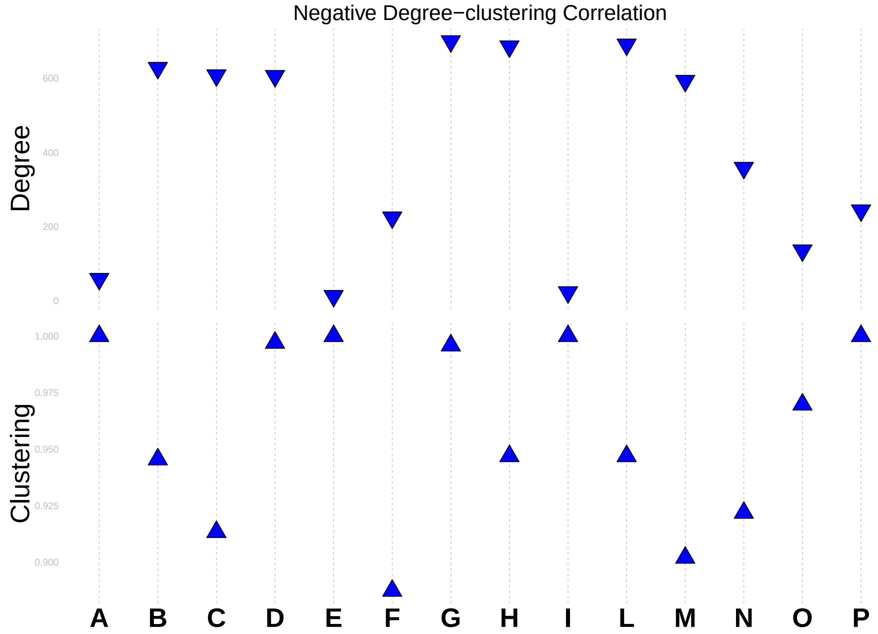


Figure 2.4: Weighted Degree and Weighted Local Clustering by Industry - ITALY 2009

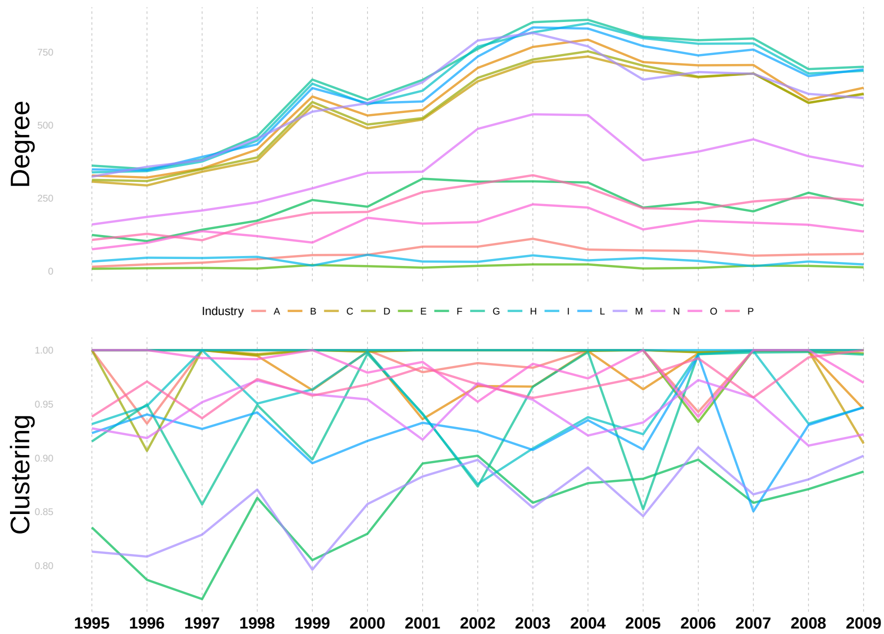


Figure 2.5: Weighted Degree and Weighted Local Clustering by Industry - ITALY

2.4.4 Econometric strategy

In this chapter we use two different econometric specifications. First, we follow Dosi et al. (2015) in exploring the link between export market shares and both technological and cost factors in a standard panel framework, with the obvious difference that we do not estimate the model in each industry separately because we are interested in inter-sectoral knowledge linkages. Secondly, once we have underlined the high persistence of export market shares over time, we move to a dynamic model with an autoregressive structure in the dependent variable akin to Amendola et al. (1993) and Laursen and Meliciani (2002). Both the two specifications may also have an evolutionary interpretation as specifying the selection dynamics linking "fitness" and expansions or contractions of export shares at the sectoral level. When a country is better in terms of cost and technology competitiveness relatively to its counterparts, it will increase its exports more than the counterparts. Fitness is captured both by cost competitiveness and technological features, notably including the relative position of each industry in the network of knowledge flows. This view also helps justify the choice of our dependent variable. Moreover, as reported in Laursen and Meliciani (2010), from an econometric point of view, exports normally grow over time (as world income does) and a variable measuring exports in absolute terms is very likely to be non-stationary. By contrast, an export market share variable is much more likely to be stationary, at least in the first moment.

The baseline model, from which we obtain the different specifications estimated in the chapter, is:

$$\begin{aligned} XMS_{ijt} &= \alpha_0 + \gamma XMS_{ijt-1} + \alpha_1 PATSH_{ijt} + \alpha_2 WAGE_{ijt} + \alpha_3 INV_{ij} + \\ &+ \alpha_4 POP_{ijt} + \beta_1 d.w_{ijt} + \beta_2 am_{ijt} + \beta_3 d.w.cit_{ijt} + \beta_4 am.cit_{ijt} + \\ &+ \eta_{1i} + \eta_{2j} + \eta_{3t} + \epsilon_{ijt}, \end{aligned} \tag{2.4}$$

where the coefficients α_h are associated to the standard control variables, β_h capture the effects of industries' positions in the inter-sectoral knowledge network and η_h represent different kinds of fixed effects we control for. Moreover, in many specifications we introduce the interaction effect between our network centrality and local clustering measures, for the national or international networks. This allows us to test whether, for an industry, the importance of being in a central position with respect to the flows of knowledge diminishes as long as it becomes more and more embedded in a tied cluster. All the variables, with exception of fixed effect dummies, are in logarithms and vary in the cross-sector, cross-country and cross-time dimensions. When an estimated coefficient in our model obtains a positive sign (as we expect in the majority of cases) this implies that when the country increases (decreases) its relative technology (knowledge flows, investment, etc.) in a given industry, the country increases (decreases) its market share in that industry. As it is standard in the literature, we expect unit labor costs to have a negative impact on export share dynamics (although this effect could be null considering that the dependent variable is expressed in current prices), while technology variables to have a positive effect on export share. The novelty of the work consists in the analysis of the role of technological factors in sustaining international competitiveness, with a particular emphasis on the effects driven by industries' position in the networks of knowledge flows and distinguishing between national and international flows.

The estimation strategy we adopt clearly differs in the case we test a dynamic model with an autoregressive component or we remain with the baseline model proposed in Dosi et al. (2015), which is simply obtained imposing $\gamma = 0$. When estimating the specification that does not consider an autoregressive element we start by pooling OLS with sector, year and country dummies. However, as it is well known, the presence of unobserved heterogeneity possibly correlated with other regressors makes our estimates biased and inconsistent; furthermore, the number of cross-sectional observations in our sample is rather restricted. To attenuate these two problems, and considering that our main variables of interest have been shown in Section 2.4.3 to vary, often considerably, over time, we estimate our model using a Fixed Effect (FE) within estimator.

Of course we know that failing of the strict exogeneity assumption would make our FE estimator inconsistent. In our context, in particular, the presence of a dynamic structure in the true data generating process is likely (Amendola et al., 1993; Laursen and Meliciani, 2002, 2010). This would imply some degree of persistence in the competitiveness of industries, suggesting that path dependence might play a non trivial role. Moreover, as a rough observation, we report that unconditional correlation between XMS and its first and second lags is relatively high (see Table 2.4). Since we use a within estimator, in presence of long enough samples, the asymptotic bias we might incur in is well known to converge to zero under suitable stability conditions. However, we only have $T = 14$ periods, which make it difficult to argue in favor of a sufficiently small bias. To account both for an autoregressive component in our model specification and to solve the presence of such a negative bias (under the assumption that $\gamma > 0$) affecting the within estimator, we move to a different strategy. In particular, we use the Blundell-Bond (BB) Generalised Method of Moments (GMM) estimator, which gives consistent estimates provided that there is no second order serial correlation among the errors, and we report tests for first and second order autocorrelation. This BB-GMM specification is preferred to the original Arellano and Bond estimator due to the high persistence in the series (see discussion in Blundell and Bond, 1998 and Laursen and Meliciani, 2010). We assume, as it is standard in this literature, exogeneity of all explanatory variables. The exogeneity of relative prices is a common hypothesis in estimating export equations and is based on the idea that the export supply price elasticities facing any individual country are infinite. Technology variables are assumed to be exogenous since they should capture structural characteristics that may respond only very slowly to changes in export shares.

2.5 Technological centrality and national boundaries

Our propositions have been confirmed by the empirical analysis. Indeed, results support both conjectures concerning the centrality of industries in the inter-sectoral knowledge space as well as the greater role of the national dimension of technological flows. Our empirical strategies (pooled OLS/FE and GMM) coherently show that centrality and clustering in the national network positively associate with export market shares, the effects are significant and the interaction term displays a negative sign.

More in detail, table 2.5 presents the estimates of both pooled and FE model specifications, taking into account national (columns 1-2) and international (columns 3-4) boundaries. Interestingly, including both geographical dimensions (columns 5-6), centrality measures, operationalized by means of weighted degree and local clustering, appear to positively and significantly explain export performances. Furthermore, only national-wise measures yield significant estimates, maintaining the existence of geographical boundaries to the diffusion of knowledge flows. The negative sign of the interaction term between the two network measures, instead, provides support for our intuition: being central in a knowledge space is far less relevant when the industry is highly connected within a cluster. Such estimates remain robust after including country, industries and year dummies.

As mentioned in the previous Section, despite being informative, static models fail to ensure unbiased estimates within this setting due to persistence in industries' export performances. To overcome methodological difficulties, we chose to employ the dynamic panel estimator (a.k.a. Blundell-Bond estimator), introduced in Section 2.4.4. Estimation results obtained using a model with an autoregressive component are collected in Table 2.6. The first two columns refer to national and international baseline model specifications. Both our propositions are robustly confirmed, even within the dynamic setting with time dummy included. However a remarkable difference applies. When the persistent nature of export performance is conveniently taken into account (i.e. including lags), it emerges that the effect of technological variables is captured by industry's relative position in the national network of knowledge flows, which is expressed through its centrality and local clustering, and by persistence in export performances, while the effects of patenting activities is not significant anymore. Such evidence points to the usefulness of our approach in capturing relevant information concerning knowledge generation and diffusion.

The redundancy of being central and well locally-clustered is confirmed and, further, we find that when both national and international network measures are included (column 2), just the former produce a significant effect on competitiveness. However, it is worth recalling that they are constructed using different data sources (IPC co-occurrence vs. patent citations), which might make the two set of regressors not fully comparable. To tackle such an issue, we have run a robustness check (columns I and II) using citations to construct both the national and international knowledge space[17]. The standardization of network measures' derivation, while dispelling any operational concerns, does not alter results, which remain fairly robust. Additional robustness checks in a static setting, related to the choice of a different centrality measure, can be found in the Appendix. The Arellano-Bond test for autocorrelation is performed and reported for each and every specification as well as Hansen-Sargan for the validity of the instruments. As a matter of fact, centrality measures as well as their interaction keep behaving as expected.

Our network approach, however fairly simple, has proven particularly useful to conclude that centrality plays a crucial role in explaining industries' export performances and that geographical proximity is a firm moderating factor.

[17]Figure A.1 in Appendix A shows how our measures (including the ones derived from national citation networks) correlate with each other.

Table 2.5: Regression Results for Model without Autoregressive Component

	Dependent variable: XMS					
	National (Co-occurrences)		International (Citations)		Final (Co-occurrences and citations)	
	(pooled)	(FE)	(pooled)	(FE)	(pooled)	(FE)
PATSH	0.071***	0.387***	0.074***	0.350***	0.064***	0.350***
	(0.021)	(0.068)	(0.021)	(0.067)	(0.021)	(0.070)
WAGE	0.016***	0.010*	0.013***	0.008	0.014***	0.008
	(0.004)	(0.005)	(0.003)	(0.005)	(0.003)	(0.005)
INV	0.046***	0.029***	0.047***	0.030***	0.045***	0.028***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
POP	−0.058*	0.002	−0.033	0.014	−0.049	0.007
	(0.032)	(0.029)	(0.032)	(0.028)	(0.032)	(0.029)
d.w	0.018***	0.013**			0.012**	0.010*
	(0.006)	(0.005)			(0.006)	(0.005)
am	0.076***	0.062**			0.076***	0.061**
	(0.029)	(0.027)			(0.029)	(0.027)
d.w.cit			0.018	0.026*	0.005	0.021
			(0.011)	(0.014)	(0.013)	(0.017)
am.cit			0.014	0.153	−0.103	0.089
			(0.116)	(0.130)	(0.141)	(0.168)
d.w:am	−0.015*	−0.014*			−0.014*	−0.014*
	(0.008)	(0.008)			(0.008)	(0.008)
d.w.cit:am.cit			−0.017	−0.031	0.0005	−0.022
			(0.017)	(0.021)	(0.019)	(0.025)
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,778	2,778	2,811	2,811	2,778	2,778
R^2	0.561	0.379	0.561	0.385	0.563	0.381
Adjusted R^2	0.551	0.347	0.551	0.353	0.553	0.349
F Statistic	69.796***	73.933***	71.934***	76.847***	67.627***	65.271***
	(df = 50; 2727)	(df = 21; 2546)	(df = 49; 2761)	(df = 21; 2579)	(df = 52; 2725)	(df = 24; 2543)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2.6: Regression Results for Model with Autoregressive Component (Blundell-Bond Estimator)

	Dependent variable: XMS			
	Baseline (Co-occurrences and citations)		Robustness (Citations)	
	(1)	(2)	(I)	(II)
XMS_{-1}	0.956^{***} (0.027)	0.956^{***} (0.027)	0.953^{***} (0.032)	0.954^{***} (0.031)
XMS_{-2}	-0.028^{*} (0.016)	-0.029^{*} (0.016)	-0.047^{***} (0.015)	-0.046^{***} (0.015)
PATSH	−0.002 (0.004)	−0.003 (0.004)	−0.002 (0.005)	−0.004 (0.005)
WAGE	0.0004 (0.001)	0.00005 (0.001)	0.001 (0.001)	0.001 (0.001)
INV	0.001 (0.004)	0.001 (0.004)	0.002 (0.005)	0.001 (0.005)
POP	0.002^{***} (0.0003)	0.002^{***} (0.0003)	0.002^{***} (0.0004)	0.002^{***} (0.0004)
d.w	0.005^{**} (0.002)	0.004^{**} (0.002)		
am	0.030^{***} (0.009)	0.029^{***} (0.009)		
d.w.cit		0.00005 (0.0004)		−0.0004 (0.001)
am.cit		−0.033 (0.026)		−0.030 (0.029)
d.w:am	-0.008^{***} (0.003)	-0.007^{**} (0.003)		
d.w.cit.control			0.005^{**} (0.002)	0.005^{**} (0.002)
am.cit.control			0.032^{**} (0.012)	0.029^{**} (0.012)
d.w.cit.control:am.cit.control			-0.008^{**} (0.003)	-0.007^{**} (0.003)
Time Dummies	Yes	Yes	Yes	Yes
Observations	2811	2811	2811	2811
AR(order1)	-5.23^{***}	-5.24^{***}	-5.29^{***}	-5.30^{***}
AR(order2)	−1.54	−1.55	−1.01	−1.09
Wald Test (coef.)	7703.89^{***}	8047.81^{***}	6071.79^{***}	6258.51^{***}
Wald Test (int.)	477.00^{***}	486.01^{***}	422.93^{***}	416.08^{***}
Sargan/Hansen (χ^2)	55.02 (df = 100)	55.01 (df = 102)	12.89 (df = 96)	12.94 (df = 98)

Note: *p<0.1; **p<0.05; ***p<0.01. The robust one-step GMM estimator is used. The number of lags used to instrument the endogenous variable go from the fourth onwards.The first two lags of our dependent variable are included in model specification. Time dummies included when specified but coefficients not reported. AR (1) and AR (2) are Arellano-Bond tests that average autocovariance in residuals of respectively order 1 and 2 are zero. Wald tests for intercepts and slopes suggest rejection of homogeneity. Sargan/Hansen accounts fro the validity of the instruments.

2.6 Discussion

This work proposes a novel factor that affects the international competitiveness of industries: the position of an industry in the inter-sectoral knowledge space. The recent literature has suggested that innovation and technological change are more relevant than cost-related factors in explaining industries' competitiveness, coherently with the interpretation of trade as a partial-disequilibrium process where heterogeneous firms compete, innovate, specialize and transfer knowledge across time and space in an imperfect and often unpredictable manner. In such a context, the impact of cost-based factors is limited. This analysis adds to stream of contributions suggesting that innovation, R&D activities and the stock of knowledge are relevant determinant of competitive advantage (Dosi et al., 2015; Laursen and Meliciani, 2000, 2002, 2010); beyond such indicators of innovation stock, we find that the position of industries within the inter-sectoral flows of knowledge matters. From our estimates, it is not the innovative effort of an industry or the direct knowledge links among industries that affect international competitiveness when the position within the inter-sectoral knowledge flow is accounted for. Rather, industry's performance is robustly and positively affected by the being central and locally well connected to other industries' knowledge stocks. Notably, our results suggest that competitive advantage positively relates to the position of an industry in the national (rather than international) knowledge space: being conveniently located within the streams of knowledge generated within the country matters more than being so in the whole knwoledge space.

Shortly, we find that (i) industries that are more central in the inter-sectoral knowledge space of their respective countries outperform their foreign competitors and that (ii) the relevant geographical dimension in determining such an effect is the national one. To obtain such results we have combined the use of firm level patent data - which have been duly aggregated into sectoral variables - and industry

- level data on exports and costs for a set of 16 OECD economies over a time span of 15 years (1995-2009). Results from our regressions robustly confirm that trade performance is positively affected by network measures characterizing the position of industries in the knowledge space. In particular, being either central or clustered in the network of knowledge flows (the knowledge space) significantly boost export market shares. However, these two effects are found to be redundant: being central in a knowledge space is far less relevant when the industry is highly connected within a cluster. Interestingly, such effects almost completely capture the role of industries' innovativeness, which turns out not to be significant when our network measures are included in the model (see tables 2.5, 2.6). It must be emphasized that we do not claim that innovation activities per se are not important in explaining trade competitiveness of industries; rather, we point out that with respect to knowledge and innovation, our approach leads to develop a variable which is more informative than the patent share of an industry, which completely neglects inter-sectoral flows of knowledge. In addition, our second proposition finds confirmation in the results which suggest that the most relevant network for an industry - i.e. the network where being central matters - is the one of national knowledge flows. We also provide some possible explanations regarding the role of the inter-sectoral knowledge space (see Section 2.3). A first mechanism involves the concept of variety of opportunities (see for example Boschma (2005) and Balland et al. (2015)): being a central industry in the flows of knowledge across industries enlarges its opportunities to innovate and to eventually exploit such innovations on the market. A second mechanism points to recombination (Ferguson and Carnabuci, 2017): if innovation requires the recombination of knowledge, then being exposed to several different technological flows coming from different industries may significantly increase innovativeness and, hence, the possibility to benefit from them in the market. Third, being exposed to a variety of knowledge flows coming from other industries may help an industry to develop technological as well as managerial capabilities to perfectly master different technologies and eventually match them with the industry's application and market context (Christensen et al., 1998; Cohen and Levinthal, 1990). Similarly, regarding the importance of the national dimension of knowledge flows we believe that the localized nature of knowledge flows and the presence of "home market bias" effects (Vernon, 1966; Linder, 1961) offer reasonable explanations for the larger importance of national rather than international connections.

More generally, our claim that industries have direct and indirect knowledge relationships with other industries which positively affect international competitiveness point to a still rather unexplored dimension of innovation and technological change: the various ways in which industries are tied together and affect each other in terms of knowledge, innovation and performance. This can be related to the broader issues of what constitutes an industry knowledge base and which are the various direct and indirect inter-sectoral channels which feed and generate this knowledge base (Breschi et al., 2003; Malerba, 2002; Dosi and Nelson, 2010). In fact, knowledge in an industry does not automatically spills over from its "production" within the industry (Dosi, 1988; Dosi et al., 2015), but it may originate and diffuse in various ways and through various channels from other industries: through vertical linkages (Hirschman, 1958; Lundvall, 1992); tacit knowledge flows (Breschi and Lissoni, 2001), movement of people and new firms that carry knowledge across industry boundaries (Adams et al., 2018) or broader links and inter-sectoral relationships at the organizational or institutional or organizational level, such as diversification or vertical integration (Helfat and Campo-Rembado, 2016; Li et al., 2018).

All such elements point to interesting areas for future of research. First, it is important to examine in detail and empirically assess the relevance of the various mechanisms proposed in this chapter through which inter-sectoral knowledge flows affect the competitiveness of an industry. Second, our analysis is focused only on 14 industries. More disaggregated analysis with more fine grained data is necessary. For instance, regional level data would provide useful information to further investigate to what extent geographical boundaries matter - we only distinguish beetween national versus international flows. Third, the number of countries examined in this work is limited and focuses on OECD countries. Our reasoning does not necessarily holds for several emerging countries in which some local industries are not developed and therefore are not present.

In conclusion, this chapter adds a novel insight to the analysis of export performance of countries and has also interesting implications for public policy. For countries, it is indeed important to promote and raise innovation and R&D in industries. However, we support the idea that they should also foster inter-industry collaborations among firms and links among industries. This second type of policy complements and does not substitute the first one: only industries and firms that are innovative and do R&D are able to benefit from inter-industry knowledge flows and increase their international competitiveness. Finally, geographical boundaries must be taken into account if we want to design effective policies.

Chapter 3

Relatedness, collaborations, and research diversification

This chapter is largely based on the following published article: Tripodi et al. (2020)

3.1 Introduction

The activities of scientists and innovators often span several areas, with choices of research endeavours driven by a variety of factors. The "essential tension" between exploration and exploitation described by Kuhn certainly characterizes research careers (Kuhn, 1977), but scientists can evolve ways to handle this trade-off. On the one hand, advances in science and technology create a "burden of knowledge" (Jones, 2009); the sheer amount of information required to move forward has grown, and larger educational costs may force scientists and innovators towards a narrower specialization. On the other hand, contemporary science is dominated by teams that bring together different expertise - albeit at a cost in terms of coordination and credit sharing (Wuchty et al., 2007). This chapter focuses on the analysis of scientists' research portfolio, investigating the roles of knowledge relatedness (among research topics) and social relatedness (among authors), as well as their interaction, as drivers of diversification.

Recent efforts to better characterize patterns in research and innovation activities produced valuable insights. For instance, based on a knowledge network created using MEDLINE articles annotated with chemical entities, Foster et al. (2015) quantitatively analyzed the dichotomy between exploration and exploitation. According to their taxonomy, each new article can expand or consolidate the knowledge space by generating a new chemical relationship (i.e., a new combination) or contribute to an existing one. Results show that research strategies (i.e., the types of articles produced) are stable over time and exploitation is preferred over exploration, despite a growing number of opportunities. Exploration is riskier, with rewards (i.e., citations) that are higher but insufficient to compensate the risk. In the domain of physics, Pan et al. (2012) focused on the temporal evolution of interdisciplinary research. The authors constructed and analyzed yearly snapshots of the connections among physics sub-fields uniquely identified through PACS codes. Results show that connectivity, and thus interdisciplinarity within physics, increased - but in a non-random way that reflects the hierarchical structure of sub-fields. In particular, *condensed matter* and *general physics* acted as hubs for the increasing number of connections. Recently, Sun and Latora (2020) proposed a novel framework, based on time-varying networks, to track knowledge flows within and across physics sub-fields. Such a method is able to highlight the increasing general trend towards interdisciplinary research as well as identify interesting patterns of influence among sub-fields over time. More directly related to our purposes, Battiston et al. (2019); Aleta et al. (2019); Jia et al. (2017) collected compelling empirical evidence on physicists' research endeavours. Battiston et al. (2019) provided a comprehensive census of academic physicists active in recent decades. The authors charted a thorough picture of the evolution of various fields in terms of number of scientists, productivity (including impact and recognitions such as Nobel prizes), team size and role of chaperones - highlighting a rich heterogeneity among specializations. Moreover, Battiston et al. (2019) mapped "migration" flows by comparing the field in which a given scientist published her first paper with the one characterizing her later research interests.

Also Aleta et al. (2019) mapped flows among physics sub-fields, with the aim of investigating the "essential tension" in the evolution of scholars' research interests. The authors defined a measure of exploration comparing early- and late-career ranges of actives, and tracked flows using origin-destination matrices among fields. Results suggest a preference for exploration over exploitation, but concentrated within the same broad area of research, and non-random transitions among different areas. Jia et al. (2017) observed that the frequency of scientists decays exponentially as one considers increasing degrees of change in interests. In order to reconstruct the macroscopic patterns that drive such evolution, the authors proposed a random walk model over a stylized knowledge space, which reproduces empirical observations thanks to the inclusion of key features such as heterogeneity, subject proximity and recency. Finally, Zeng et al. (2019) analysed the dynamics of "topic switching" by exploring co-citing networks. Results suggest a growing propensity to switch among topics but also that such a strategy might hamper productivity, especially for early-career researchers.

Despite the growing body of evidence and stylized facts provided by this literature, much remains to be done to disentangle and quantify the roles of different contributing factors. To make progress in this direction, we investigate scientists' research portfolio diversification by quantifying potential drivers of exploration, or, to put it differently, the hurdles faced by scientists when they move out of their immediate specialization. We use a network approach to compute a measure of similarity among research sub-fields, define a measure of social relatedness and track scientists' diversification patterns. We build our empirical strategy upon the intuition of Breschi et al. (2003), who used patent data to explore the nature and degree of coherence in firms' technological diversification.

Our analysis proceeds as follows. First, we test and reject the hypothesis that research portfolio diversification is random. Second, we use regression techniques to characterize how subject and social proximity affect diversification, controlling for possible confounding factors. Third, we quantify the relative importance of our relatedness measures. We provide robust empirical evidence that knowledge and social relatedness are both significant statistical predictors of diversification, as is their interaction - which corroborates the notion that collaborations modulate knowledge acquisition, especially when scientists move far from their own specialization. Like many of the articles mentioned above, we analyze data concerning physicists. This focus is due in part to the central role of physics among the *hard* sciences, and in part to the reliability of data collected labeling articles through the PACS codes. Nevertheless, our approach is fully general and could be used in different domains.

3.2 APS overview

We use the American Physical Society (APS) dataset to reconstruct the activities of 197,682 physicists who published at least one paper in one of the APS outlets in the period ranging from 1977 to 2009 (see Section 3.9.1 for details). All articles in APS journals are classified according to hierarchical codes that map into physics fields and sub-fields (i.e., PACS codes). For our analyses (see Section 3.9.5), we filter out authors and sub-fields that appear only sporadically in the data. Specifically, we focus on 105,558 authors who published at least two articles, covering a minimum of two sub-fields over a restricted set of 68 PACS which appear in at least four articles.

Figure 3.1 provides a general description of the data and some insights. Figure 3.1-**a** shows the popularity, in terms of number of articles, of fields and sub-fields (one- and two-digit level PACS codes, respectively). As expected, PACS popularity is highly heterogeneous and reflects the prominence of *condensed matter* research in the last decades. Figure 3.1-**b** shows scientists' degree of diversification and their relative specialization, as defined in Section 3.9.3. The research portfolio of most scholars in our dataset is fairly limited in scope, with a large majority of scientists

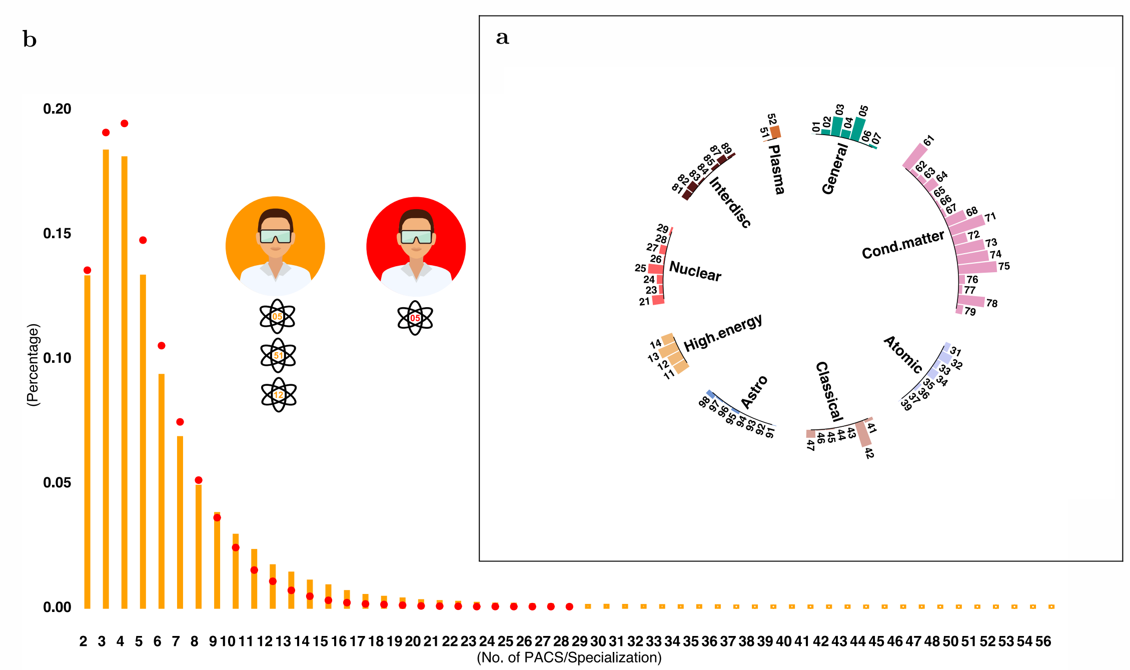


Figure 3.1: Popularity of fields and scientists' degree of diversification/specialization. **(a)** Circular bar-chart showing the number of articles assigned to each sub-field in the one-digit PACS codes, taking into account their hierarchical structure. The chart highlights the popularity of *Condensed Matter* research in both size and scope. **(b)** Distribution of scientists' degree of diversification (the number of sub-field they explored; orange bars) and of their relative specialization (the number of sub-fields in which they have a scientific advantage; red dots). Scientists explore several sub-fields, but specialize in only a few - despite the existence of some individuals with a truly interdisciplinary path, by and large research portfolios are fairly limited in scope. Inset: pictorial description of a scientist who explored three sub-fields (orange) but has only one specialization (PACS 05: red).

diversifying in no more than 5 sub-fields. The choice of subjects, however, is not random - as we demonstrate in the next Section.

3.3 Diversification is not random

Do scientists, much like firms (Teece et al., 1994; Breschi et al., 2003), shape their research portfolios based on specific strategies and constraints? To address this question quantitatively, we draw a parallel with ecology: as species may co-occur in distinct sites, sub-fields may overlap in research portfolios. Measuring the relatedness of species based on their geographical co-occurrence is analogous to measuring the relatedness of sub-fields based on their overlap in scientists' ranges of activity.

Thus, the PACS-Authors binary bipartite network resembles a presence-absence matrix (Veech, 2013). The monopartite projection of this bipartite network (see Section 3.9.2) on the PACS layer carries a critical piece of information: for each pair of PACS, it tells us how many scientists are active in both sub-fields irrespective of the number of articles, drawing a diversification network.

We can assess this network contrasting it against an appropriate null model. Which sub-fields overlaps are over- or under-represented relative to what we would expect under the assumption that scientists picked research topics at random, but taking into account the popularity of sub-fields? Under a random model, the probability that x scientists are active both in sub-field a and in sub-field b, given that S_a and S_b scientists are active in these sub-fields, follows a hypergeometric distribution (Tumminello et al., 2011)

$$P(X = x) = \frac{\binom{S_a}{x}\binom{S-S_a}{S_b-x}}{\binom{S}{S_b}} \tag{3.1}$$

where S is the total number of scientists in the sample.

Figure 4.3 describes the steps of our procedure. Starting from the bipartite network (panel **a**), we derive its monopartite projection (panel **b**) and test whether the resulting structure is non-random, summarizing statistically validated diversification patterns (panel **c**). Out of 2,278 pairs of PACS, 72% are classified as non-random with a Bonferroni-corrected p-value < 0.05. Of these, 1,151 pairs show a positive association and 486 a negative one. Given the severity of the Bonferroni correction (i.e., power decreases significantly as the number of tests increases) and possible issues related to dependency, we also employ the *False Discovery Rate* (FDR): Benjamini-Hochberg and Benjamini-Yekutieli correction (see Section B.2 and Table B.2). These results strongly support a coherent nature of scientists' diversification choices, but do not provide a direct quantification of the role played by specific features in shaping such coherence. Next, we investigate potential drivers of diversification considering measures of cognitive and social proximity.

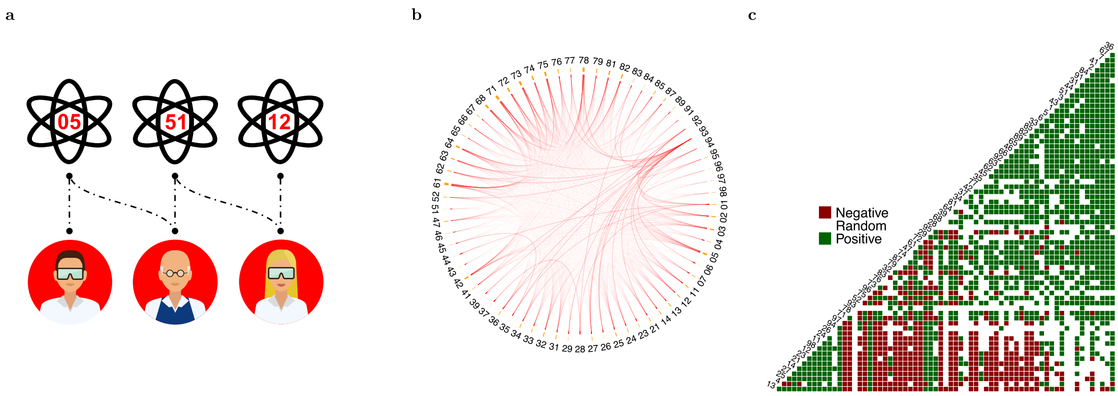


Figure 3.2: Diversification patterns. **(a)** A stylized picture of the original PACS-Authors bipartite network representing scientists' diversification patterns. **(b)** The diversification network (the monopartite projection on PACS): links represents the number of scientists active in each pair of sub-fields. **(c)** Visual summary of the hypergeometric test, providing evidence of the coherent nature of scientists' diversification choices: 72% of pairs are classified as non-random ($p < 0.05$ after Bonferroni correction).

3.4 Knowledge and social relatedness predict diversification

The relationships among scientific fields, like those among technologies, can be mapped using network science tools. To chart a knowledge space we need a measure of distance between fields. Several different metrics have been proposed to quantify the relatedness of technologies or scientific domains (see Bowen and Jianxi (2016) for a review). When we consider the monopartite projection on the PACS layer of the bipartite PACS-Articles network, counting the co-occurrences of all pairs of PACS produces a first approximation of the relatedness of sub-fields. A similar approach was used in Lamperti et al. (2019) for patent data. However, we need a measure of proximity that: (i) does not depend on the absolute popularity of the fields, and (ii) is symmetric. The most straightforward metric that fulfils both requirements is the cosine similarity (see Figure 3.3-**a**/**b**/**c**, Section 3.9.4). As expected, the proximity matrix has a clear hierarchical block structure, with blocks largely overlapping with fields. Interestingly, several off block elements show the proximity of sub-fields belonging to different PACS fields.

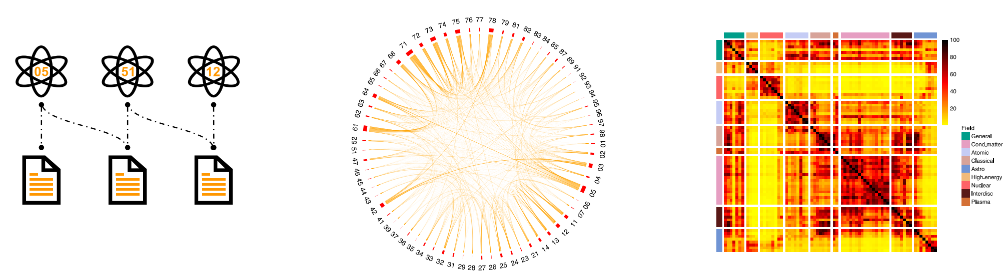
As science becomes an increasingly "social" enterprise, it is also important to capture the relatedness of scholars, which can be done by analysing co-authorships (Wuchty et al., 2007). Similar to what we did for knowledge relatedness, we construct a measure of social relatedness starting from the bipartite Authors-Articles network. The monopartite projection on the Authors defines the co-authorship network from which we compute our desired metric. In addition, to investigate whether diversification is associated with the exploitation of social relationships, we include information on authors' specialization as node attributes in the network and we introduce a dummy SR_{ib} equal to 1 if scientist i can reach sub-field b through direct social interactions (see Figure 3.3-**d**, Section 3.9.4).

Next, we evaluate the effects of knowledge and social relatedness on diversification with logistic regression. The binary dependent variable encodes whether a scientist is active in a sub-field, the main explanatory variables are our measures of cognitive and social proximity, and a control is introduced for the core field. In practice, each scientist is assigned to a core sub-field (specialization) and can possibly diversify in one or more target sub-fields different from her own (see Section 3.9.3). In this first set of regressions, each scientist appears 67 times, one for every possible target PACS different from her own specialization (see Section 3.9.5 for more details).

Figure 3.4 provides evidence that both social and knowledge relatedness are associated with scientists' diversification strategies. Social relatedness matters irrespective of the field, as scientists who can acquire new knowledge through social relationships are more likely to be active in a sub-field different form their own specialization (panel **a**). Also knowledge relatedness increases the probability of a scientist being active out of her own specialization, and again this is true for all fields (panel **b**). These results strongly suggest that cognitive and social proximity do contribute to shaping diversification strategies.

Knowledge relatedness

a b c



Social relatedness

d

Scientist	Specialization	Co-authors
Alice	45	Bob, Cecil, David, Paula
Bob	48	Alice, Cecil, David
Cecil	33	Alice, Bob
David	34	Alice, Bob
Paula	21,45	Alice

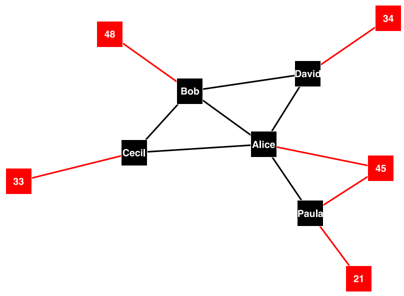


Figure 3.3: Knowledge and social relatedness. **(a)** A stylized example of the bipartite PACS-Articles network. **(b)** The PACS co-occurrence network (monopartite projection on PACS codes). **(c)** The cosine similarity matrix, which "maps" the physics knowledge space and identifies clusters corresponding to fields. **(d)** A table illustrating how co-authorship and specialization information are combined to produce the augmented co-authorship network shown in the figure, which includes nodes attributes (specializations). The nodes represent individual scientists (in black) and specializations (in red). Our measure of social relatedness (SR_{ib}) is defined as a dummy that captures whether scholar i can reach a certain sub-field b through social interactions; $SR_{ib} = 1$ if $d(i,b) = 2$, where $d(i,b)$ is the geodesic distance between scholar i and sub-field b. For instance, $SR_{David,45} = 1$ since David could directly exchange knowledge with Alice (specialized in sub-field 45), while $SR_{David,21} = 0$.

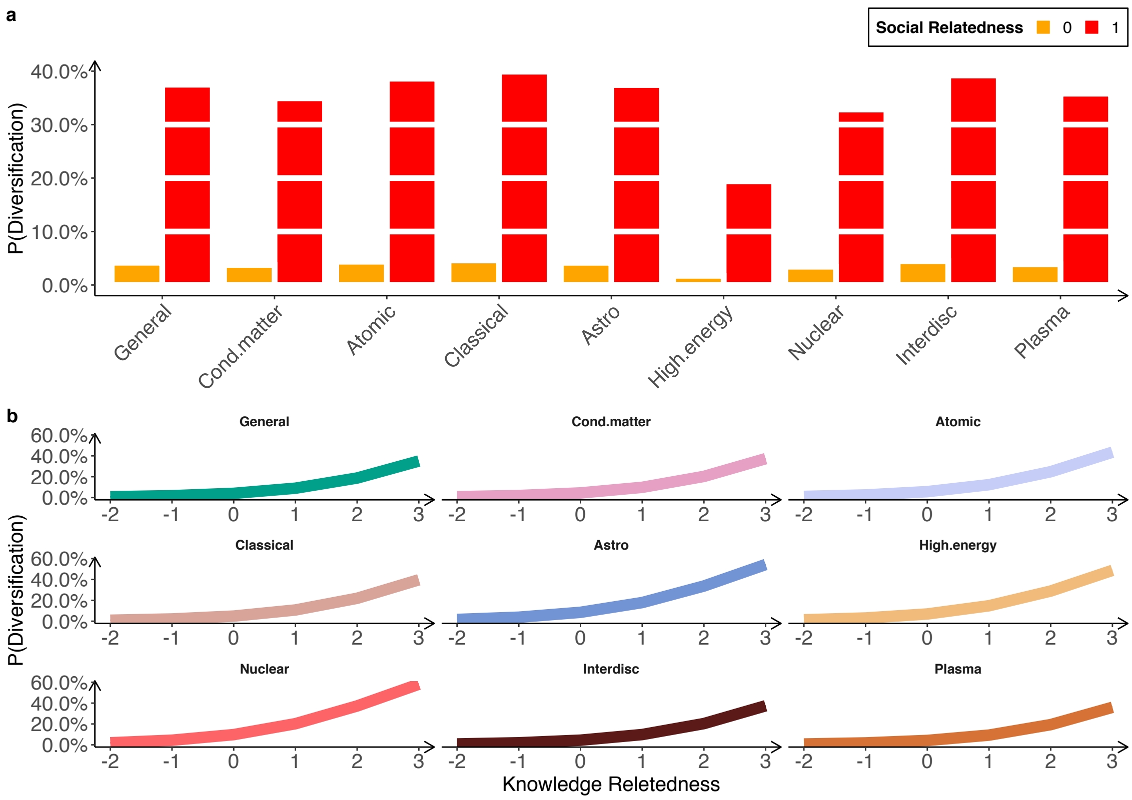


Figure 3.4: Probabilities of scientists diversifying in a sub-field different from their own specialization. Predicted probabilities of a scientist being active in a sub-field different from her own specialization as a function of **(a)** (binary) social relatedness, and **(b)** (standardized) knowledge relatedness. Results are obtained by fitting a logistic regression with only one control variable - the scientist' core field. All coefficients are statistically significant ($p < 0.01$).

3.5 Model extensions and robustness checks

To move further in our investigation of research portfolio diversification, we broaden our analysis in several ways. *First*, we expand our logistic regression model including a larger set of control variables, such as the number of co-authors or the popularity and citations of the target sub-field (see Table B.3 for a complete list). All numerical variables in the expanded model are normalized, and log-transformed to reduce right-skew when necessary (see Section 3.9.5 for more details). Since the effect of knowledge relatedness on the probability of diversification may be modulated by social relatedness, we also include an interaction term in our analysis.

Second, we tackle two potential limitations of our original analysis; that is, defining a single specialization for each scientist (while core specializations may actually be multiple), and not separating sub-field movements within and between fields, i.e., one-digit PACS codes (which may be differently affected by various features). We run additional model fits allowing scientists to have multiple specializations (see Section 3.9.3) and separating within and between field diversification. Specifically, we perform the following fits: (i) single specialization with full diversification, (ii) multiple-specialization with full diversification, (iii) single specialization with within field diversification, (iv) multiple specialization with within field diversification, (v) single specialization with between field diversification and (vi) multiple specialization with between field diversification.

Third, we account for the fact that the data employed in our fits are "clustered", with several observations associated to each scientist and a potential heteroskedasticity across clusters/scientists. We estimate clustering-robust standard errors using the clustered sandwich estimator from the **R** package *sandwich* (Zeileis, 2004).

Fits for specifications (i)-(iv), all including the interaction between knowledge and social relatedness and clustering corrected standard errors, are summarized in Table 3.1, confirming the high significance of the relatedness metrics in shaping research diversification. Figure 3.5 focuses on the full diversification case. Panels **a** (single specialization, (i)) and **c** (multiple specialization, (ii)) show the log-odds difference in the probability of diversification as a function of knowledge and social relatedness, accounting for all controls. Social relatedness positively affects the chances of diversification and the effect is moderated by knowledge relatedness in both specifications, though more markedly in (i) than in (ii). Panels **b** (for (i)) and **d** (for (ii)) further illustrate this, showing how the estimated coefficient of social relatedness decreases as knowledge relatedness increases. This result indicates that when diversifying toward "close" sub-field, the role of social relatedness becomes less crucial.

Table 3.1: Regression results. Coefficients of the logistic regressions of Eq. 3.8, i.e. the model including the interaction term between knowledge and social relatedness, under different specialization settings. The table reports clustering corrected standard errors (in parenthesis) and significance level:*p<0.1; **p<0.05; ***p<0.01.

	Dependent variable: Prob(diversification)					
	Full Diversification		Within Field Diversfication		Between Field Diversfication	
	single	multiple	single	multiple	single	multiple
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Knowledge Relatedness	0.936^{***}	0.688^{***}	0.184^{***}	0.121^{***}	0.702^{***}	0.511^{***}
	(0.003)	(0.009)	(0.005)	(0.013)	(0.003)	(0.011)
Social Relatedness	2.827^{***}	4.243^{***}	2.272^{***}	3.968^{***}	2.914^{***}	4.284^{***}
	(0.006)	(0.019)	(0.008)	(0.021)	(0.008)	(0.021)
field core-Atomic	-0.332^{***}	-0.428^{***}	0.056^{**}	-0.276^{***}	-0.303^{***}	-0.385^{***}
	(0.010)	(0.007)	(0.025)	(0.021)	(0.010)	(0.008)
field core-Classical	-0.490^{***}	-0.477^{***}	-1.001^{***}	-0.932^{***}	-0.313^{***}	-0.328^{***}
	(0.010)	(0.007)	(0.029)	(0.023)	(0.010)	(0.008)
field core-Cond.matter	-1.088^{***}	-0.761^{***}	-1.110^{***}	-0.892^{***}	-1.263^{***}	-0.903^{***}
	(0.012)	(0.009)	(0.024)	(0.020)	(0.017)	(0.013)
field core-General	-0.722^{***}	-0.537^{***}	-0.927^{***}	-0.823^{***}	-0.632^{***}	-0.422^{***}
	(0.011)	(0.007)	(0.028)	(0.021)	(0.012)	(0.008)
field core-High.energy	0.219^{***}	0.168^{***}	1.806^{***}	1.176^{***}	-0.360^{***}	-0.060^{***}
	(0.010)	(0.006)	(0.027)	(0.023)	(0.013)	(0.008)
field core-Interdisc	-0.557^{***}	-0.553^{***}	-0.357^{***}	-0.724^{***}	-0.365^{***}	-0.367^{***}
	(0.010)	(0.007)	(0.026)	(0.021)	(0.011)	(0.008)
field core-Nuclear	0.463^{***}	0.164^{***}	0.969^{***}	0.692^{***}	0.068^{***}	-0.161^{***}
	(0.010)	(0.006)	(0.024)	(0.021)	(0.011)	(0.009)
field core-Plasma	-0.269^{***}	-0.419^{***}	-0.155^{**}	-0.361^{***}	-0.074^{***}	-0.256^{***}
	(0.013)	(0.008)	(0.068)	(0.058)	(0.015)	(0.009)
# of PACS	0.882^{***}	0.806^{***}	0.769^{***}	0.497^{***}	1.003^{***}	0.944^{***}
	(0.002)	(0.003)	(0.005)	(0.004)	(0.004)	(0.004)
# of papers	0.010^{***}	0.113^{***}	0.065^{***}	0.252^{***}	-0.032^{***}	0.050^{***}
	(0.002)	(0.003)	(0.005)	(0.004)	(0.004)	(0.004)
# of co-authors	-0.406^{***}	-0.347^{***}	-0.240^{***}	-0.145^{***}	-0.488^{***}	-0.444^{***}
	(0.002)	(0.004)	(0.004)	(0.004)	(0.003)	(0.006)
PACS target popularity	1.130^{***}	0.611^{***}	1.370^{***}	0.774^{***}	1.108^{***}	0.559^{***}
	(0.002)	(0.002)	(0.005)	(0.003)	(0.003)	(0.002)
Δ crowd	0.239^{***}	0.358^{***}	0.131^{***}	0.345^{***}	0.320^{***}	0.393^{***}
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Δ PACS citations	-0.273^{***}	-0.332^{***}	-0.208^{***}	-0.313^{***}	-0.369^{***}	-0.354^{***}
	(0.002)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)
Δ field citations	-0.156^{***}	-0.070^{***}	✗	✗	-0.196^{***}	-0.143^{***}
	(0.004)	(0.004)			(0.006)	(0.005)
KR:SR	-0.255^{***}	-0.061^{***}	-0.047^{***}	-0.001	-0.234^{***}	-0.067^{***}
	(0.004)	(0.010)	(0.007)	(0.013)	(0.005)	(0.011)
Constant	-3.812^{***}	-5.903^{***}	-1.882^{***}	-4.250^{***}	-4.168^{***}	-6.165^{***}
	(0.010)	(0.020)	(0.022)	(0.028)	(0.010)	(0.022)
Observations	7,072,386	35,968,615	1,000,230	5,407,404	6,072,156	30,154,990
Log Likelihood	−1,086,281.000	−7,303,198.000	−334,697.300	−2,166,803.000	−716,398.900	−4,971,497.000
Akaike Inf. Crit.	2,172,600.000	14,606,434.000	669,430.600	4,333,642.000	1,432,836.000	9,943,033.000

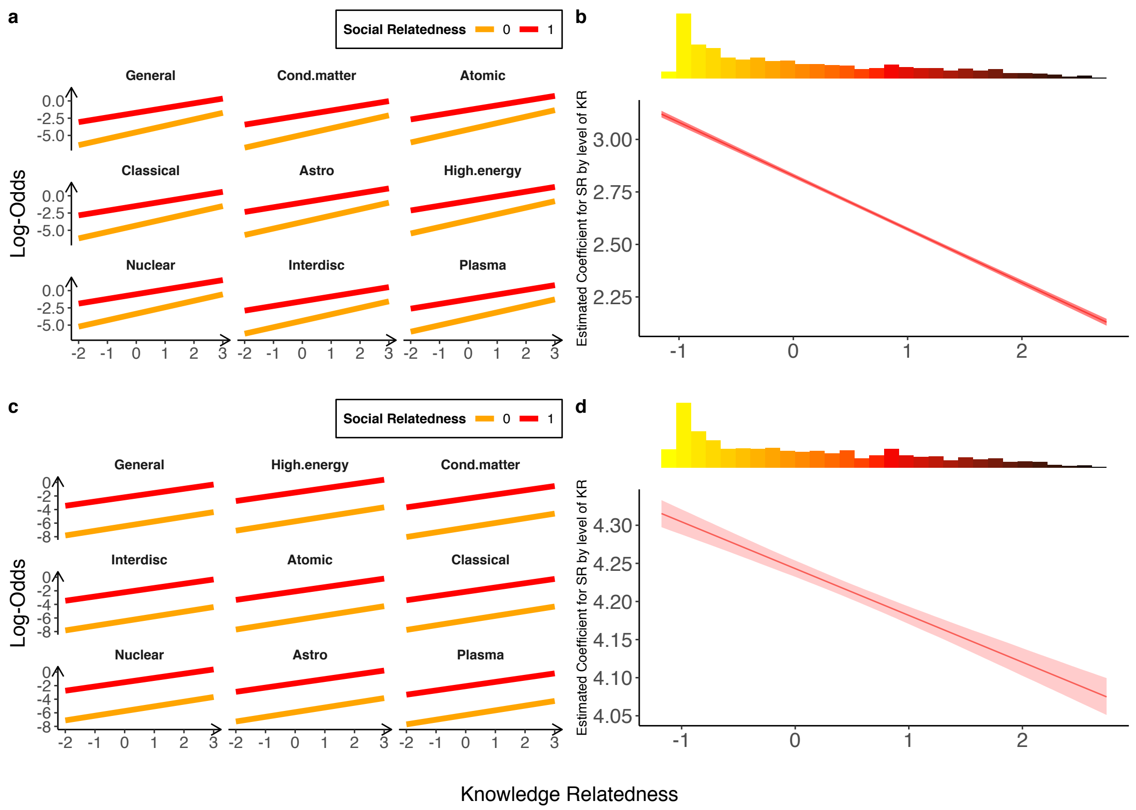


Figure 3.5: Scientists' research portfolio diversification: full diversification, single and multiple specialization. **(a)** Log-odds as function of (binary) social relatedness and (standardized) knowledge relatedness, accounting for multiple control variables, for the single specialization specification (i). **(b)** Estimated coefficient for social relatedness conditional on knowledge relatedness, and distribution of knowledge relatedness (on top, similarity color coded as in Figure 3.3-**c**), for the single specialization specification (i)). **(c), (d)** Same as (a) and (b) for the multiple specialization specification (ii).

Next, we contrast scientists moving within their specialization field (between two sub-fields, i.e. two-digit PACS codes, belonging to the same field, i.e. one-digit PACS code; e.g. PACS 12 *Specific theories and interaction models; particle systematics* and PACS 13 *Specific reactions and phenomenology*, both belonging to PACS 1 *High Energy* physics) and scientists moving out of their field and towards a completely different subject (i.e. a different one-digit PACS code). These choices may be driven by different factors. Scientists moving within their field may be less dependent on external collaborations, since such a diversification strategy requires a smaller learning effort. Our estimates do highlight differences. Looking at the within field diversification case, single specialization (Table 3.1, (iii)), we see that knowledge and social relatedness, as well as their interaction, are still significant - but the magnitude of the coefficients is smaller with respect to the full diversification case. When we consider multiple specialization (Table 3.1, (iv)), coefficients shrink even further and the interaction is no longer significant (see also Figure B.2). On the contrary, looking at the between field diversification case, the general trends outlined for the full diversification case are confirmed - including the negative interaction term remaining sizeable and significant for both single and multiple specialization (see Table 3.1, (v) and (vi), and Figure B.3). These results are in line with expectations: while having a co-author in a different sub-field may well be useful, knowledge is not a barrier to entry when scientists move within the same general area of inquiry. This explains why the interaction between social and knowledge relatedness becomes less prominent or non-significant in our estimates.

3.6 Quantifying the relative importance of knowledge and social relatedness

Can we quantify the (relative) role of knowledge and social relatedness in explaining research portfolio diversification? How important are these quantities when evaluated in the presence of several control covariates, and under a range of model specifications? To answer these questions we follow two approaches.

First, we run a LASSO feature selection procedure to gauge the relative importance and role of different predictors by tracking how they are excluded/included in a model as one varies the regularization penalty. Since our predictors include categorical variables (i.e., groups of dummies), as well as naturally grouped variables (e.g., scientists' individual characteristics, sub-fields' popularity and competition, etc.) we run a *group* LASSO algorithm (Yuan and Lin, 2006) with features grouped as shown in Table B.3. Moreover, to counteract collinearity and finite sample issues which can render the LASSO unstable (Mullainathan and Spiess, 2017), we split our data forming ten random subsamples of 1,000 scientists each, and repeat the group

LASSO fit on each of the subsamples for all the considered model specifications. Panels **(a)-(f)** of Figure 3.6 show the (grouped) coefficient norms as a function of the penalization parameter λ. Results clearly demonstrate the crucial role played by social and knowledge relatedness. They also confirm that the role of knowledge relatedness weakens markedly in the case of within-field diversification (panels **(c)** and **(d)**).

Second, we compute the *Relative Contributions to Deviance Explained* (RCDEs; see Section 3.9.5 for details). This index captures what percentage of the logistic regression deviance is captured by a predictor. Panel **g** of Figure 3.6 strongly supports a prominent role for social relatedness, with RCDEs around or above 30% across all specifications. The RCDEs of knowledge relatedness are smaller, around 5-10%, and again become negligible in the case of within-field diversification. In summary, our results provide additional evidence that both social and knowledge proximity shape scientists' diversification strategies, but highlight social interactions as the dominant channel through which knowledge is exchanged and acquired.

3.7 Digging deeper: multidisciplinarity and time

Next, we tackle two additional potential limitations of our original analysis, which might overestimate the probability of diversification for truly multidisciplinary scientists and suffer from reverse causality issues. To investigate diversification into truly unexplored sub-fields, we fitted the model specification (i) (see Section 3.5) considering scientists' specialization (see Section 3.9.3) and limiting their diversification choices to sub-fields in which they have no revealed scientific advantage (see Section B.5). To at least partially address causality in the effects of knowledge and social relatedness on diversification, we included a temporal dimension: we split the original dataset in three time periods, re-computed our measures of relatedness in each, and used them to predict scientists' diversification introducing time

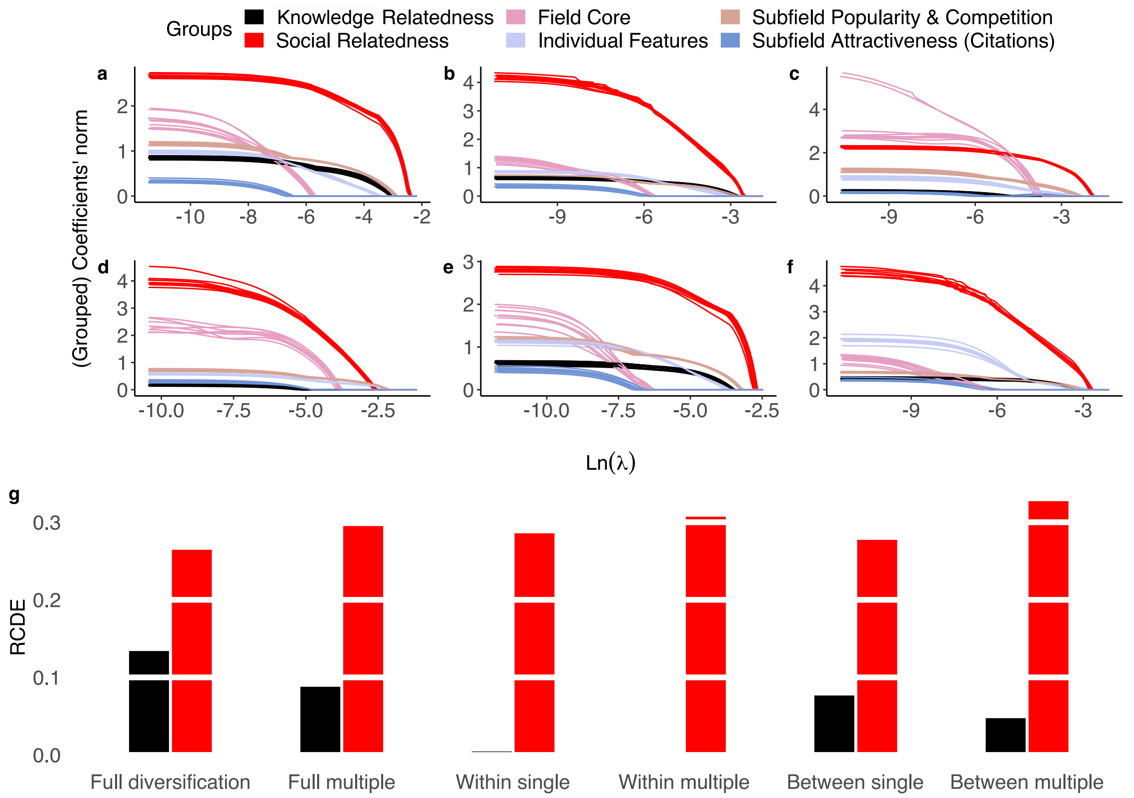


Figure 3.6: Relative importance of predictors. (a)-(f) Group LASSO paths for (a) full diversification, single specialization; (b) full diversification, multiple specialization; (c) within-field diversification, single specialization; (d) within-field diversification, multiple specialization; (e) between-field diversification, single specialization; (f) between-field diversification, multiple specialization. In each panel, variables in the same group are color coded, and their average coefficient norm is plotted (as a single path) against the penalty parameter (log λ). The multiple paths for each color correspond to separate group LASSO runs on 10 random sub-samples of 1,000 scientists. **(g)** Relative Contributions to Deviance Explained for knowledge relatedness (black) and social relatedness (red) across all fits.

lags (see Section B.5). In both exercises, results confirmed our previous findings: social relatedness shapes scientists' diversification strategies more than knowledge relatedness.

Finally, and again related to time, our findings may be influenced by underlying trends in the temporal evolution of PACS co-occurrence networks - and thus knowledge proximity. A detailed study of the evolution of relationships among subfields, which is of course of interest *per se*, is beyond the scope of the present article. Nevertheless, to gather at least some approximate sense of its potential impact, we recomputed our measure of knowledge relatedness separately for each of the different decades in the original dataset. Based on results shown in Section B.4, the physics knowledge space remained rather stable over the time span considered. A valuable alternative approach to take into account the temporal evolution of the physics knowledge space is provided by Chinazzi et al. (2019).

3.8 Discussion

Scientists try to balance the "tension" between exploitation and exploration, but the exploration phase is, to some extent, constrained by the "burden of knowledge". To tackle the rising complexity of producing new knowledge, scientists adapt their diversification strategies leveraging social interactions; that is, proximity to other scientists. Our analysis attempts to identify and quantify drivers of research portfolio diversification. Based on data concerning a very large sample of physicists we find that, while knowledge relatedness plays a role, contemporary science is a profoundly social enterprise. When scientists move out of their specialization, they do so through collaborations. And the further the move, the more these collaborations matter.

Limitations in the methodology we employed for this study point towards needed future developments. First and foremost, we are not assessing causal effects; we analyse research diversification patterns irrespective of the mechanisms which determine the similarity among sub-fields and the co-authorship network. Indeed, knowledge relatedness and collaborations may themselves be affected by scientists' diversification strategies. We believe that the observed negative interaction between knowledge and social relatedness helps us rule out, at least partially, the contingency of reverse causality for social relatedness: if diversification were causally driving the link, we would expect a positive interaction. There is no reason to believe that new collaborators are easier to find in sub-fields far from a scientist's own specialization; in fact, the opposite may be more likely - the closer the sub-fields, the higher the chances to collaborate. Moreover, since the structure of the knowledge space appears fairly stable over time, the direction of causality is more likely from subject proximity to diversification - not the other way around. Additional analyses with methods that fully exploit the temporal trajectories of scientists' activities will be instrumental to elucidating the causal interplay between individual strategies and collaborations. In the Supplement we do provide results for the checks we were able to run based on the data and methods at our disposal.

Another critical development will be expanding the investigation to scientific and/or technological domains beyond physics - shedding further light on behaviours and potential sources of heterogeneity. Our initial focus on physics was due to its central role in the natural sciences and to the availability of reliable and abundant data. Nevertheless, the approach used in this study is fully applicable to different domains. Patents and publications records would both be useful grounds to validate and extend our results - thus providing a quantitative benchmark to inform science and technology policy.

From a policy perspective, our current results already provide some insights. They support the notion that social interactions constitute the core medium to foster new scientific venues, allowing scientists to overcome knowledge barriers. Thus, social interactions should be a focus of efforts aimed at improving cross-disciplinary team formation. Institutions should strive to create environments that favor social proximity and collaboration, and funding for interdisciplinary research should reward matches among scholars specialized in very distant domains.

3.9 Data and Methodology

3.9.1 Data

We use the American Physical Society (henceforth APS) dataset, which is maintained by the APS and publicly available for research purposes upon request (see APS website).

Each article in the dataset is labeled with up to 5 PACS codes. As an example, the PACS code *42.65.-k* refers to *nonlinear optics*; the first digit represents a broad field (Classical Physics), and the second a more specific sub-field (Optics). A brief description of the one-digit level fields is provided in Table B.1. In our analyses, we work at the level of sub-fields; our measure of knowledge relatedness is based on similarity of PACS at two-digit level. Based on our aims (analysing research diversification strategies), we created a dataset based on two requirements: (i) the ability to reconstruct the career of each individual, and (ii) a standardized classification system for each article. (i) poses several issues related to name disambiguation, which have been successfully investigated in previous studies. We rely on the disambiguated dataset made available by Sinatra et al. (2016). (ii) concerns the classification scheme applied to physics articles. The PACS classification has been broadly employed from 1970 to 2016, but then the APS adopted a different labelling procedure (Physics Subject Headings; PhySH). We limit our analysis to a period entirely covered by the PACS system. Our final dataset includes information regarding 197,682 scholars that published at least one article in one of the 9 APS journals in the period ranging from 1977 to 2009. Figure B.1 shows the number of papers (panel **a**) and the number of papers per author (panel **b**) over time.

3.9.2 Monopartite projections of bipartite networks

A bipartite network is a graph whose nodes can be divided into two distinct sets (layers) such that no edge connects a pair of nodes belonging to the same set. A binary undirected bipartite network is identified by a rectangular biadjacency matrix **b** of dimensions $N_R \times N_C$. The number of rows N_R is the number of nodes in layer R, and the number of columns N_C is the number of nodes in layer C (Saracco et al., 2017). Being binary simply means that the elements of the matrix are

$$b_{rc} = \begin{cases} 1 & \text{if node } r \in \text{R and } c \in \text{C are linked} \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

The weighted monopartite projection on one of the layers is constructed counting so-called V-motifis: we draw a link in the projected network if two nodes share a neighbour in the bipartite network. For instance, to derive the weighted monopartite projection on layer R, we count co-occurences in the bipartite network and construct the square $N_R \times N_R$ matrix $\mathbf{M}$ with elements

$$m_{rr'} = \sum_{c=1}^{N_C} b_{rc} b_{r'c} \tag{3.3}$$

For our analyses, we derive weighted monopartite projections from three binary bipartite networks; namely, Subfields-Articles, Authors-Articles and Subfields-Authors.

3.9.3 Scientists' specializations

Our analyses require us to assign specializations (single or multiple) to individuals. Unfortunately, there is no standard way to approach this problem - in part because, unlike articles or patents which can often be unambiguously linked to a limited number of classes, scientists can explore the knowledge space quite extensively. For our purposes, a suitable assignment should take into account both the relative specialization of a scientist and the distribution of publications across areas. Share-based metrics can be used to construct effective assignments. An instance is the Revealed Scientific Advantage (RSA) recently used in Battiston et al. (2019), which is akin to a metric originally used in Balassa (1965) to analyse comparative international trade advantages among countries. We consider the normalized metric; for each author i and sub-field (two-digit PACS) s this is defined as

$$RSA_{is} = \frac{\frac{w_{i,s}}{\sum_s w_{i,s}}}{\frac{\sum_i w_{i,s}}{\sum_{i,s} w_{i,s}}}, \tag{3.4}$$

where $w_{i,s}$ is the number of articles author i has published in sub-field s. By construction, $RSA_{is} \in [-1, 1]$, and a positive value indicates an advantage for author i in sub-field s. To assign a single specialization to i, we simply take $s(i) = argmax_s\{RSA_{is}\}$.

To assign multiple specializations to i, we take $S(i) = \{s \ s.t. \ RSA_{is} > 0\}$. In this case we actually create a fictitious "copy" of i for each of the sub-fields in $S(i)$ - keeping all individual characteristics but the specialization for each copy. This overcomes possible biases stemming from classification errors or marked heterogeneity in the distribution of articles across sub-fields.

3.9.4 Measures of knowledge and social relatedness

We define knowledge relatedness among sub-fields (two-digit PACS) from the bipartite network PACS-Articles. Specifically, we derive the monopartite projection on the PACS layer (a 68×68 co-occurrence matrix) and then apply the cosine similarity to construct a knowledge relatedness matrix. The procedure is illustrated in Figure 3.3: panel (**a**) shows a stylized example of the bipartite network PACS-Articles, panel (**b**) shows the network of co-occurrences of all pairs of PACS (the monopartite projection on the PACS layer), and panel (**c**) shows the cosine similarity matrix describing proximity among physics sub-fields.

We define social relatedness from the initial co-authorship network $G(V, E)$. Specifically, we build an augmented graph $G'(V', E')$ to integrate scientists' specializations: for each node (author) $V \in G$, we create an *individual* node in G' and for each edge $E \in G$ we draw the corresponding edge in G'. Then for each PACS s, we create an *attribute* node in G'. Next, we add further edges to G' considering the specialization(s) of each scientist and creating an edge between her individual node and the her specialization(s)'s attribute node(s) (panel (**d**) of Figure 3.3 provides a simple example). Finally, we capture social relatedness with a binary variable based on whether an author has at least one coauthor specialized in a sub-field different from her own; that is

$$SR_{is} = \begin{cases} 1 & \text{if } d(i,s) = 2 \\ 0 & \text{otherwise} \end{cases} \tag{3.5}$$

where $d(i, s)$ is the geodesic distance between scientist i and sub-field s in the augmented graph.

3.9.5 Modeling and assessment of predictors' contributions

Consider an author i specialized in the sub-field a. The probability that she is also active in sub-field $b \neq a$ is modeled as

$$p := f(KR_{ab}, SR_{ib}, \mathbf{IF_i}, \mathbf{SC_b}, \mathbf{Cit_b}) \tag{3.6}$$

where KR_{ab} is the knowledge relatedness between the two sub-fields, SR_{ib} is the social relatedness between the author and the sub-field b, $\mathbf{IF_i}$ is a vector of author's characteristics, $\mathbf{SC_b}$ is a vector of variables capturing the sub-field popularity and competition (i.e., for each sub-field, number of papers and number of specialized scientists), and $\mathbf{Cit_b}$ is a vector of variables capturing the relative attractiveness of the sub-field. A full list of the variables comprised in these vectors is provided in Table B.3. We reformulate the model as a logistic regression and consider two baseline specifications, with and without the interaction term between knowledge and social relatedness:

$$Y = ln(\frac{p}{1-p}) = \alpha + \beta KR_{ab} + \gamma SR_{ib} + \boldsymbol{\theta} \cdot \mathbf{IF_i} + \boldsymbol{\eta} \cdot \mathbf{SC_b} + \boldsymbol{\phi} \cdot \mathbf{Cit_b} \tag{3.7}$$

$$Y = ln(\frac{p}{1-p}) = \alpha + \beta KR_{ab} + \gamma SR_{ib} + \zeta(KR_{ab} \times SR_{ib}) + \boldsymbol{\theta} \cdot \mathbf{IF_i} + \boldsymbol{\eta} \cdot \mathbf{SC_b} + \boldsymbol{\phi} \cdot \mathbf{Cit_b} \tag{3.8}$$

For both the single-and multiple-specialization settings, we fit these logistic regressions in three scenarios; namely, *full* (no constraint on sub-fields a and b), *within field* (a and b in the same field; i.e. one-digit PACS code) and *between field* (a and b in different fields) diversification.

In order to quantify the roles of knowledge and social relatedness, we compute the *Relative Contribution to Deviance Explained* (RCDE) for each of these variables (Campos-Sánchez et al., 2016). For a generic predictor X this is defined as

$$RCDE_X = \frac{(D_{null} - D_{full}) - (D_{null} - D_{full \setminus X})}{(D_{null} - D_{full})} \tag{3.9}$$

where D_{null} is the null deviance, D_{full} is the residual deviance of the full model (including all predictors) and $D_{full \setminus X}$ is the residual deviance of the model obtained by removing X (in our case KR or SR). The RCDE thus quantifies the percentage of the total logistic deviance attributable X.

Chapter 4

Science, technology and climate change

This chapter is largely based on the following working paper: (Tripodi et al., 2022)

4.1 Introduction

There is increasingly robust evidence that meeting ambitious climate targets, perhaps with limited temperature overshooting (Riahi et al., 2021), will require removing large stocks of carbon dioxide from the atmosphere (Allen et al., 2019; Shukla et al., 2022). Tackling climate change by removing CO2 from the atmosphere has been a tantalizing idea for quite some time (Baes et al., 1980). Planting trees, or more precisely, designing forest management programs, has probably been the first solution to arise (Dyson, 1977). Over time though, a broader set of technical solutions have been developed, generally going under the label of *Negative Emissions Technologies* (NETs).

Recently, as stressed in the last IPCC report, a large majority of Integrated Assessment Models (IAMs) mention NETs as a pivotal element to meet the Paris Agreement requirements and thus tackle global warming (Rogelj et al., 2015; Clarke et al., 2014; Shukla et al., 2022). According to these models, the transition toward zero emissions will require the extensive deployment of NETs to balance the inevitable difficulties of cutting short-term emissions even more drastically (Van Vuuren et al., 2017). Furthermore, NETs might contribute to smooth out the so-called green transition, which will prove challenging from an economic, social, technological, and, of course, political perspective (de Coninck et al., 2018).

As of today, there are doubts on the possibility of immediate large-scale deployment of NETs, and their use as technical or policy panacea could not only be implausible, but even hazardous (Anderson and Peters, 2016; Van Vuuren et al., 2018; Grubler et al., 2018; Lane et al., 2021). The inclusion of these technologies in the design of climate policy pathways could risk delivering misleading guidelines if it underestimates the long and uncertain process that moves from basic research to the systemic diffusion of complex technical artifacts (Tavoni and Socolow, 2013; Fuss et al., 2014; Vaughan and Gough, 2016; Dosi, 1988; Dosi and Nelson, 2010; Probst et al., 2021). In addition, little is known about how NETs at full regime could interact with other Sustainable Development Goals (SDGs) (Fuss et al., 2014; Fuhrman et al., 2020). NETs are indeed a peculiar set of technologies, whose economic value and market size largely depends on the strength of current and future climate policy, as well as from the global trajectory of emissions (Meckling and Biber, 2021). Against this backdrop, the available evidence about how different NETs could develop and diffuse is inconclusive.

Our analysis provides new evidence about the relationships between scientific research in NETs, its diffusion and policy coverage, as well as their technological developments. In particular, we quantify the likelihood that scientific advances in NETs research (i) stimulate the production of further knowledge, (ii) foster technological innovation, and (iii) enter the policy debate. Moreover, we investigate the geographical distribution of NETs-related knowledge production, using relative comparative advantage and network analytic measures to identify the scientific specializations of countries and single out the main research hubs of the global innovation system.

Our work contributes to a recent stream of studies acknowledging a relatively marginal role of NETs-related research within the broader climate discourse (Minx et al., 2017), and emphasise the need to better understand the scientific trends, the diffusion and up-scaling issues of NETs (Minx et al., 2018; Fuss et al., 2018;

Nemet et al., 2018), as well as their broader economic challenge. Different NETs have been mostly evaluated along five dimensions (Figure 4.1**B**): negative emissions potential (i.e., Gt Ceq per year), energy and natural resource requirements (i.e., land and water use) and economic costs (US$ per t Ceq) (Smith et al., 2016; Smith, 2016). Overall, no universally superior option has been identified (Rueda et al., 2021). This chapter provides novel dimensions to the multi-faceted comparison of various carbon removal technologies and provides the first estimates of knowledge spillovers generated by research in NETs.

We focus on the following list of options (see Figure 4.1**A** and Table C.1 for a summary description): Afforestation and Reforestation (AR), Bio-energy with Cabon Capture and Storage (BECCS), Biochar, Blue carbon (BC), Direct Air Capture (DAC), Enhanced weathering (EW), Ocean fertilization (OF), and Soil carbon sequestration (SCS). DAC does not explicitly include storage options (Fuss et al., 2016); see Section 4.7.1 and C.1 for more details.

We measure knowledge spillovers by using citations networks, as is standard in the innovation and applied economics literature (Jaffe et al., 1993; Dechezleprêtre et al., 2013; Jaffe and De Rassenfosse, 2019). Given the critical role played by climate-related technologies, we move beyond the standard citation counts to incorporate knowledge flows to practical innovations (i.e., patents) and the public discourse (i.e., policy documents) (Ahmadpoor and Jones, 2017; Yin et al., 2021a,b). We also include the broader public impact of NETs research through different media channels to take into account a more complete and multidimensional set of knowledge spillovers. More in detail, by analyzing 20 years of academic literature via network and regression techniques (Verdolini and Galeotti, 2011; Popp, 2016), we first provide a quantitative comparison of the impact of different NETs. Next, we focus on knowledge spillovers of NETs research in science, technology, and policy. Finally, we provide additional geographical and network analyses to study the spatial heterogeneity of cities and countries that can serve as research hubs for supporting future collaborations.

In extreme synthesis, by unpacking the multidimensional impact of knowledge spillovers, this work suggests the existence of coordination gaps between science, technology and policy in the domain of carbon removal solutions. Our results show that (i) knowledge spillovers in science play a non-negligible role in the development of negative emissions solutions, (ii) in terms of impact, NETs are characterized by great heterogeneity, and only very few options are substantially linked to marketplace inventions, and (iii) negative emissions research activities are geographically concentrated around hubs with different specialisations from the viewpoint of the global division of labour. Interestingly, DAC appears as the most promising solution concerning technological developments (as indicated by patent citations); however, it is still relatively overlooked by policymakers (as indicated by policy reports citations).

4.2 Knowledge base and spillovers: the landscape of negative emissions research

Technological and scientific breakthroughs are often the result of knowledge recombination processes, wherein past scientific advances become themselves knowledge components of future, often unexpected, innovative research paths (Dosi, 1982, 1988; Fleming, 2001; Xiao et al., 2021). Our exploration of the NETs' research landscape starts by mapping the knowledge base (i.e., scientific fields on which NETs rely upon) and the potential spillover directions (i.e., scientific sub-fields influenced by NETs research developments). To identify them, we collected a large amount of bibliometric information related to NETs articles published in scientific journals (see Section 4.7.1). We retrieve NETs papers by querying Web of Science (WoS) on the basis of keywords and their combinations in titles and abstracts (Fuss et al., 2016; Minx et al., 2017). From 1998 to 2017, we collect 3301 published articles, distinguishing eight different NETs and considering a general residual category. Figure 4.1**c** shows the growing number of publications per year, with details for the different NETs. Next, we collect citations data from scientific papers, patents, and policy documents, along with non-technical media mentions (e.g., in social media, newspapers, blogs). To do so, we integrate several data sources, namely: Web of Science (WoS), Reliance on Science (RoS) and Altmetric (see Sections 4.7.1 and C.1

A

Negative Emissions Technology	NET	References*
Afforestantion & Reforestation	**AR**	[Smith et al., 2016, Fuss et al., 2018]
Bio-energy with Carbon Capture and Storage	**BECCS**	[Smith et al., 2016, Fuss et al., 2018]
Biochar	**Biochar**	[Smith, 2016, Fuss et al., 2018]
Blue Carbon	**BC**	[Fuss et al., 2018, Bertram et al., 2021]
Direct Air Capture	**DAC**	[Smith et al., 2016, Fuss et al., 2018]
Enhanced Weathering	**EW**	[Smith et al., 2016, Fuss et al., 2018]
Ocean Fertilization	**OF**	[Strong et al., 2009, Fuss et al., 2018]
Soil Carbon Sequestration	**SCS**	[Smith, 2016, Fuss et al., 2018]

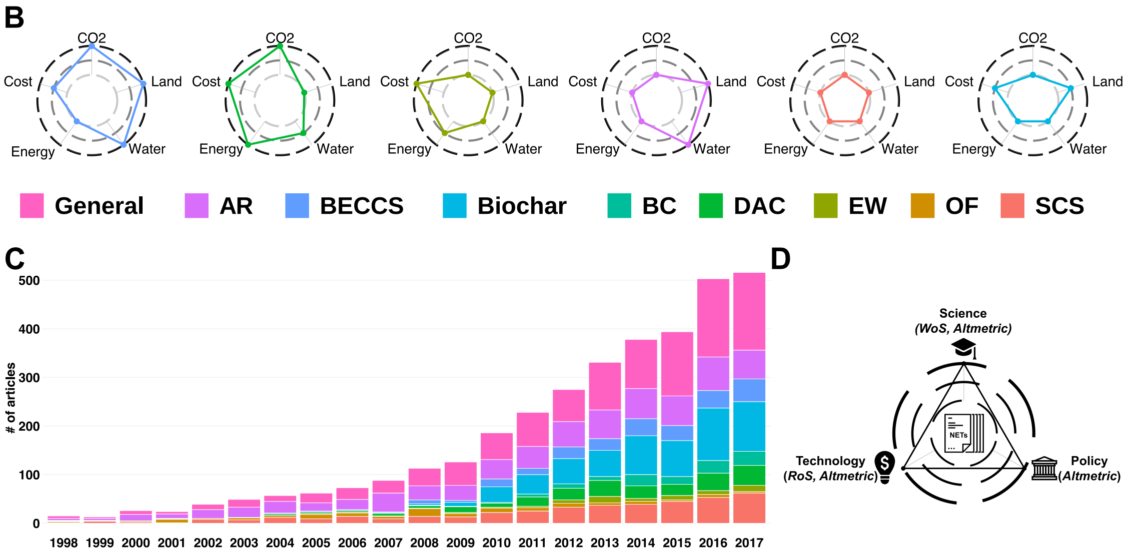


Figure 4.1: Negative emissions research. **(A)** The list of eight NETs included in our analysis. **(B)** A multidimensional comparison among different NETs (authors assessment adapted from (Smith et al., 2016; Smith, 2016; Fuss et al., 2018)). **(C)** NETs articles from 1998 to 2017 collected though WoS text search. The category *General* is defined as a residual class including articles that match NETs keywords but do not specifically include words patterns in their titles or abstracts. **(D)** A stylized representation of the diverse sources of data necessary to keep track of knowledge flows to science, technology and policy. **(*)** The aforementioned references provide a detailed review of each NET. Summary radars concerning OF and BC not included (fewer conclusive information currently available (Strong et al., 2009; Bertram et al., 2021)).

for more details). Figure 4.1**d** provides a schematic representation of the different sources of data used in our analysis to keep track of the multidimensionality of knowledge spillovers.

Negative emissions technologies are not all alike: crucial differences have been reported in relation to measurement, verification, accounting, and durability of carbon stored (Joppa et al., 2021), as well as to costs and requirements (Smith et al., 2016; Smith, 2016). Against this background, we investigate the heterogeneity that characterizes NETs' knowledge base and spillover directions (Figure 4.2). Nature-based and technology-based approaches differ in both aspects. Figures 4.2**A,B,C,D** show a qualitative comparison between two nature-based methods (i.e., forest management and soil carbon sequestration) with the most popular technological solutions (BECCS and DAC). As expected, nature-based NETs are scientifically grounded in soil science and ecology, while solutions such as BECCS and DAC are engineering-driven methods. More interestingly, NETs build on different scientific fields, and the directions of potential spillovers follow accordingly. To better illuminate this, we show the overlap rates among subjects most frequently reported in the knowledge base (Figure 4.2**E**) and in set of spillover directions of each NETs pair (Figure 4.2**F**). Overall, our descriptive observation signals a prominent feature of NETs: the scientific heterogeneity of their knowledge base closely reflects the direction of spillovers effects. Some NETs can certainly be compatible in applications, but they are not synergic in the knowledge they develop and build upon.

In the following Sections, we investigate the impact of negative emissions research on several dimensions, revealing that NETs generate substantial but heterogeneous spillovers, and that research activities are not evenly distributed from a geographical perspective.

4.3 Multidimensional impact of NETs research

As mentioned above, NETs comprise a heterogeneous group of carbon capture solutions stemming from a diversified range of scientific disciplines. In this Section, by exploiting the richness of different sources of data, we characterize, for the first time,

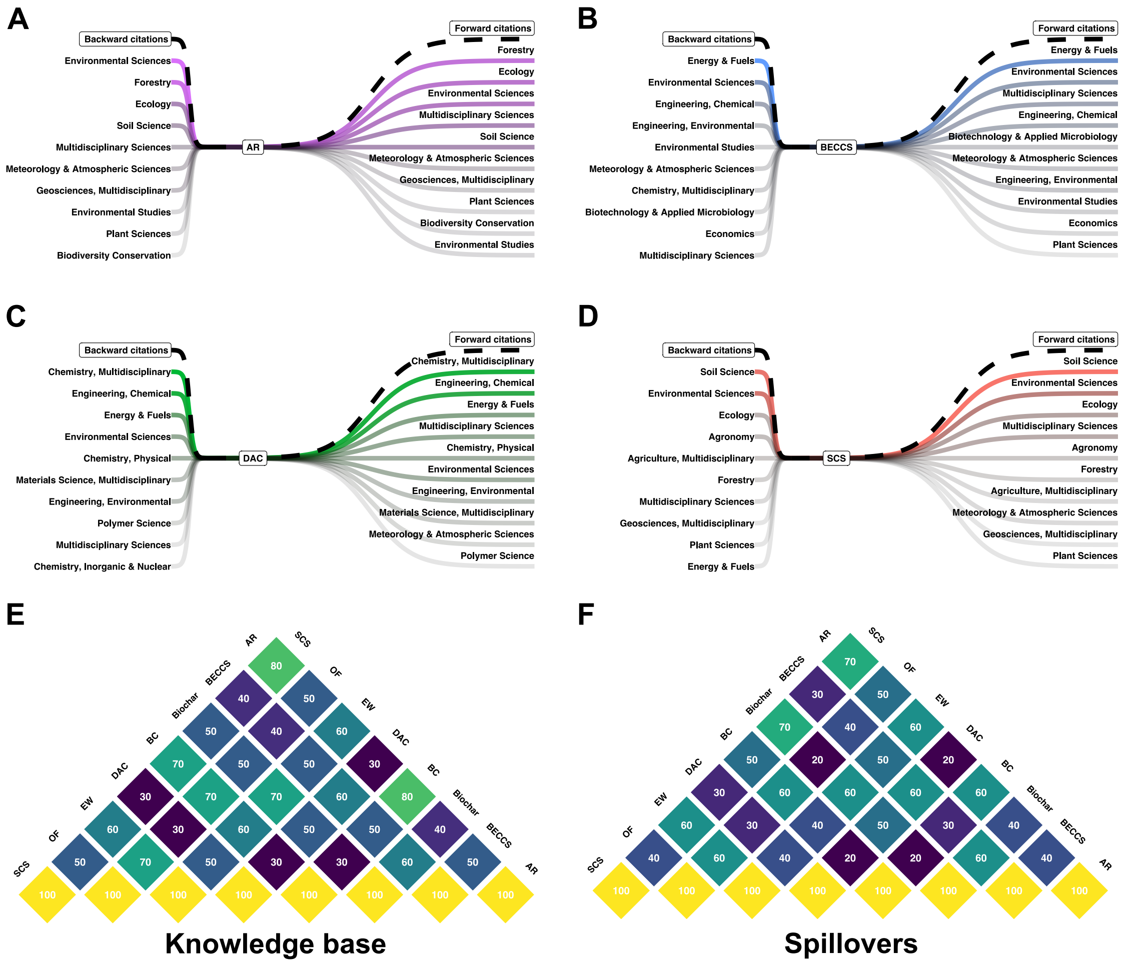


Figure 4.2: NETs knowledge base and spillovers. **(A–D)** Flows diagrams for AR, BECCS, DAC, and SCS. Top 10 WoS subjects that affect (backward citations) and are affected (forward citations) by NETs research. The first ten subjects comprise the large majority of citations (see figure C.6 for details). **(A)** Afforestation and Reforestation – AR. **(B)** Bio-energy with Carbon Capture ans Storage – BECCS. **(C)** Direct Air Capture – DAC. **(D)** Soil Carbon Sequestration – SCS. **(E)** Matrix of overlapping subjects in NETs knowledge base (% values). **(F)** Matrix of overlapping subjects in NETs knowledge spillovers (% values). Full list of flows charts included in C.5.

the multidimensional nature of NETs impact, measuring their spillovers within and beyond their scientific reach.

Our quantitative comparison among scientific articles relies on identifying suitable control groups. Therefore, we employ a matching procedure to construct a "baseline" control, including articles published in the same year and the same journal, not directly related to NETs. In addition, to better characterize the role of NETs within the broader climate change academic debate, we construct a second control group (i.e., "climate control"), following the same strategy but focusing on the climate change literature. (see Section 4.7.1 and C.2 for more detailed information related to our matching strategy). It is worth noticing that our matching procedure has the purpose of balancing the comparison taking into account articles of the same age and ideally the same quality. However, such a matching scheme does not guarantee an exact counterfactual; it ensures that we compare articles with some key common characteristics.

Using Altmeric data, we compute the normalized number of mentions for each NET to gauge how the different streams of research are covered in academic, policy, technical, and media outlets. Figure 4.3 summarizes a first quantitative comparison in terms of impact (with the control group fixed at 1): each radar chart (Figure 4.3**A**–**I**) shows the multidimensional impact profile that characterizes research articles belonging to different NETs. We perform the same empirical exercise using as benchmark the climate control group (see Figure C.7). Two main observations must be made: first, as NETs are intrinsically different, their impact mirrors such differences both from a qualitative and a quantitative perspective. Some negative emissions solutions have momentum beyond the academic realm, with some of them, such as BECCS or Blue Carbon, being relatively popular in policy documents and media outlets. In addition, EW research has been discussed on social media such as Facebook. Second, very few options are linked to practical technological developments (i.e., mentions in patents), the only exception being DAC. Given the crucial role of the nexus between science, technology, and policy for developing specialized climate solutions, we investigate these three dimensions in greater detail in the next Section.

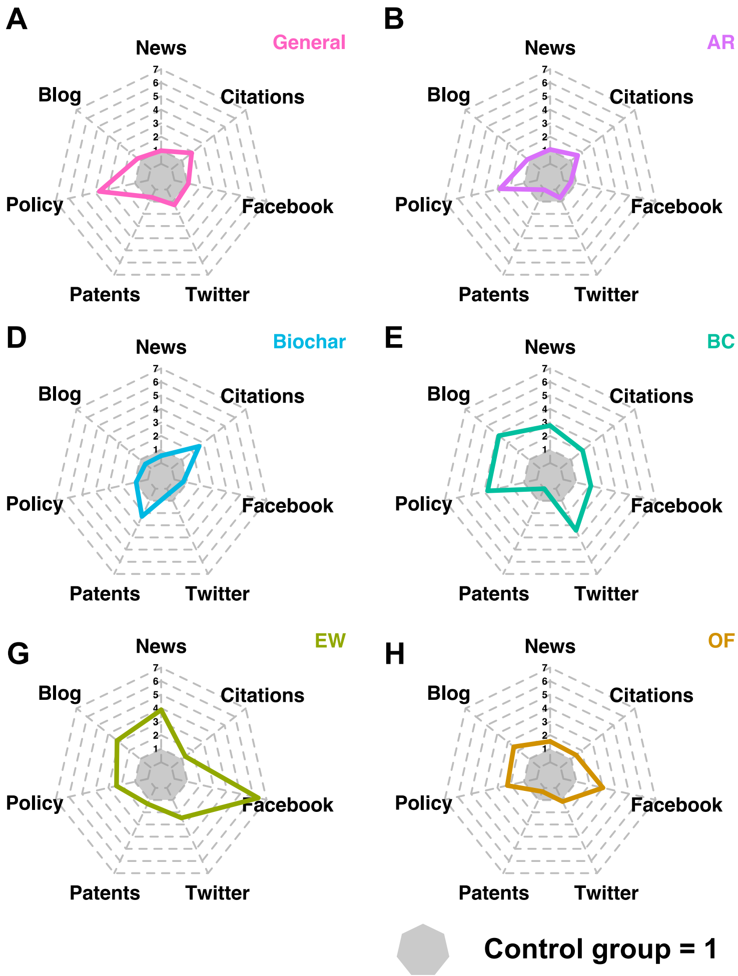


Figure 4.3: Multidimensional coverage of NETs research. **(A–I)** Radar charts for each NET, showing multidimensional spillovers (control group fixed at 1). **(A)** General. **(B)** Afforestation and reforestation – AR. **(C)** Bio-energy with Carbon Capture and Storage – BECCS. **(D)** Biochar. **(E)** Blue Carbon – BC. **(F)** Direct Air Capture – DAC. **(G)** Enhanced weathering – EW. **(H)** Ocean fertilization – OF. **(I)** Ocean fertilization – OF.

4.4 Quantifying knowledge flows to science, technology and policy

It has widely been argued that tackling climate change will require novel scientific research, practical technological innovations as well as policy support.[1] This Section quantifies the impact that knowledge accumulation in NETs produces on technology, policy, and science itself.

We rely on econometric methods based on generalized linear models to estimate the size of knowledge spillovers. Our preferred specifications employ negative binomial regressions for citations counts and logistic regressions for citation likelihoods. We run a set of regressions on one-to-one matched samples to check the stability of our results and to quantify the uncertainty around our estimates (see Section 4.7.2 and C.2 for econometric and matching details). Results are summarized in Figure 4.4. In particular, Figure 4.4**A** highlights that several negative emissions options generate relative more spillovers than control groups. For instance, Biochar, BECCS and DAC articles collect, on average, 2.59, 1.84 and 1.83 times more citations than the non-NET control group, respectively. However, it is just for few NETs that scientific advances significantly impact on technological development (Figure 4.4**B**). Namely, DAC and Biochar research is somehow related to patenting activities, with a significant gap in favor of DAC. Indeed, DAC scientific advances are 7.89 times more likely to be cited by a patent. Contrarily, when looking at the probability of being cited by a policy document, BECCS and BC stand out among all the options (see Figure 4.4**C**).

To better quantify the variability of our point estimates as possible control groups vary, we compute the confidence interval around our mean effects size (i.e., β_k^*). Table 4.1 summarizes the mean effects for each coefficient across different runs of our statistical model (i.e., point estimates for all our NETs) and its variability. Our estimates prove relatively stable to possible differences in the matched control groups. Nevertheless, as far as scientific spillovers are concerned, we notice some differences

[1] See, for example, calls for attention by the EU Commission and the UK government.

between the baseline and climate control. As expected, climate change is a very active area of research, leading to smaller coefficients in our setting. In addition, we re-estimate our model including controls related to fields (or combination of fields) and whether articles are open access (see Figure C.14 and C.15). To further check the robustness of our results, we run our analysis using alternative control groups, different data sources, and alternative models (see Section C.6 for all the details). The insights of our empirical investigation are confirmed irrespective of specifications, data sources, and alternative measures. While there is plenty of evidence that citations, at least partially, capture positive knowledge spillovers for science and technology advances (Fortunato et al., 2018a; Jaffe et al., 2000), little is known about the references in policy documents. Hence, to better understand the role of citations coming from policy documents, we select a subset of policy reports to measure their overall sentiment. Our analysis shows that the overall sentiment of the documents citing NETs articles is positive (see Section C.3).

Our results already bring important implications for climate and innovation policy: NETs constitute an active research area with great potential and attract substantial attention within the scientific community. Nevertheless, our multidimensional spillovers estimates signal that most NETs hardly move beyond the scientific realm: only DAC research turns into marketplace innovations. In addition, the policy dimension seems to be relatively disconnected from the general scientific and technological trends. Finally, to better understand the trends that characterize NETs research efforts, we focus on the geography of NETs research and collaborations in the next Section.

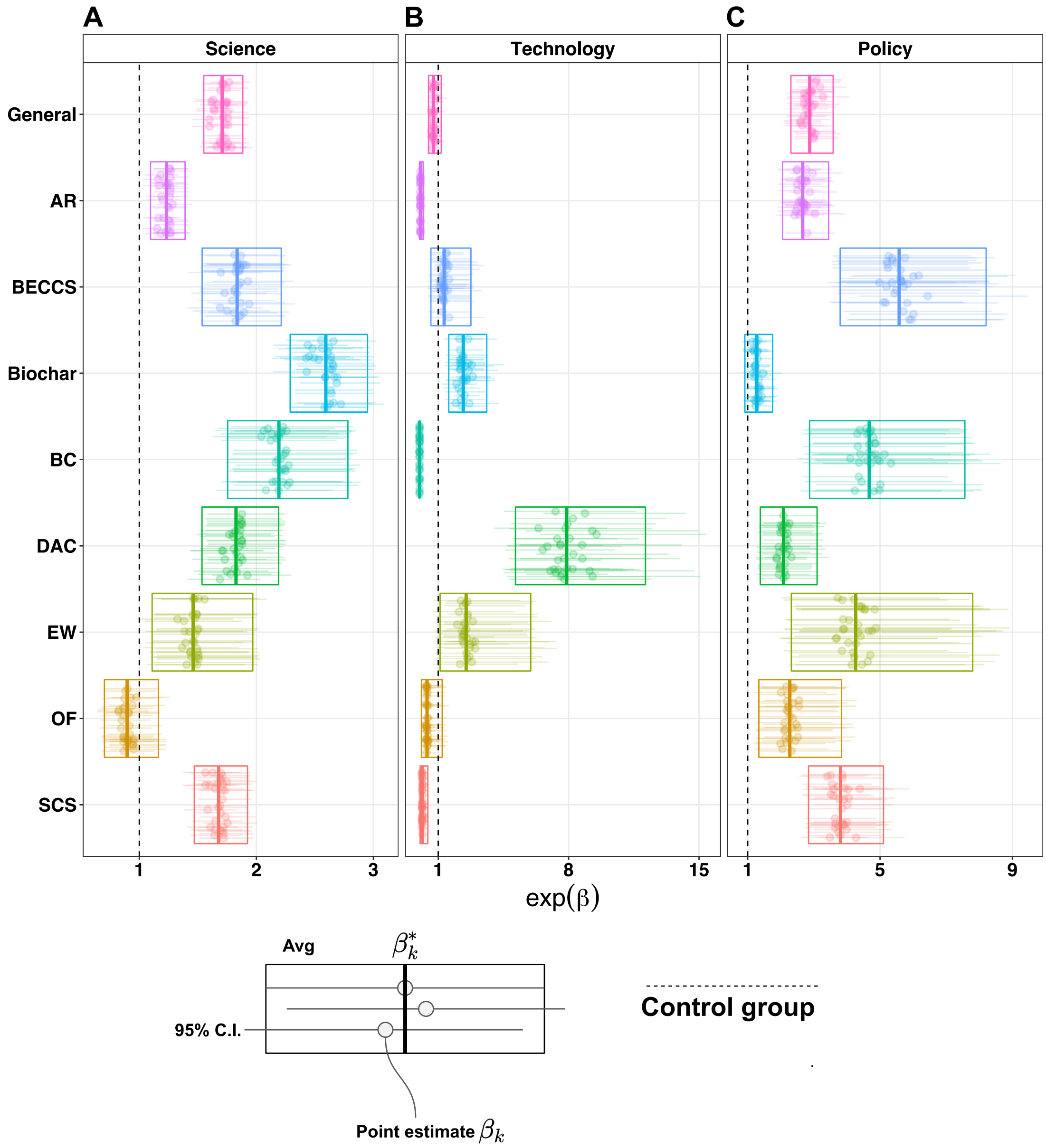


Figure 4.4: NETs spillovers to science, technology and policy. Coefficients (exponentiated) of the regression models of Eq. (1). Results are obtained by fitting 30 negative binomial regressions (**A**) and 30 logistic regressions (**B–C**) on one-to-one matched samples with year dummies. **(A)** Incident Rate Ratio (IRR) for each NET on the number of scientific citations. **(B)** Odds Ratio (OR) for each NET on the probability of being cited by a patent (BC estimates set to zero since there is no patent documents citing BC papers). **(C)** Odds Ratio (OR) for each NET on the probability of being cited by a policy document.

Table 4.1: Point estimates variability for baseline and climate control. IRRs and ORs (i.e. average exponentiated coefficients β_k^*) estimated through regression models – Eq. (1) – and relative variability of point estimates β_k [C.I. 95%].

	Baseline control			Climate control		
NET	Science	Technology	Policy	Science	Technology	Policy
General	1.71	0.75	2.88	1.40	0.99	1.93
	[1.55,1.88]	[0.720,0.77]	[2.81,2.94]	[1.39,1.42]	0.961,1.03]	[1.90,1.96]
AR	1.23	0.042	2.66	0.96	0.06	1.75
	[1.10,1.39]	[0.0412,0.04]	[2.61,2.72]	[0.95,0.97]	[0.0629,0.07]	[1.73,1.78]
BECCS	1.84	1.32	5.58	1.55	1.43	3.74
	[1.54,2.21]	[1.27,1.37]	[5.45,5.71]	[1.53,1.56]	[1.38,1.49]	[3.68,3.80]
Biochar	2.59	2.34	1.28	2.18	3.26	0.88
	[2.29,2.95]	[2.26,2.43]	[1.25,1.31]	[2.16,2.20]	[3.14,3.38]	[0.864,0.89]
BC	2.19	✗	4.67	1.78	✗	3.19
	[1.75,2.78]		[4.57,4.78]	[1.76,1.80]		[3.13,3.25]
DAC	1.83	7.89	2.08	1.82	12.3	1.47
	[1.54,2.19]	[7.59,8.19]	[2.04,2.13]	[1.17,1.20]	[11.9,12.8]	[1.44,1.49]
EW	1.46	2.50	4.27	1.19	4.02	2.50
	[1.11,1.97]	[2.41,2.59]	[4.13,4.40]	[1.17,1.20]	[3.85,4.18]	[2.46,2.53]
OF	0.896	0.401	2.27	0.70	1.03	1.34
	[0.702,1.16]	[0.39,0.42]	[2.21,2.33]	[0.69,0.71]	[1.00,1.06]	[1.31,1.37]
SCS	1.68	0.138	3.81	1.37	0.20	2.49
	[1.47,1.92]	[0.133,0.14]	[3.72,3.89]	[1.36,1.39]	[0.198,0.210]	[2.45,2.53]

Note: No valid estimates for BC in Technology due to absence of citations from patent documents.

4.5 The geography of NETs research collaborations

So far, we have provided empirical evidence on the heterogeneity of NETs research in terms of knowledge base, scientific impact, and spillovers to practical applications.

Empirical evidence shows that proximity matters for complex activities and, more precisely, that innovation is disproportionately concentrated in cities (Carlino et al., 2007; Catalini, 2018; Balland et al., 2020). So, in this Section, we turn our attention to the geography of negative emissions research. First, we geo-localize NETs scientific articles using author affiliation data from WoS. Then, we derive countries' relative specializations by looking at the geographical distribution of research activities. Finally, we map scientific collaborations (at both the country and city level) to eventually shed some light on the identification of potential research hubs (see Section 4.7.1 for more details on the geo-localization of NETs articles).

Figure 4.5**A** depicts the aggregate geographical distribution of research activities. The map shows the total number of articles related to NETs, the centrality (i.e., nodes' strengths) both at the city and country level and the overall collaboration network. For the sake of clarity, we filter out cities that appear less the ten times in our sample (see Section 4.7.3 for network construction details). Beijing stands out as the city associated with the most significant number of articles and appears to be the most central city in the collaboration network. At the country level, though, the USA maintain their role as the primary research hub worldwide. However, the aggregate collaboration network can hardly allow us to dig deeper into a single technology, as it might be influenced by the distribution of articles across specific NETs. Therefore, we focus on different NETs separately. First, to better capture the relative specialization of countries in different NETs, we compute the Relative Scientific Advantage (RSA, see 4.7.3 for more details). Figure 4.5**B** summarizes the values of the RSA for a subset of countries, signaling, for instance, the greater specialization of European countries and the USA concerning engineering-based options such as BECCS and DAC. Intuitively, relative specializations still

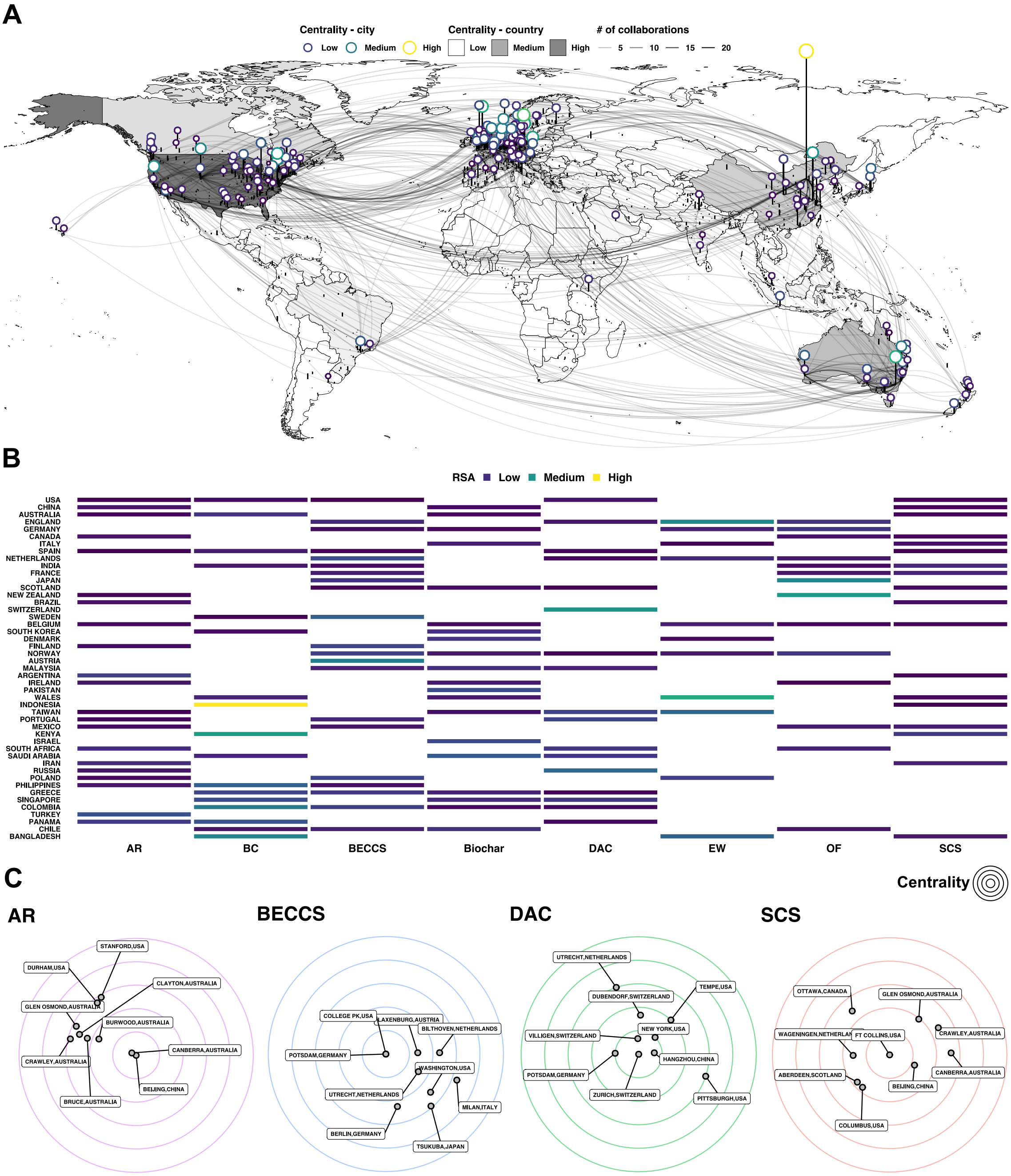


Figure 4.5: NETs geography and collaborations. **(A)** Geographical distribution of NETs research activities across cities (i.e., total number of NETs publications). Cities and countries centrality (i.e., node strength) scores are computed by analyzing the aggregate collaboration networks. **(B)** Revealed Scientific Advantage of selected countries (white spaces indicate values lower than 1). **(C)** Centrality ranking in the research collaboration networks related to AR, BECCS, DAC, and SCS.

underline the links between research potentials and local opportunities. According to the RSA, Switzerland appears primarily specialized in DAC research, while Indonesia – one of the largest reserves of coastal forests – is almost fully specialized in

BC. Next, we construct collaboration networks for all the NETs in our sample. Formally, we identify the largest connected component (i.e., the largest subset of nodes that can be reached from one another) and pin down the most central cities for each specific negative emissions option. As in Figure 4.2, we focus on AR, BECCS, DAC, and SCS (see Section 4.7.3 for all NETs). Figure 4.5**C** points out that basic network measures can already allow us to spot different geographical specializations: Beijing and Canberra result as the most central locations as far as AR is concerned. Postdam and College Park are the most important hubs for BECCS research, while Fort Collins stands above in the SCS research. Finally, Zurich appears as the most central city for DAC research. Interestingly, the company that first made it to the market with a commercial DAC solution was founded as spin-off of the ETH in Zurich (as of today, several companies are active in the DAC sector). The Zurich example highlights the importance of basic scientific research in developing technologically viable climate solutions and the role of geographical proximity between science and technology hubs. Innovative activities benefit from co-location, allowing scientists and inventors to form collaborations and share valuable knowledge. The potential research hubs identified above might well pave the way to accelerate advances in NETs.

4.6 Discussion

The urgent need for a rapid scale-up of NETs development and deployment should go hand in hand with extensive R&D efforts worldwide. Indeed, keeping track of the knowledge flows generated by negative emissions research would be crucial to inform scientists, market players, and policymakers on the potential opportunities for such technologies in the next few years. Our analysis provides a first quantitative comparison among different negative emissions solutions from the science-innovation nexus standpoint. Looking at knowledge spillovers within and outside the academic world, we find that negative emissions research is highly heterogeneous and spread across different hubs. Only a few options will eventually turn into marketplace inventions. As of today, DAC appears to be the most promising as far as practical technological innovations are concerned.

A quantitative benchmark of multidimensional spillovers for NETs can be considered as a starting point to evaluate the potential impact of NETs technological trajectories (or different climate-related technologies) from science to practical applications. In other words, it is an instrument that can be used by climate scientists and policymakers to keep track of scientific and technological trends from a systematic quantitative perspective.

Our empirical analysis is not without limitations, and some of these limitations point towards future research directions. First, the scientific literature on negative emissions is growing fast and in an interdisciplinary way. We follow a well-defined query strategy for the retrieval of data, relying on specific patterns and keywords. However, identifying the relevant articles and their disciplinary span might need more advanced criteria in the future. Machine learning/NLP models might prove helpful in finding better clusters of articles and consequently the direction of their spillovers. Second, we employ a matching procedure and propose an intrinsically stable re-sampling strategy to compare similar articles robustly. Nevertheless, we do not identify causal mechanisms or the impact of funding on the trajectories of such negative emission options. Relatively small advances might lead to sudden and sizeable changes for early-stage technologies. We can not rule out the possibility that some universally superior NET will appear in the following years or that some technological breakthroughs would make some existing ideas more likely to be patented. Finally, citations are only an imperfect measure of knowledge spillovers. Although our methodology relies on different data sources, our quantification might still be subject to possible measurement errors.

From a policy perspective, our findings provide at least two clear insights. First, when considering the applicability of a diversified portfolio of NETs, their knowledge base, spillovers and trajectory of development should be considered carefully. Indeed, our analysis support evidence of little synergies between various NETs. Second, given the current distance of negative emissions research from the technological frontier, the prospective diffusion of NETs at scale would benefit from both conventional and unconventional innovation policies (Meckling and Allan, 2020; Hanna et al., 2021; Gross and Sampat, 2021b,a). In practice, R&D subsidies, public procurement, grants as well as the reinforcement of university-industry linkages could be coupled with the proposal of innovation prizes (e.g., XPRIZE) and Advance Market Commitments (AMC, e.g., Frontier), previously used to serve different scientific and policy purposes (Kremer et al., 2020). In addition, the evidence of strong positive knowledge spillovers could support a mission-oriented approach towards NETs (Mazzucato, 2013). Innovation can play an essential role in dealing with the climate change crisis (Nature-editorial, 2022); however, science, technology, and policy need to be better coordinated to boost the efficacy of research endeavors.

4.7 Data and Methodology

4.7.1 Data & matching

To track down the evolution of NETs research, we use three main sources of data: Web of Science (WoS), Reliance on Science (RoS), and Altmetric. WoS is a large global citations database collecting millions of research articles information and maintained by the private company Clarivate. RoS is a publicly available database that includes citations from patents to scientific articles Marx and Fuegi (2020). Altmetric is a curated database that collects metrics complementary to standard citation-based data, such as mentions on a diverse set of outlets. Altmetric data can be freely available upon request for scientific purposes.

To identify the first sample of NETs relevant articles, we look at keywords, titles, and abstracts in WoS, as previously done in the literature Minx et al. (2017, 2018). We retrieve a total of 3301 scientific articles from 1998 to 2017 for 8 different NETs. Note that the queries we used to filter DAC articles do not explicitly include storage (see C.1), contrary to BECCS and in line with previous studies Minx et al. (2017, 2018). All the articles that match the keywords search with no explicit reference in their titles or abstracts are included in the NET category *General* (see section C.1 to find further descriptions of the sample and the full queries). Most of the articles retrieved from WoS are also covered in Altmetric ($\sim$ 62%). From WoS and

Altmetric we can collect all cited and citing articles of our focal NETs sample. We use both RoS and Altmetric to recover patents-articles citation links, and Altmetric to keep track of all mentions from policy documents, mainstream media outlets as well as blogs and social media platforms such as Facebook or Twitter.

As far as the geo-localization of NETs scientific output is concerned, starting from authors' affiliation data, we use OpenStreetMap and the R package *tmap* to identify the coordinates of the cities linked to the publications in our sample. After a manual inspection, we can geo-localize a total of 3255 articles ($\sim 98\%$ of the initial set of articles).

To quantify multidimensional spillovers of NETs research, taking care of possible sources of bias, we employ an exact matching procedure and construct two controls groups: a "baseline" and a "climate" specific sample. First, for each focal negative emissions paper, we select up to 10 articles published in the same year and the same journal (the final pool of articles includes about $23k$ articles). Then, for our regression analysis, we further refine our procedure to match articles one-to-one. In detail, to check the stability of our results we create 30 sub-samples with replacement. We repeat the aforementioned procedure to construct a second set of control groups, specifically designed to match climate-change related articles. We retrieve climate-specific papers by querying WoS as in Grieneisen and Zhang (2011) (the final pool of climate-specific articles includes about $20k$ articles). Sections C.2 and C.1 describe in greater detail the matching scheme, the queries to collect the climate-specific control, and the overall compatibility among our different sources of data.

4.7.2 GLM regressions

To quantify knowledge spillovers in sections 4.4, we employ generalized linear regression models. After the construction of our one-to-one matched sub-samples, we estimate a negative binomial regression to model citation counts coming from scientific papers. Next, we use logistic regressions to model the probability of being cited by a patent or a policy document. The baseline specification can be written as follows:

$$g(E(S_{ikt}|NET_{ik}, T_{it}, \mathbf{X_i})) = \alpha + \sum_k \beta_k NET_{ik} + \sum_t \gamma_t T_{it} + \delta \mathbf{X_i} \qquad (4.1)$$

where S_{ikt} is the number of forward citations (alternatively, the occurrence of a citation from a patent or a policy document), NET_{ik} refers to the corresponding technology and T_{it} represents a year dummy, and $\mathbf{X_i}$ a vector of control variables such as free accessibility of the articles or sub-fields categories (see section C.6 for more details). Within this setting, the link function allows us to derive the relationship between the linear predictions and the expected value of the response variable (in our case a measure of knowledge spillovers). The link functions used for the binomial and negative binomial case are the following:

if $g(\cdot) = log\frac{\mu}{1-\mu}$ with $\mu = E(S_{ikt}|NET_{ik}, T_{it}) \rightarrow$ Logistic regression

if $g(\cdot) = log\mu$ with $\mu = E(S_{ikt}|NET_{ik}, T_{it}) \rightarrow$ Poisson/Negative Binomial regression

In practice, we estimate the models 30 times, to check the stability our results as the matched control groups vary. The boxes in figure 4.4 highlight the average effect: $\beta^*_k = \frac{1}{30}\sum_{c=1}^{30} \beta_{kc}$, where $c = \{1, \ldots, 30\}$ represent different matched control groups. The lower and upper bound of the boxes are instead the average confidence interval $\langle C.I. \rangle$, corresponding to the average value of the 95% confidence intervals across our estimates. In Table 4.1 we collect β_k and quantify the range of variation of these coefficients around their mean β^*_k (C.I. 95%).

4.7.3 Geographical specialization and collaboration networks

We employ the Revealed Scientific Advantage (RSA) to gauge countries' relative specialization. Such a metric was initially developed in Balassa (1965) to analyze comparative international trade advantages among countries (i.e., Revealed Comparative Advantage – RCA). Later it has been extensively used in several applications beyond trade Tripodi et al. (2020); Hidalgo (2021). Within our setting, for each country or location l and NET k, this is defined as

$$RSA_{lk} = \frac{\frac{w_{l,k}}{\sum_k w_{l,k}}}{\frac{\sum_l w_{l,k}}{\sum_{l,k} w_{l,k}}}, \tag{4.2}$$

where $w_{l,k}$ is the number of articles published in country l covering NET k. RSA values greater the 1 signal relative specialization.

By exploiting the geo-localization of NETs articles, we construct collaboration networks among cities to better understand where and how novel developments in negative emissions take place. The most straightforward way to analyze collaborations at different geographical levels is by using bipartite networks. A bipartite network is defined as a graph in which nodes are split into two separate sets (or layers). No link connects pairs of nodes that belong to the same layer. In our case, the two layers represent articles and cities, respectively. The binary case is simply described by a bi-adjacency matrix of dimensions $N_A \times N_C$. The number of rows N_A is the number of nodes in layer A (i.e., articles), and the number of columns N_C is the number of nodes in layer C (i.e., cities), as follows:

$$b_{ac} = \begin{cases} 1 & \text{if node } a \in \text{A and } c \in \text{C are linked} \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

In this setting, we draw a link in the bipartite network if any of the authors of a NET research article a is affiliated with an institution of a given city c. A weighted monopartite projection on the article layer is constructed by counting the co-occurrences in the bipartite network and takes the form of a square $N_C \times N_C$ matrix $\mathbf{M}$ with elements:

$$m_{cc'} = \sum_{a=1}^{N_A} b_{ac} b_{a'c} \tag{4.4}$$

Before computing our centrality measures (i.e., nodes' strength), we first derive the largest connected component to filter out unconnected nodes (or groups of irrelevant nodes). Having information about cities, we can also derive the aggregate network at the country level. See section C.7 for an additional description of our network analysis results.

‘If I have seen further it is by standing on the shoulders of Giants”

Isaac Newton (1675)

Chapter 5

Final Remarks

Science and innovation are our most valuable weapons to face and tackle the most pressing issues of our time (Jones, 2021). As pointed out in various parts of this dissertation, studying the production and diffusion of knowledge is challenging from a technical standpoint. Ideas are “public goods”, and markets fail to provide sufficient resources for novel research avenues. Moreover, standard economic models hardly capture the actual scope of knowledge flows in a globalized economy. Fortunately, big data allow us to zoom into the origins of scientific and technical advances, exploiting micro-level information, and to identify what factors drive the behavior of innovators (i.e., scientists, inventors, and firms) – offering a promising opportunity to tackle the complexity that characterizes the production and diffusion of knowledge. Across the chapters of this thesis, we investigated how social, technological, and geographical proximity affect innovative activities. By employing network analysis and statistical models, we could keep track of knowledge flows at different scales (i.e., industries, scientists, research areas) and link them to knowledge production and diffusion.

In chapter II, we provided evidence that the position of industries in the intersectoral knowledge space boosts their competitiveness. However, geography still moderates the transmission of tacit knowledge. Next, in chapter III, we showed that knowledge and social proximity drive scientists' research portfolio diversification, but science is more and more a social enterprise. When scientists move far from their core specialization, they rely more heavily on collaborations. Lastly, in chapter

IV, we focused on one of the leading climate change mitigation technologies: Negative Emissions Technologies (NETs). We quantified the multidimensional nature of knowledge spillovers generated from basic negative emissions research. Our analysis suggests that NETs are still relatively far from practical, innovative applications, except for Direct Air Capture (DAC), which shows the most significant potential in terms of technological developments. Interestingly, policymakers overlook DAC research relative to other NET options, preferring, e.g., Bio-energy with Carbon Capture and Storage (BECCS); currently, DAC is close to the technological frontier but distant from the social and policy debate.

Overall, through the chapters of this thesis, we argued that proximity - in several dimensions - still plays a crucial role in shaping how knowledge flows and why it matters. Yet, science and technology are inherently complex systems, defined by intertwined and evolving networks of ideas, publications, inventions, individual innovators, and organizations. The evidence collected in this dissertation only represents a starting point – with all the limitations we articulated in each chapter. The essential question then becomes: where do we go from here? We believe there are at least two potential directions worth exploring.

First, it would be valuable to build a general framework to characterize the heterogeneity of innovative activities – akin to those used for mobility patterns and social interactions. In other words, a data-driven framework to analyze mobility in the knowledge space. Thanks to dimension reduction tools, we can map scientific fields and technologies into locations of an abstract knowledge space, where patent and publication sequences can be charted as trajectories. Accordingly, human mobility models and measures can be adapted to analyze patterns of technological innovation and scientific discovery in a rich and novel fashion. Moreover, once such trajectories in the knowledge space are obtained, many insightful mobility measures, such as the radius of gyration or the entropy of individual whereabouts, become readily extendable to this context. The first step in this direction is to investigate how mobility patterns in the knowledge space affect productivity and impact of authors, disentangling the role of exploration and specialization.

Second, it would be valuable to quantify the role of scientific advances in climate-related technologies and linking them to practical innovation. As in other historical challenges (e.g., COVID-19, World War II), innovation may play an essential role in dealing with the climate crisis. Indeed, the so-called *green* transition will inherently be a technological transition whose hurdles and complexities are hard to quantify. As a first potential step toward a quantitative assessment of the science-technology nexus in mitigation technologies, we plan to investigate how geographical, social, and technological proximity between scientists and inventors affect the development of mitigation technologies.

In the long-run, economic prosperity critically depends on a "black-box" called technological progress. Thus, a quantitative understanding of how scientific and innovative endeavors unfold is crucial for designing effective policy responses to global challenges. Interdisciplinarity will be key for boosting the effectiveness of our research efforts. As Christopher Freeman pointed out in 1974 "Innovation is far too important to be left to scientists and technologists. It is also far too important to be left to economists or social scientists."(Freeman, 1974).

Appendices

Appendix A

Supplementary Information Chapter2

A.1 Additional measures & figures

The empirical analysis confirmed a prevalent effect of the "National" dimension in a knowledge space. Table A.1 summarizes some basic results we obtained using two alternative centrality measures and estimating the corresponding specifications by pooled OLS. In particular, we consider degree centrality (unweighted) and eigenvector centrality. Formally, for a network $\Gamma = (\mathcal{I}, \mathcal{L})$ with adjacency matrix $\mathbf{A}$ whose elements are such that:

$$a_{ij} = \begin{cases} 1 & \text{if } (i, j) \in \mathcal{L} \\ 0 & \text{Otherwise} \end{cases}$$

we can write degree centrality as the number of ties that involve a given node:

$$d_i = \sum_{j \in \mathcal{I}} a_{ij}$$

Furthermore, eigenvector centrality can be written as the principal eigenvector of the adjacency matrix of the network:

$$\lambda \mathbf{v} = \mathbf{A}\mathbf{v}$$

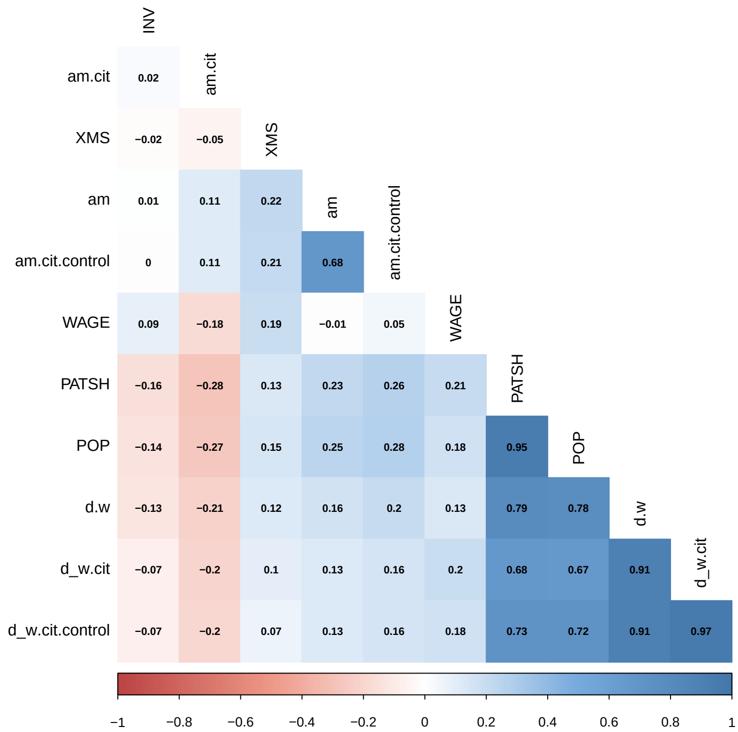


Figure A.1: Correlation Matrix including Robustness Measures (.control)

where λ is the eigenvalue and $\mathbf{v}$ the eigenvector. Of course, different centrality measures produce different results since they capture different features of the network (see Section 2.4.3 and Borgatti (2005) for a brief and intuitive description of centrality measures). Nevertheless, we can see that the alternative specifications we use provide insights that go in the same direction of the effects highlighted and discussed in the main text. In particular, both degree and eigenvector centrality at national level are positively associated with export market shares, although in a much less robust way.

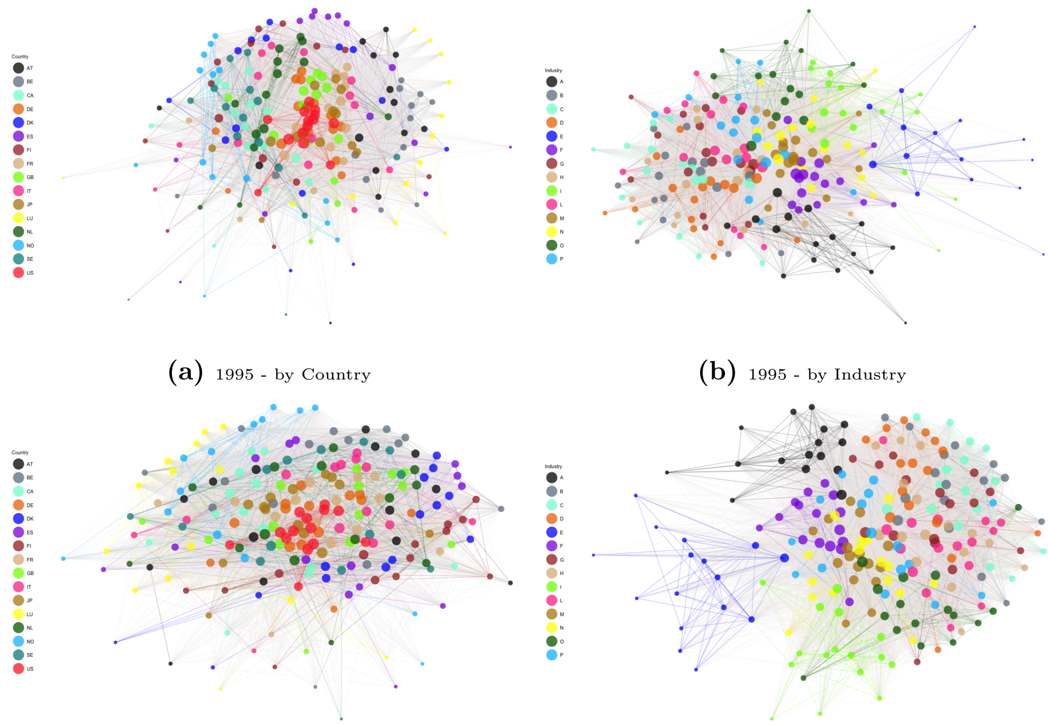
Figure A.1 summarizes how our network measures correlate with each other. Control variables, used as robustness check (d.w.cit.control and am.cit.control) refer, respectively, to weighted degree and local clustering, computed at the national level but using citations instead of co-occurrences.

Figure A.2 describes two knowledge networks based on patent citations in two different points in time (1995 and 2009). To be as clear as possible, we aggregate data by country and by industry. Edges' colors are set according to intra-industry

Table A.1: Alternative Centrality Measures

	Dependent variable: XMS					
	Degree			Eigenvector		
	(National)	(International)	(Final)	(National)	(International)	(Final)
PATSH	0.127^{***}	0.118^{***}	0.118^{***}	0.117^{***}	0.122^{***}	0.115^{***}
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.020)
WAGE	0.015^{***}	0.015^{***}	0.014^{***}	0.016^{***}	0.016^{***}	0.015^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
INV	0.053^{***}	0.053^{***}	0.052^{***}	0.050^{***}	0.051^{***}	0.051^{***}
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
POP	−0.035	−0.034	−0.035	−0.034	−0.032	−0.033
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
d	0.007^{***}		0.003			
	(0.002)		(0.003)			
ev				0.009^{***}		0.009^{***}
				(0.003)		(0.003)
am	0.043^{***}		0.045^{***}	0.033^{***}		0.038^{***}
	(0.013)		(0.013)	(0.013)		(0.013)
d.cit		0.007^{***}	0.006^{**}			
		(0.003)	(0.003)			
ev.cit					0.002	−0.006
					(0.008)	(0.008)
am.cit		−0.073	−0.083		-0.123^{**}	-0.130^{***}
		(0.052)	(0.052)		(0.049)	(0.049)
Observations	2,778	2,778	2,778	2,778	2,778	2,778
R^2	0.556	0.556	0.558	0.556	0.554	0.557
Adjusted R^2	0.546	0.546	0.547	0.546	0.545	0.547
F Statistic	71.163^{***} (df = 48; 2729)	71.116^{***} (df = 48; 2729)	68.764^{***} (df = 50; 2727)	71.140^{***} (df = 48; 2729)	70.740^{***} (df = 48; 2729)	68.571^{***} (df = 50; 2727)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$



(a) 1995 - by Country

(b) 1995 - by Industry

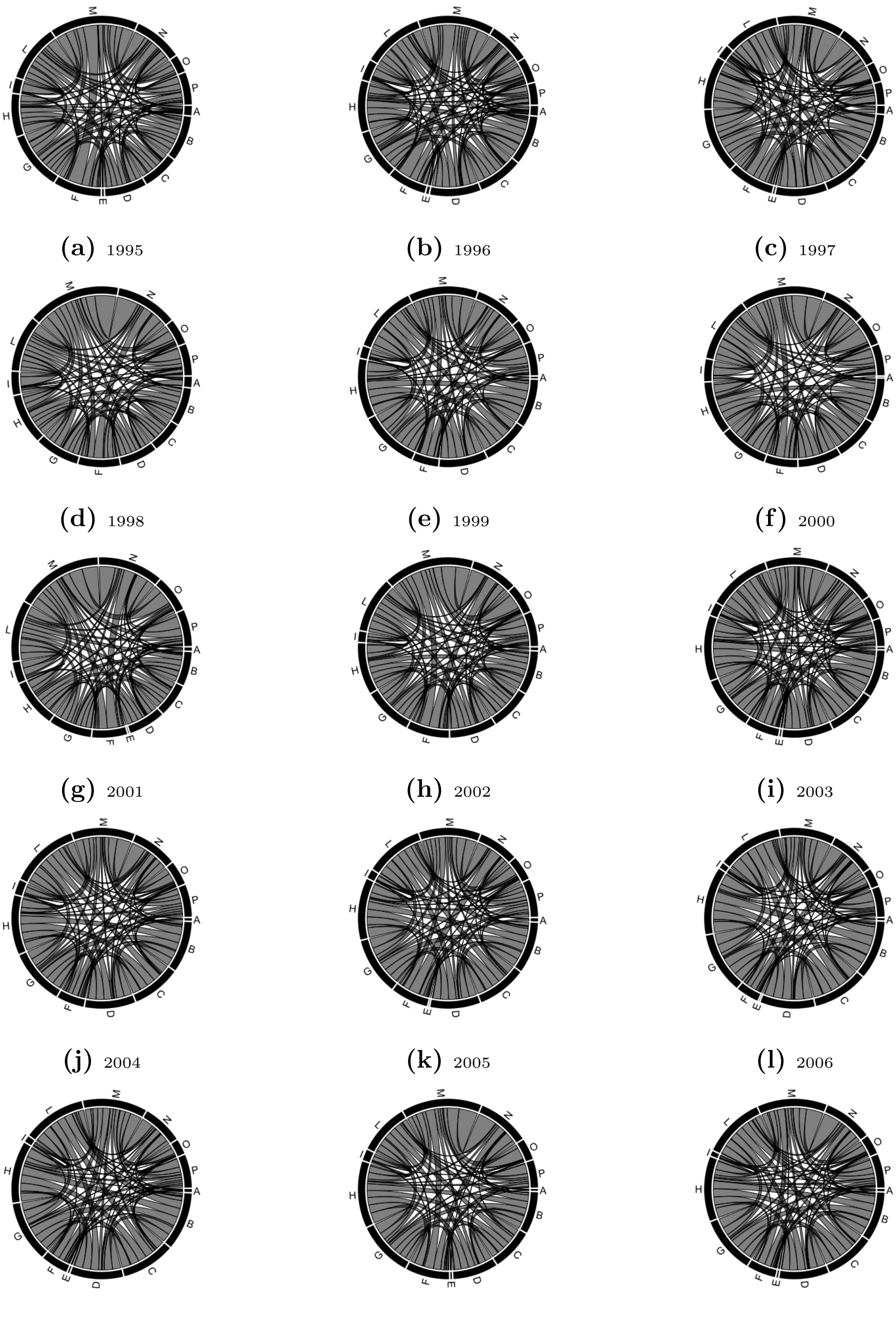
(c) 2009 - By Country

(d) 2009 - by Industry

In graphs (a) and (b) industry E for Luxembourg has been dropped for ease of visualization.

Figure A.2: Citation Networks by Country and by Industry

(intra-country) relationships in order to capture citations' patterns. It's worth noticing that both intra-industry and intra-country links can help us to understand the underline mechanism of network formation. Figure A.3:A.18 show the "national" knowledge space evolution over time across all countries in our dataset.



(a) 1995 (b) 1996 (c) 1997

(d) 1998 (e) 1999 (f) 2000

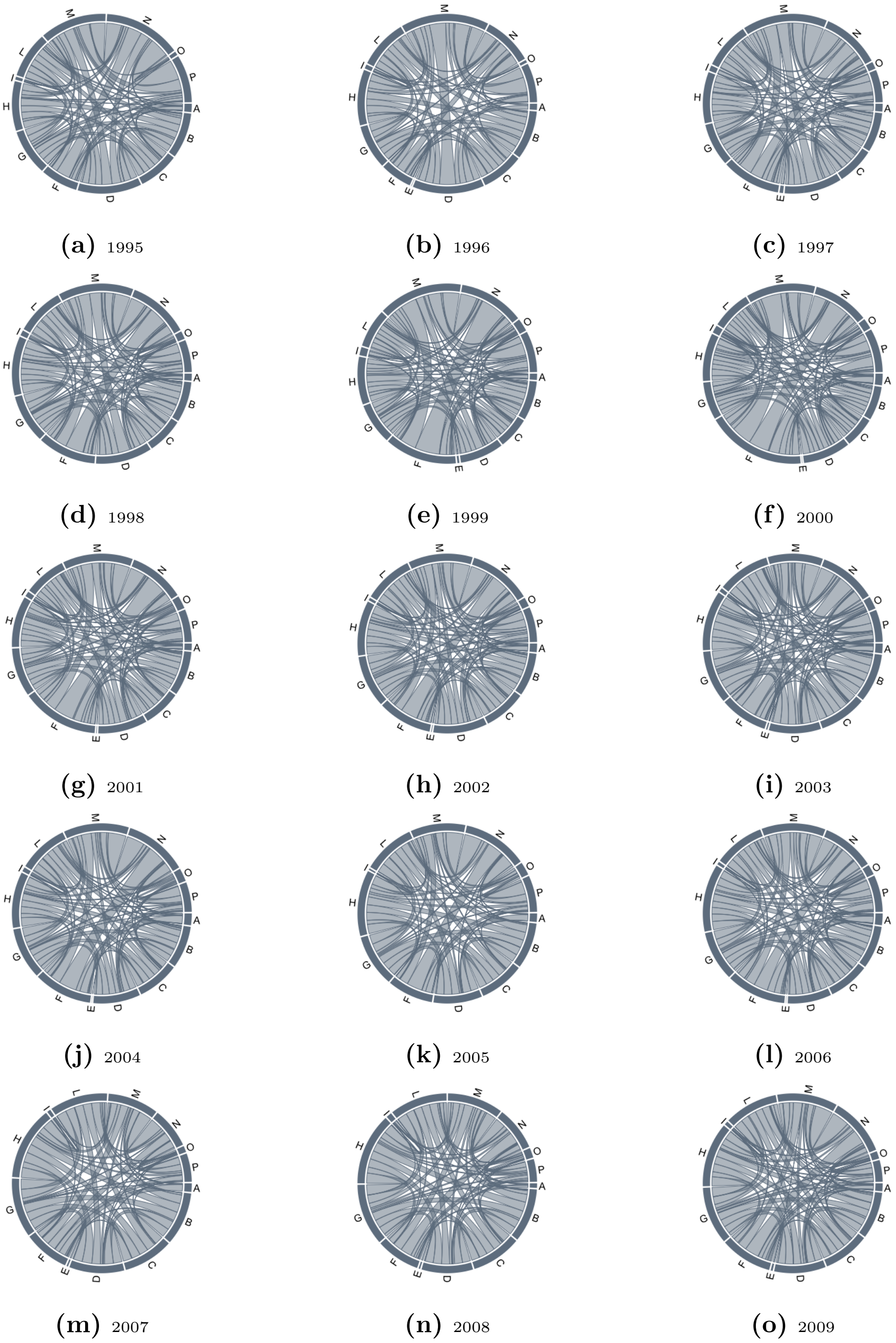
(g) 2001 (h) 2002 (i) 2003

(j) 2004 (k) 2005 (l) 2006

(m) 2007 (n) 2008 (o) 2009

(Co-occurences)

Figure A.3: Austria - Knowledge Space



(a) 1995 **(b)** 1996 **(c)** 1997

(d) 1998 **(e)** 1999 **(f)** 2000

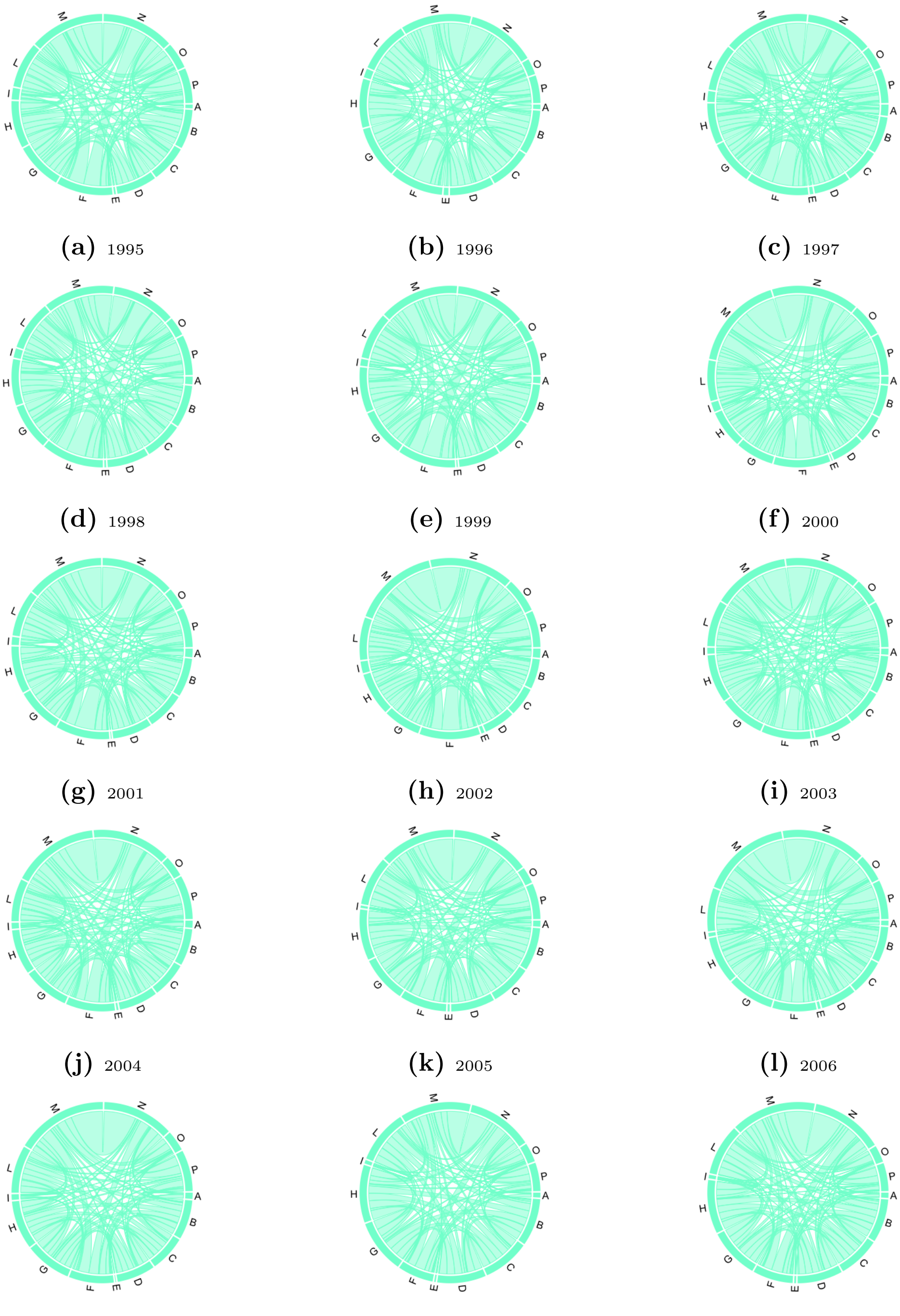
(g) 2001 **(h)** 2002 **(i)** 2003

(j) 2004 **(k)** 2005 **(l)** 2006

(m) 2007 **(n)** 2008 **(o)** 2009

(Co-occurences)

Figure A.4: Belgium - Knowledge Space



(a) 1995 **(b)** 1996 **(c)** 1997

(d) 1998 **(e)** 1999 **(f)** 2000

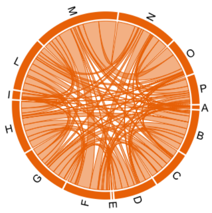
(g) 2001 **(h)** 2002 **(i)** 2003

(j) 2004 **(k)** 2005 **(l)** 2006

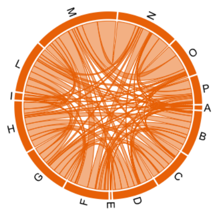
(m) 2007 **(n)** 2008 **(o)** 2009

(Co-occurences)

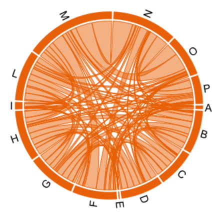
Figure A.5: Canada - Knowledge Space



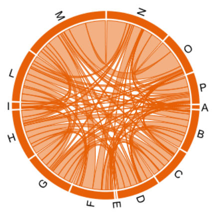
(a) 1995



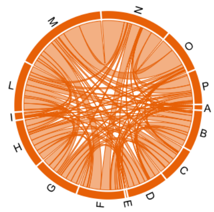
(b) 1996



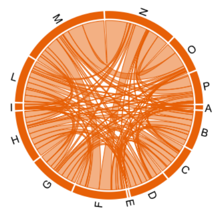
(c) 1997



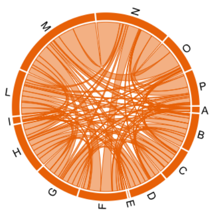
(d) 1998



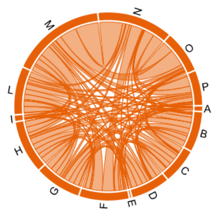
(e) 1999



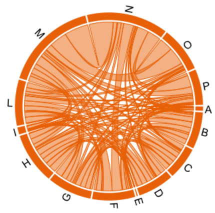
(f) 2000



(g) 2001

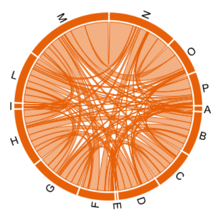


(h) 2002

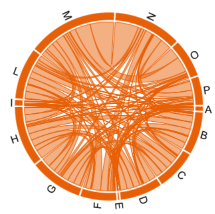


(i) 2003

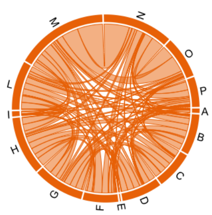
(j) 2004



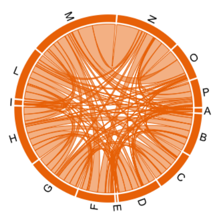
(k) 2005



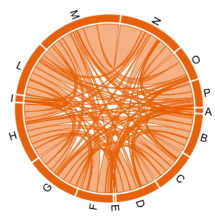
(l) 2006



(m) 2007



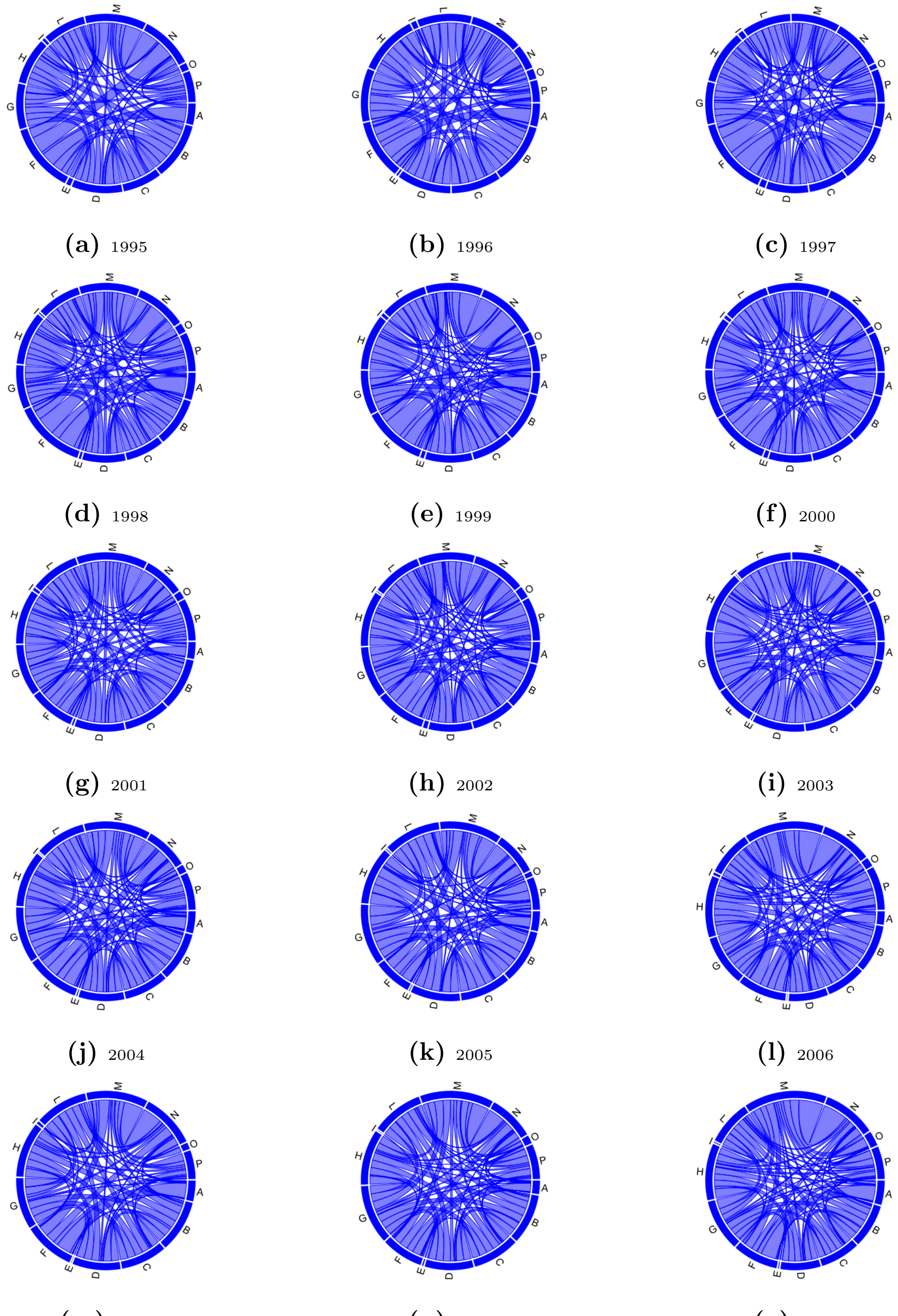
(n) 2008



(o) 2009

(Co-occurences)

Figure A.6: Germany - Knowledge Space



(a) 1995 (b) 1996 (c) 1997

(d) 1998 (e) 1999 (f) 2000

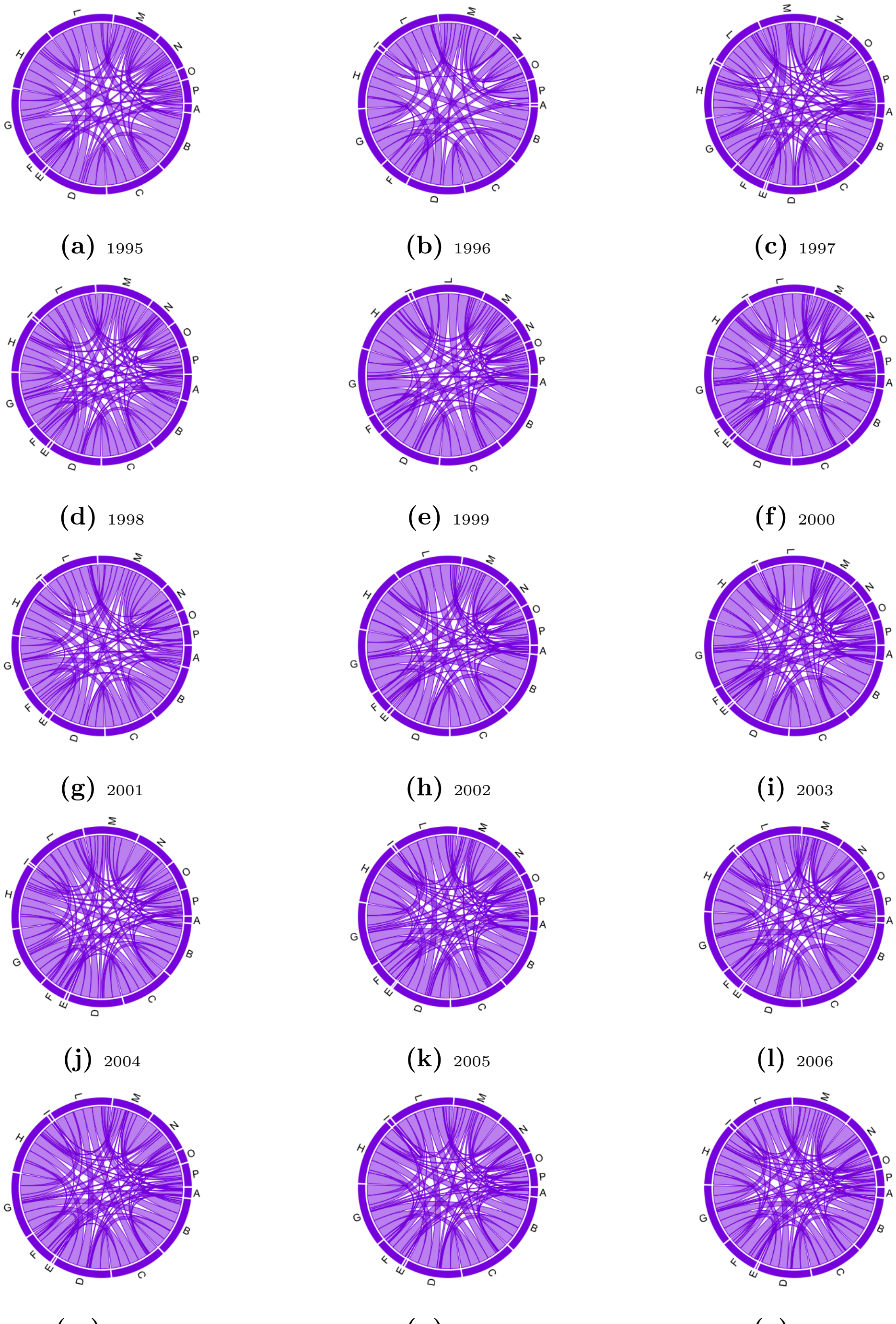
(g) 2001 (h) 2002 (i) 2003

(j) 2004 (k) 2005 (l) 2006

(m) 2007 (n) 2008 (o) 2009

(Co-occurences)

Figure A.7: Denmark - Knowledge Space



(a) 1995 (b) 1996 (c) 1997

(d) 1998 (e) 1999 (f) 2000

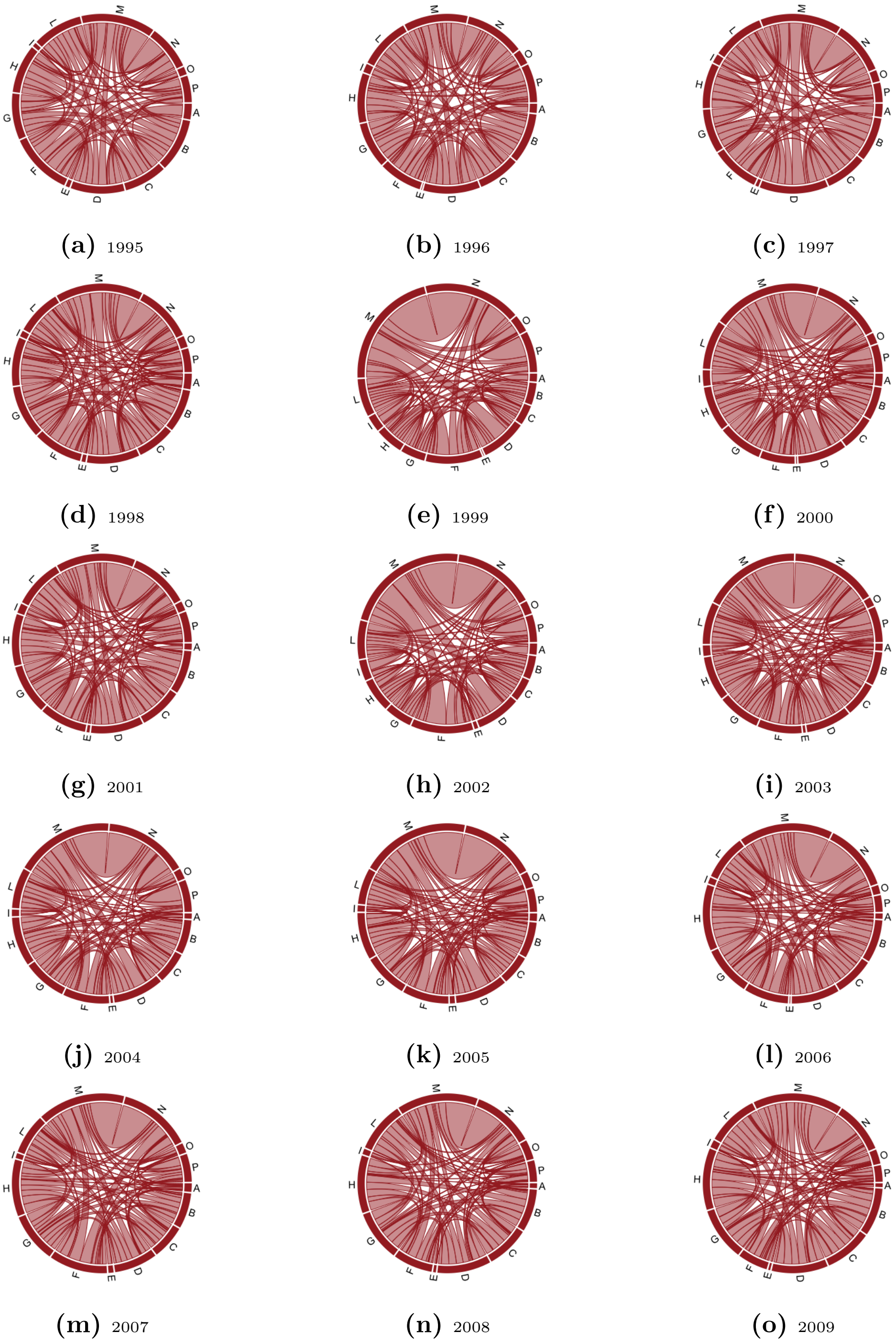
(g) 2001 (h) 2002 (i) 2003

(j) 2004 (k) 2005 (l) 2006

(m) 2007 (n) 2008 (o) 2009

(Co-occurences)

Figure A.8: Spain - Knowledge Space



(a) 1995 (b) 1996 (c) 1997

(d) 1998 (e) 1999 (f) 2000

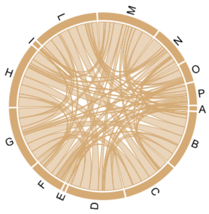
(g) 2001 (h) 2002 (i) 2003

(j) 2004 (k) 2005 (l) 2006

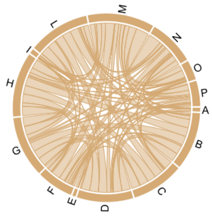
(m) 2007 (n) 2008 (o) 2009

(Co-occurences)

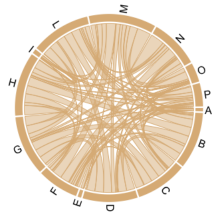
Figure A.9: Finland - Knowledge Space



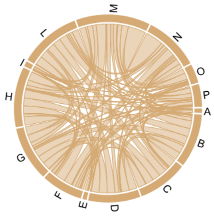
(a) 1995



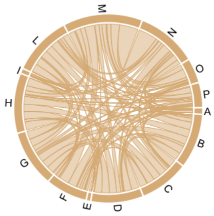
(b) 1996



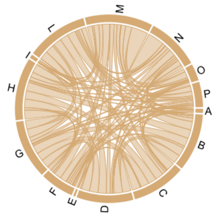
(c) 1997



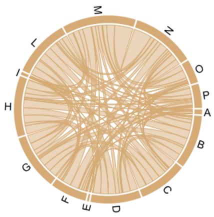
(d) 1998



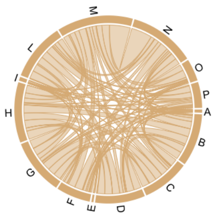
(e) 1999



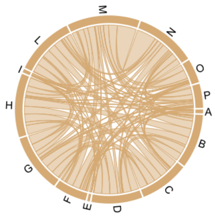
(f) 2000



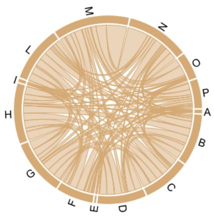
(g) 2001



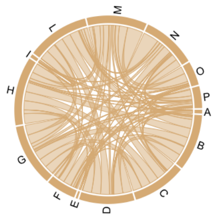
(h) 2002



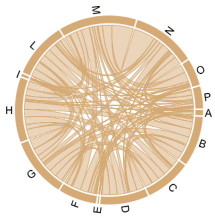
(i) 2003



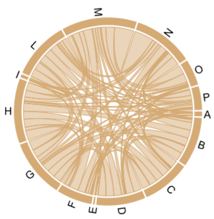
(j) 2004



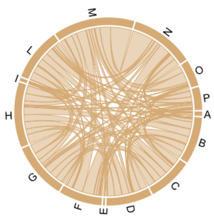
(k) 2005



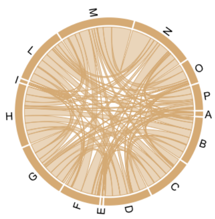
(l) 2006



(m) 2007



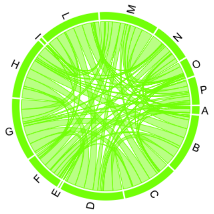
(n) 2008



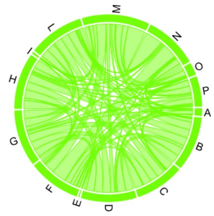
(o) 2009

(Co-occurences)

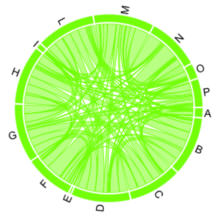
Figure A.10: France - Knowledge Space



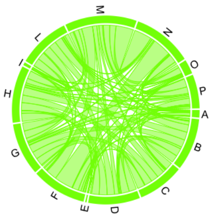
(a) 1995



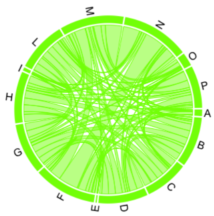
(b) 1996



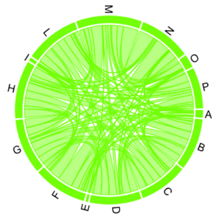
(c) 1997



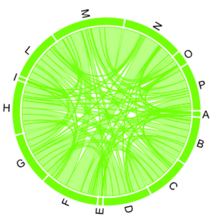
(d) 1998



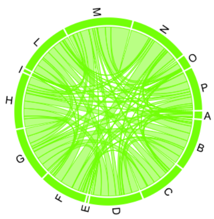
(e) 1999



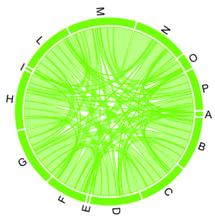
(f) 2000



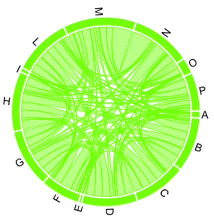
(g) 2001



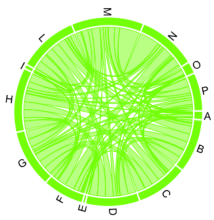
(h) 2002



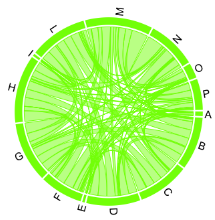
(i) 2003



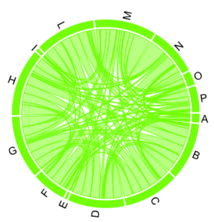
(j) 2004



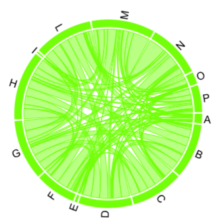
(k) 2005



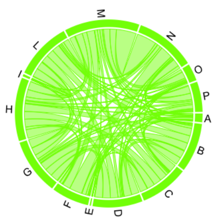
(l) 2006



(m) 2007



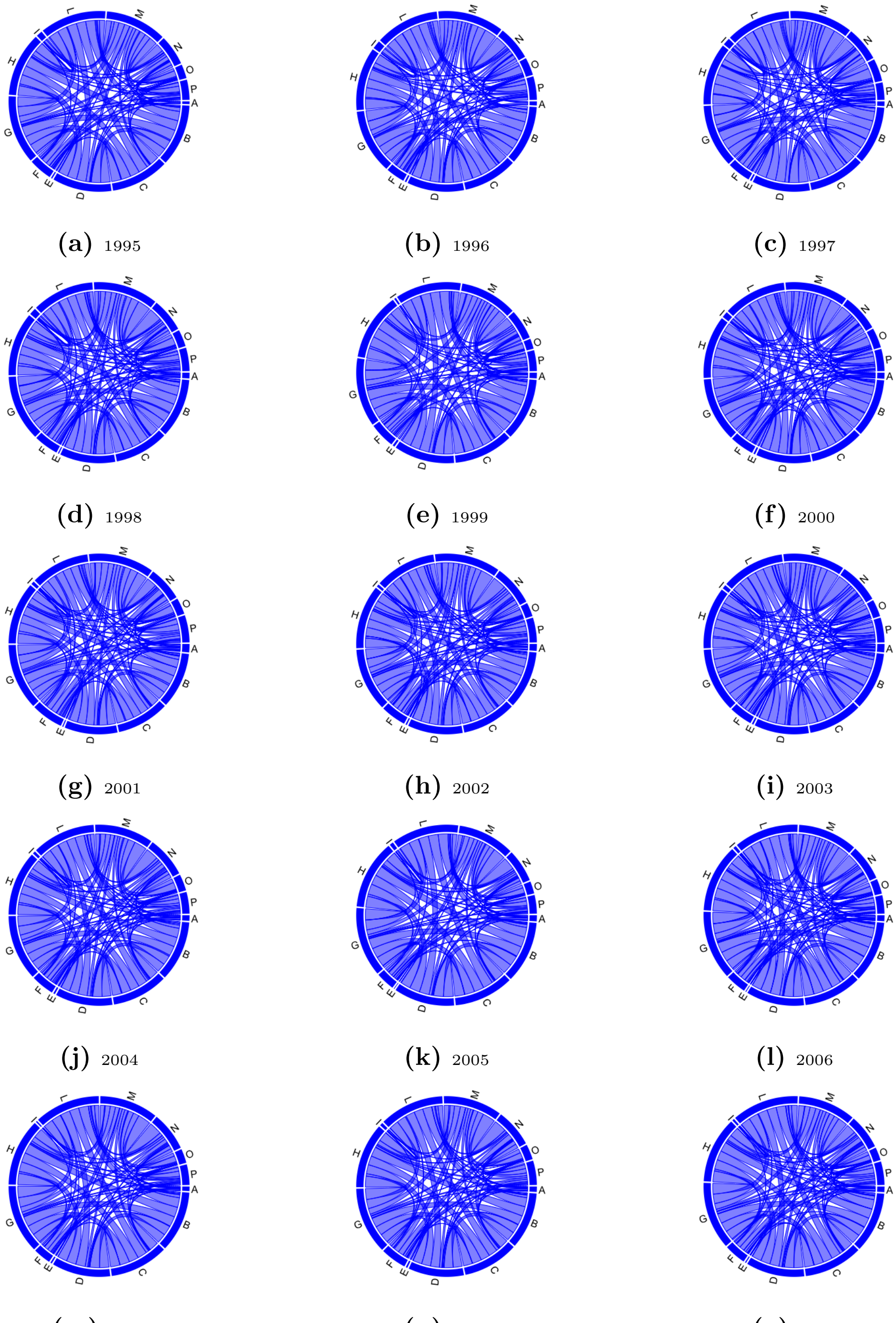
(n) 2008



(o) 2009

(Co-occurences)

Figure A.11: United Kingdom - Knowledge Space



(a) 1995 (b) 1996 (c) 1997

(d) 1998 (e) 1999 (f) 2000

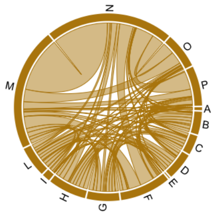
(g) 2001 (h) 2002 (i) 2003

(j) 2004 (k) 2005 (l) 2006

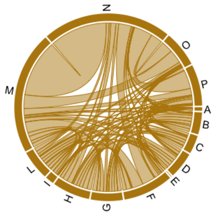
(m) 2007 (n) 2008 (o) 2009

(Co-occurences)

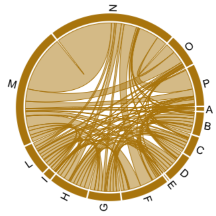
Figure A.12: Italy - Knowledge Space



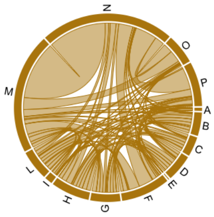
(a) 1995



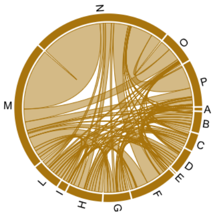
(b) 1996



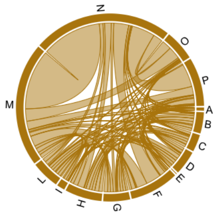
(c) 1997



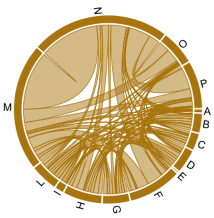
(d) 1998



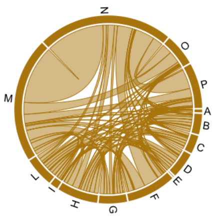
(e) 1999



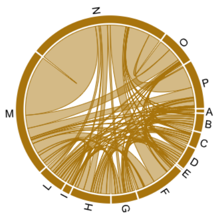
(f) 2000



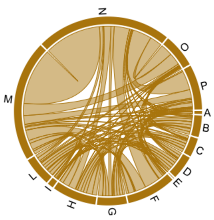
(g) 2001



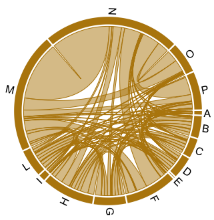
(h) 2002



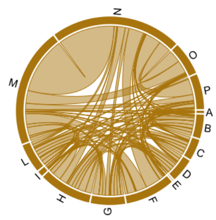
(i) 2003



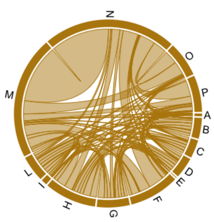
(j) 2004



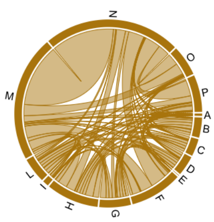
(k) 2005



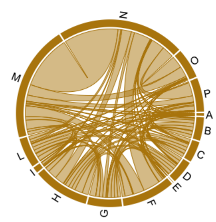
(l) 2006



(m) 2007



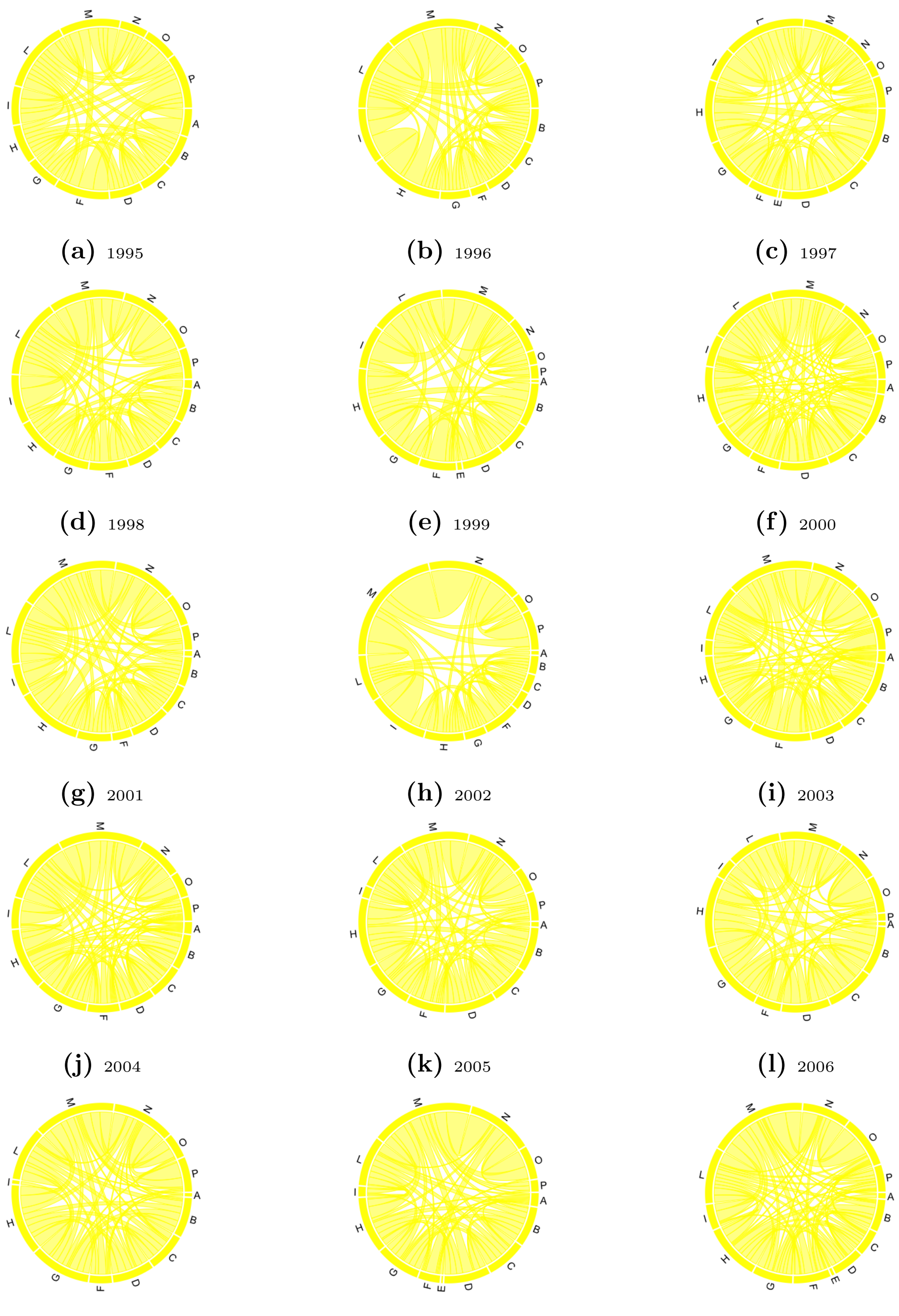
(n) 2008



(o) 2009

(Co-occurences)

Figure A.13: Japan - Knowledge Space



(a) 1995 (b) 1996 (c) 1997

(d) 1998 (e) 1999 (f) 2000

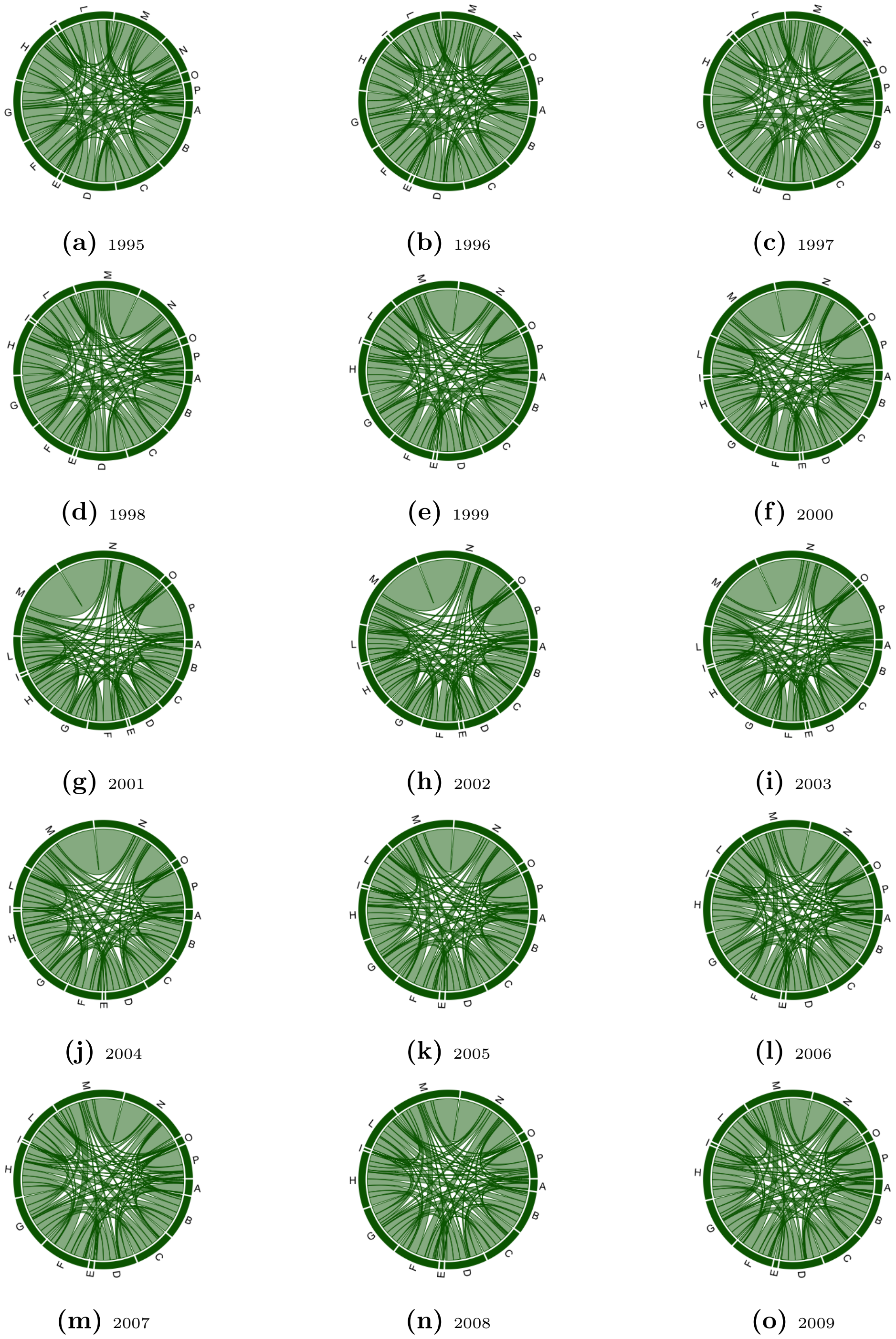
(g) 2001 (h) 2002 (i) 2003

(j) 2004 (k) 2005 (l) 2006

(m) 2007 (n) 2008 (o) 2009

(Co-occurences)

Figure A.14: Luxembourg - Knowledge Space



(a) 1995 **(b)** 1996 **(c)** 1997

(d) 1998 **(e)** 1999 **(f)** 2000

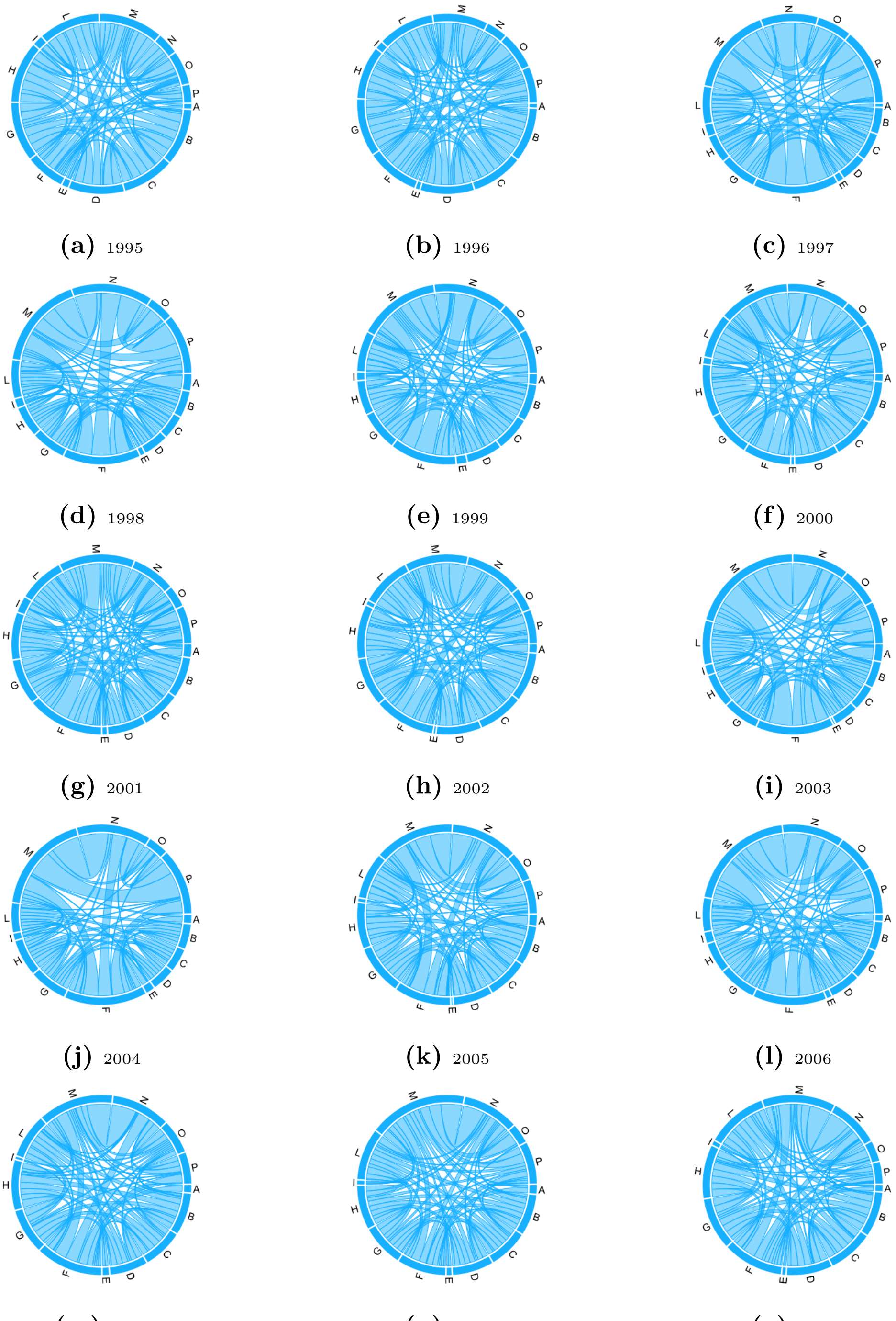
(g) 2001 **(h)** 2002 **(i)** 2003

(j) 2004 **(k)** 2005 **(l)** 2006

(m) 2007 **(n)** 2008 **(o)** 2009

(Co-occurences)

Figure A.15: Netherlands - Knowledge Space



(a) 1995 (b) 1996 (c) 1997

(d) 1998 (e) 1999 (f) 2000

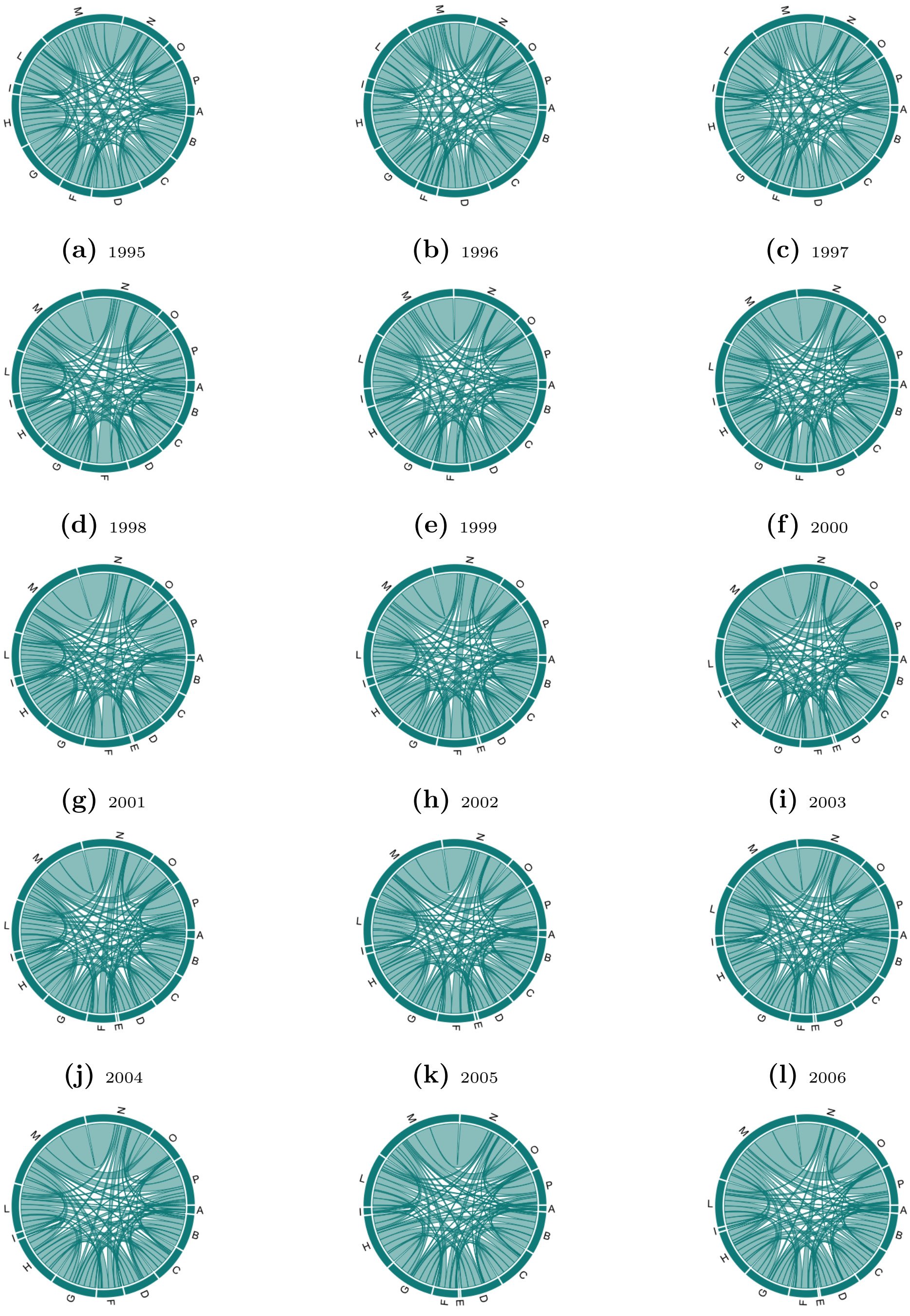
(g) 2001 (h) 2002 (i) 2003

(j) 2004 (k) 2005 (l) 2006

(m) 2007 (n) 2008 (o) 2009

(Co-occurences)

Figure A.16: Norway - Knowledge Space



(a) 1995 (b) 1996 (c) 1997

(d) 1998 (e) 1999 (f) 2000

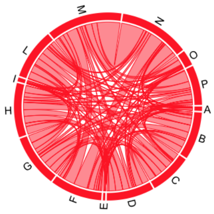
(g) 2001 (h) 2002 (i) 2003

(j) 2004 (k) 2005 (l) 2006

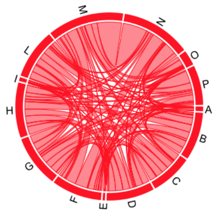
(m) 2007 (n) 2008 (o) 2009

(Co-occurences)

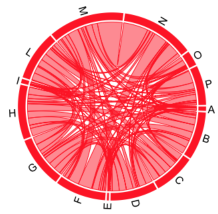
Figure A.17: Sweden - Knowledge Space



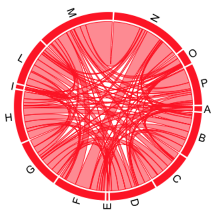
(a) 1995



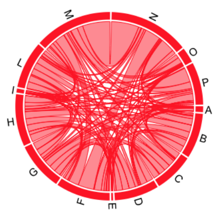
(b) 1996



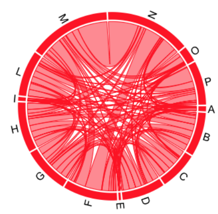
(c) 1997



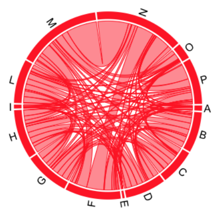
(d) 1998



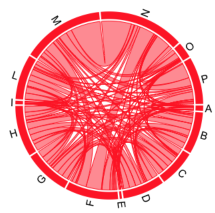
(e) 1999



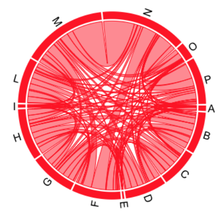
(f) 2000



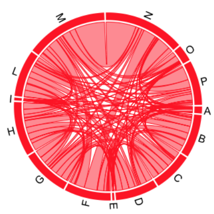
(g) 2001



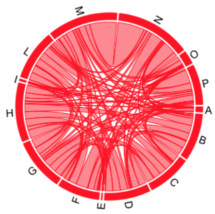
(h) 2002



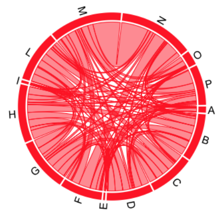
(i) 2003



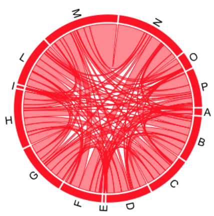
(j) 2004



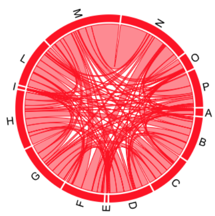
(k) 2005



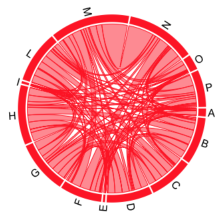
(l) 2006



(m) 2007



(n) 2008



(o) 2009

(Co-occurences)

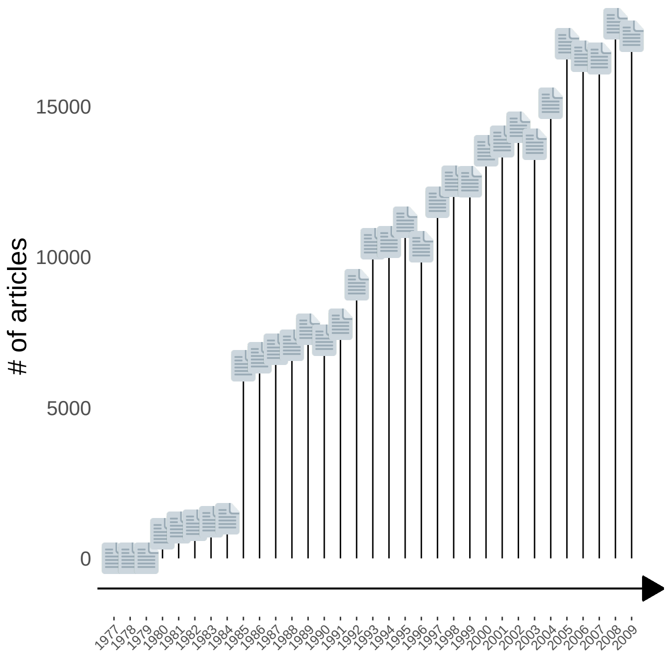
Figure A.18: United States - Knowledge Space

Appendix B

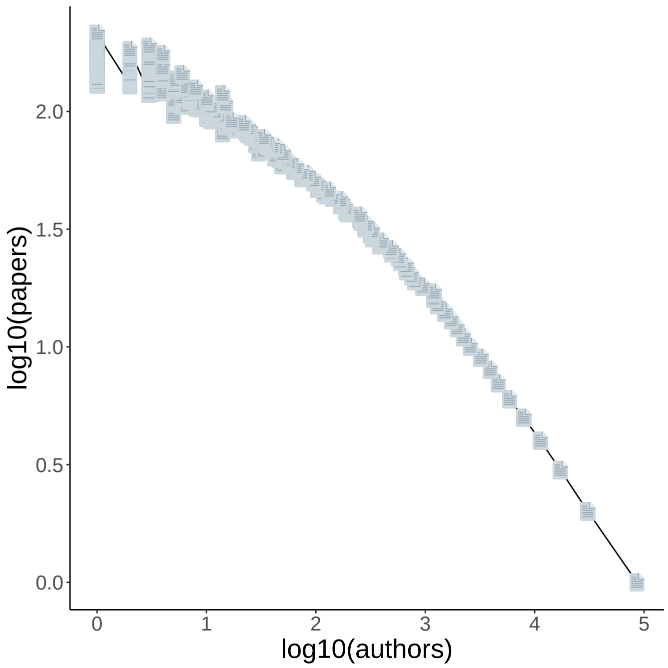
Supplementary Information Chapter3

B.1 Data

The American Physical Society (APS) grants access to data containing information about papers published in 9 journals: Physical Review A, B, C, D, E, I, L, ST and Review of Modern Physics. The APS makes available, under request, two datasets including over 450,000 articles metadata and citations from 1893 onwards. Each article has a unique identifier and most of them contain reference codes that map into physics sub-fields (PACS codes). As mentioned in Section 3.9.1, we make use of such a classification to keep track of scientists' diversification patterns. Moreover, we use a disambiguated list of authors made available by Sinatra et al. (2016). As a result, we analyse a sub-sample for which we have access to all the necessary information: it includes more than 300,000 articles published by 197,682 authors over the period 1977-2009. Figure B.1 provides simple statistical properties of the dataset.



a



b

Figure B.1: Statistical properties of the APS data. **a**, The time series of papers over time shows that the number of papers published in APS outlets increased substantially from 1977 to 2009. **b**, The distribution of the number of papers per author is fat-tailed: the large majority of authors published just few articles while some authors have been extremely productive.

Table B.1: One-digit PACS codes

PACS	Field	Description
0	General	Mathematical Methods, Quantum Mechanics, Relativity, Nonlinear Dynamics and Metrolog
1	High-energy	Physics of Elementary Particles and Fields
2	Nuclear	Nuclear Structure and Reactions
3	Atomic	Atomic and Molecular Physics
4	Classical	Electromagnetism, Optics, Acoustics, Heat Transfer, Classical Mechanics and Fluid Dynamics
5	Plasma	Physics of Gases, Plasmas and Electric Discharges
6 - 7	Condensed Matter	Structural, Mechanical and Thermal Properties, Electronic Structure and Electrical, Magnetic and Optical Properties
8	Interdisc	Interdisciplinary Physics and Related Areas of Science and Technology
9	Astro	Astrophysics, Astronomy and Geophysics

B.2 Test of randomness - multiple hypothesis correction

As stated in Section 3.3, under the hypothesis that scientists diversify their research portfolio at random, the probability that exactly x authors are active in two sub-fields follows a hypergeometric distribution (Tumminello et al., 2011). A clear advantage of such a formulation is that we can easily associate a p-value to each element (i.e., link in the projected network) ans evaluate the statistical significance. However, since we are performing hypothesis testing, we need to set a level of statistical significance accordingly.

Table B.2: Test of randomness in scientists' research portfolio diversification

	Positive	**Negative**	**% Non-Random**
No correction	1361	616	86.8
Bonferroni	1151	486	71.8
BH	1339	580	84.2
BY	1264	547	79.4

Note: 2,278 pairs analyzed, 68 sub-fields, 197,682 scientists. Analysis performed employing the **R** package *cooccur*.

B.3 Additional estimation results

As mentioned Section 3.9.5 and 3.4, we use a multivariate logistic regression to estimate the probability that a scientist diversifies in a sub-field different from her own specialization. Table B.3 summarizes our independent variables and includes information about our grouping strategy. Here, we provide results for each and every specification: (i) single specialization (full diversification), (ii) multiple-specialization (full diversification), (iii) single specialization (within field diversification), (iv) multiple specialization (within field diversification), (v) single specialization (between field diversification) and (vi) multiple specialization (between field diversification).

Table B.3: Variables and grouping strategy

Name	Group	Description
Knowledge relatedness	1 - KR	Cosine similarity among sub-fields
Social relatedness	2 - SR	Scientist' co-authors specialized in the sub-field different from her core one (dummy)
Field core	3 - IF	macro-field specialization (categorical)
# of PACS	4 - IF	Number of PACS explored
# of papers	4 - IF	Number of papers published
# of co-authors	4 - IF	Number of co-authors
PACS target popularity	5 - SC	Number of articles assigned to the target sub-field
Δ crowd	5 - SC	Difference in the number of specialized scientists between core and target sub-field
Δ PACS citations	6 - Cit	Difference in the number citations between core and target sub-field
Δ field citations	6 - Cit	Difference in the number citations between core and target macro-field

Full diversification - Specification (i) and (ii) Results are summarized in Table B.4 and B.5, where the first column refers to the baseline model (without the interaction term between social and knowledge relatedness), column (2) refers to the model including the interaction term while column (3) presents the same results with clustering corrected standard errors. Figure 3.5-**a**/**c** show the differences in the probability of diversification as a function of knowledge and social relatedness, taking into account all the control variables. Figure 3.5-**b**/**d** provide evidence of the moderating role played by the similarity across sub-fields on the estimated coefficient of social relatedness

Table B.4: (i) Single specialization - full diversification.

	Dependent variable: P(diversification)		
	Baseline	Interactions	Robust SE
	(1)	(2)	(3)
Knowledge Relatedness	0.846***	0.936***	0.936***
	(0.002)	(0.002)	(0.003)
Social Relatedness	2.647***	2.827***	2.827***
	(0.004)	(0.005)	(0.006)
field core-Atomic	−0.332***	−0.332***	−0.332***
	(0.009)	(0.009)	(0.010)
field core-Classical	−0.480***	−0.490***	−0.490***
	(0.010)	(0.010)	(0.010)
field core-Cond.matter	−1.094***	−1.088***	−1.088***
	(0.010)	(0.010)	(0.012)
field core-General	−0.710***	−0.722***	−0.722***
	(0.010)	(0.010)	(0.011)
field core-High.energy	0.221***	0.219***	0.219***
	(0.011)	(0.011)	(0.010)
field core-Interdisc	−0.546***	−0.557***	−0.557***
	(0.009)	(0.009)	(0.010)
field core-Nuclear	0.438***	0.463***	0.463***
	(0.009)	(0.009)	(0.010)
field core-Plasma	−0.258***	−0.269***	−0.269***
	(0.014)	(0.014)	(0.013)
# of PACS	0.891***	0.882***	0.882***
	(0.003)	(0.003)	(0.002)
# of papers	−0.007**	0.010***	0.010***
	(0.003)	(0.003)	(0.002)
PACS target popularity	1.130***	1.130***	1.130***
	(0.003)	(0.003)	(0.002)
Δ crowd	0.239***	0.239***	0.239***
	(0.002)	(0.002)	(0.002)
# of co-authors	−0.382***	−0.406***	−0.406***
	(0.003)	(0.003)	(0.002)
Δ PACS citations	−0.272***	−0.273***	−0.273***
	(0.003)	(0.003)	(0.002)
Δ field citations	−0.167***	−0.156***	−0.156***
	(0.003)	(0.003)	(0.004)
KR:SR		−0.255***	−0.255***
		(0.004)	(0.004)
Constant	−3.749***	−3.812***	−3.812***
	(0.008)	(0.008)	(0.010)
Observations	7,072,386	7,072,386	7,072,386
Log Likelihood	−1,088,731.000	−1,086,281.000	−1,086,281.000
Akaike Inf. Crit.	2,177,498.000	2,172,600.000	2,172,600.000
Note:			*p<0.1; **p<0.05; ***p<0.01

Table B.5: (ii) Multiple-specialization - full diversification

	Dependent variable:		
	Y		
	Baseline	Interactions	Robust SE
	(1)	(2)	(3)
Knowledge Relatedness	0.628***	0.689***	0.689***
	(0.001)	(0.005)	(0.009)
Social Relatedness	4.221***	4.243***	4.243***
	(0.005)	(0.005)	(0.019)
field core-Atomic	−0.427***	−0.427***	−0.427***
	(0.004)	(0.004)	(0.007)
field core-Classical	−0.475***	−0.475***	−0.475***
	(0.005)	(0.005)	(0.007)
field core-Cond.matter	−0.761***	−0.760***	−0.760***
	(0.004)	(0.004)	(0.009)
field core-General	−0.537***	−0.537***	−0.537***
	(0.004)	(0.004)	(0.007)
field core-High.energy	0.165***	0.165***	0.165***
	(0.005)	(0.005)	(0.006)
field core-Interdisc	−0.552***	−0.552***	−0.552***
	(0.004)	(0.004)	(0.007)
field core-Nuclear	0.163***	0.163***	0.163***
	(0.004)	(0.004)	(0.006)
field core-Plasma	−0.409***	−0.409***	−0.409***
	(0.007)	(0.007)	(0.008)
# of PACS	0.768***	0.768***	0.768***
	(0.001)	(0.001)	(0.003)
# of papers	0.117***	0.117***	0.117***
	(0.001)	(0.001)	(0.003)
PACS target popularity	0.611***	0.611***	0.611***
	(0.001)	(0.001)	(0.002)
Δ crowd	0.359***	0.359***	0.359***
	(0.001)	(0.001)	(0.003)
# of co-authors	−0.346***	−0.346***	−0.346***
	(0.001)	(0.001)	(0.004)
Δ PACS citations	−0.333***	−0.333***	−0.333***
	(0.001)	(0.001)	(0.003)
Δ field citations	−0.071***	−0.070***	−0.070***
	(0.001)	(0.001)	(0.004)
KR:SR		−0.062***	−0.062***
		(0.005)	(0.010)
Constant	−5.855***	−5.877***	−5.877***
	(0.006)	(0.007)	(0.020)
Observations	35,562,394	35,562,394	35,562,394
Log Likelihood	−7,299,777.000	−7,299,692.000	−7,299,692.000
Akaike Inf. Crit.	14,599,590.000	14,599,421.000	14,599,421.000
Note:			*p<0.1; **p<0.05; ***p<0.01

Within field diversification - Specification (iii) and (iv). Figure B.2-**a**/**b** plots the results for the single specialization case: knowledge and social relatedness are still significant as well as their interaction, but the magnitude of the coefficients is smaller with respect to the full diversification case. In addition, when we consider the multiple specialization case (Figure B.2-**c**/**d**), coefficients shrink further and the interaction term between social and knowledge relatedness is no longer significant (see Table B.6 and B.7 for details).

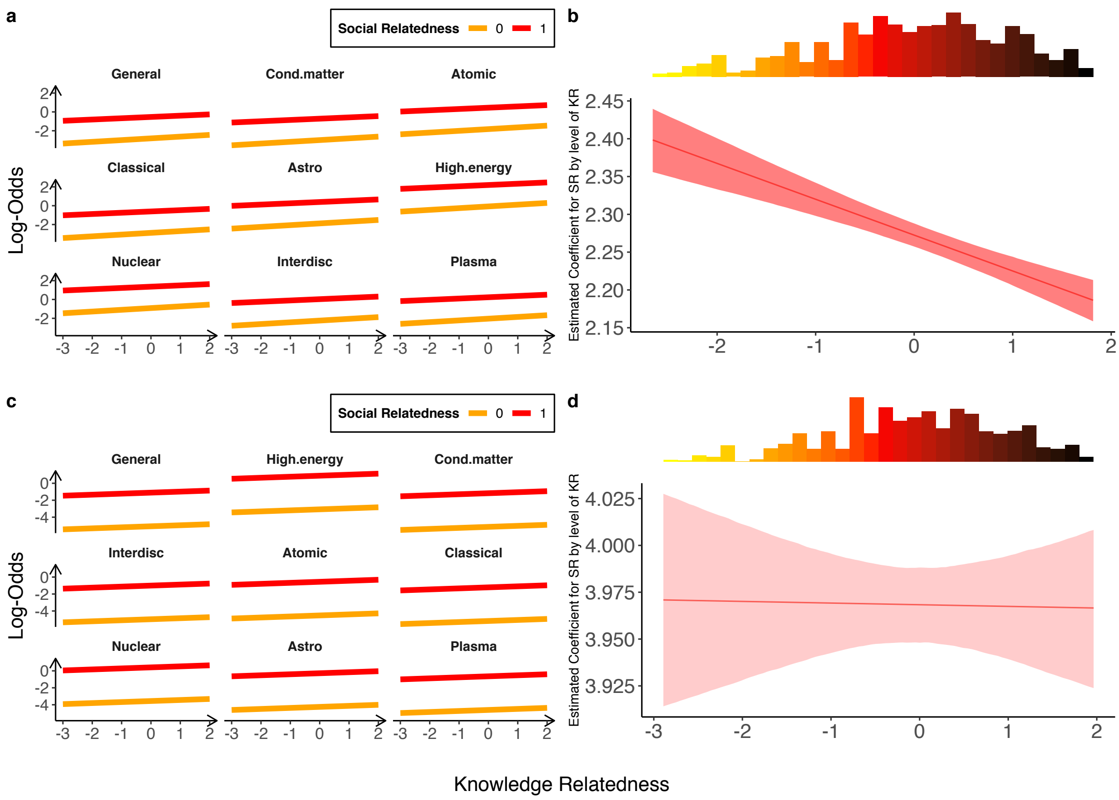


Figure B.2: Scientists' research portfolio diversification - (within field diversification) single and multiple specialization. **a**, Log-odds as a function of social and (standardized) knowledge relatedness, controlling for all the confounding variables - specification (iii). **b**, Estimated coefficient for social relatedness conditional on (standardized) knowledge relatedness - specification (iii). **c**, Log-odds as a function of social and (standardized) knowledge relatedness, controlling for the all confounding variables - specification (iv). **d**, Estimated coefficient for social relatedness conditional on (standardized) knowledge relatedness - specification (iv). **b** and **d** include the distribution of the conditional variable (i.e., knowledge relatedness). The color palette is in accordance with the similarity matrix (Figure 3.3-**c**.

Between field diversification - Specification (v) and (vi). As far as the between field diversification is concerned, the general trends in terms of social and

Table B.6: (iii) Single specialization - within field diversification

	Dependent variable:		
	Y		
	Baseline	Interactions	Robust SE
	(1)	(2)	(3)
Knowledge Relatedness	0.166***	0.184***	0.184***
	(0.004)	(0.004)	(0.005)
Social Relatedness	2.265***	2.272***	2.272***
	(0.008)	(0.008)	(0.008)
field core-Atomic	0.059***	0.056**	0.056**
	(0.022)	(0.022)	(0.025)
field core-Classical	−1.003***	−1.001***	−1.001***
	(0.026)	(0.026)	(0.029)
field core-Cond.matter	−1.108***	−1.110***	−1.110***
	(0.021)	(0.021)	(0.024)
field core-General	−0.931***	−0.927***	−0.927***
	(0.026)	(0.026)	(0.028)
field core-High.energy	1.809***	1.806***	1.806***
	(0.025)	(0.025)	(0.027)
field core-Interdisc	−0.353***	−0.357***	−0.357***
	(0.024)	(0.024)	(0.026)
field core-Nuclear	0.978***	0.969***	0.969***
	(0.021)	(0.021)	(0.024)
field core-Plasma	−0.149**	−0.155**	−0.155**
	(0.068)	(0.068)	(0.068)
# of PACS	0.769***	0.769***	0.769***
	(0.005)	(0.005)	(0.005)
# of papers	0.064***	0.065***	0.065***
	(0.006)	(0.006)	(0.005)
PACS target popularity	1.372***	1.370***	1.370***
	(0.006)	(0.006)	(0.005)
Δ crowd	0.130***	0.131***	0.131***
	(0.004)	(0.004)	(0.003)
# of co-authors	−0.239***	−0.240***	−0.240***
	(0.005)	(0.005)	(0.004)
Δ PACS citations	−0.209***	−0.208***	−0.208***
	(0.005)	(0.005)	(0.004)
KR:SR		−0.047***	−0.047***
		(0.007)	(0.007)
Constant	−1.883***	−1.882***	−1.882***
	(0.020)	(0.020)	(0.022)
Observations	1,000,230	1,000,230	1,000,230
Log Likelihood	−334,720.800	−334,697.300	−334,697.300
Akaike Inf. Crit.	669,475.700	669,430.600	669,430.600
Note:			*p<0.1; **p<0.05; ***p<0.01

Table B.7: (iv) Multiple specialization - within field diversification

	Dependent variable:		
	Y		
	Baseline	Interactions	Robust SE
	(1)	(2)	(3)
Knowledge Relatedness	0.121***	0.121***	0.121***
	(0.009)	(0.009)	(0.013)
Social Relatedness	3.968***	3.968***	3.968***
	(0.010)	(0.010)	(0.021)
field core-Atomic	−0.276***	−0.276***	−0.276***
	(0.011)	(0.011)	(0.021)
field core-Classical	−0.932***	−0.932***	−0.932***
	(0.013)	(0.013)	(0.023)
field core-Cond.matter	−0.892***	−0.892***	−0.892***
	(0.011)	(0.011)	(0.020)
field core-General	−0.823***	−0.823***	−0.823***
	(0.012)	(0.012)	(0.021)
field core-High.energy	1.176***	1.176***	1.176***
	(0.012)	(0.012)	(0.023)
field core-Interdisc	−0.724***	−0.724***	−0.724***
	(0.012)	(0.012)	(0.021)
field core-Nuclear	0.692***	0.692***	0.692***
	(0.011)	(0.011)	(0.021)
field core-Plasma	−0.361***	−0.361***	−0.361***
	(0.041)	(0.041)	(0.058)
# of PACS	0.497***	0.497***	0.497***
	(0.002)	(0.002)	(0.004)
# of papers	0.252***	0.252***	0.252***
	(0.002)	(0.002)	(0.004)
PACS target popularity	0.774***	0.774***	0.774***
	(0.002)	(0.002)	(0.003)
Δ crowd	0.345***	0.345***	0.345***
	(0.002)	(0.002)	(0.003)
# of co-authors	−0.145***	−0.145***	−0.145***
	(0.002)	(0.002)	(0.004)
Δ PACS citations	−0.313***	−0.313***	−0.313***
	(0.002)	(0.002)	(0.003)
KR:SR	−0.001	−0.001	−0.001
	(0.009)	(0.009)	(0.013)
Constant	−4.250***	−4.250***	−4.250***
	(0.014)	(0.014)	(0.028)
Observations	5,407,404	5,407,404	5,407,404
Log Likelihood	−2,166,803.000	−2,166,803.000	−2,166,803.000
Akaike Inf. Crit.	4,333,642.000	4,333,642.000	4,333,642.000
Note:			*p<0.1; **p<0.05; ***p<0.01

cognitive proximity are confirmed. Moreover, the negative interaction term remains statistically significant and not negligible in magnitude for both model specifications (single and multiple specialization). Figure B.3, Table B.8 and Table B.9 summarize the results.

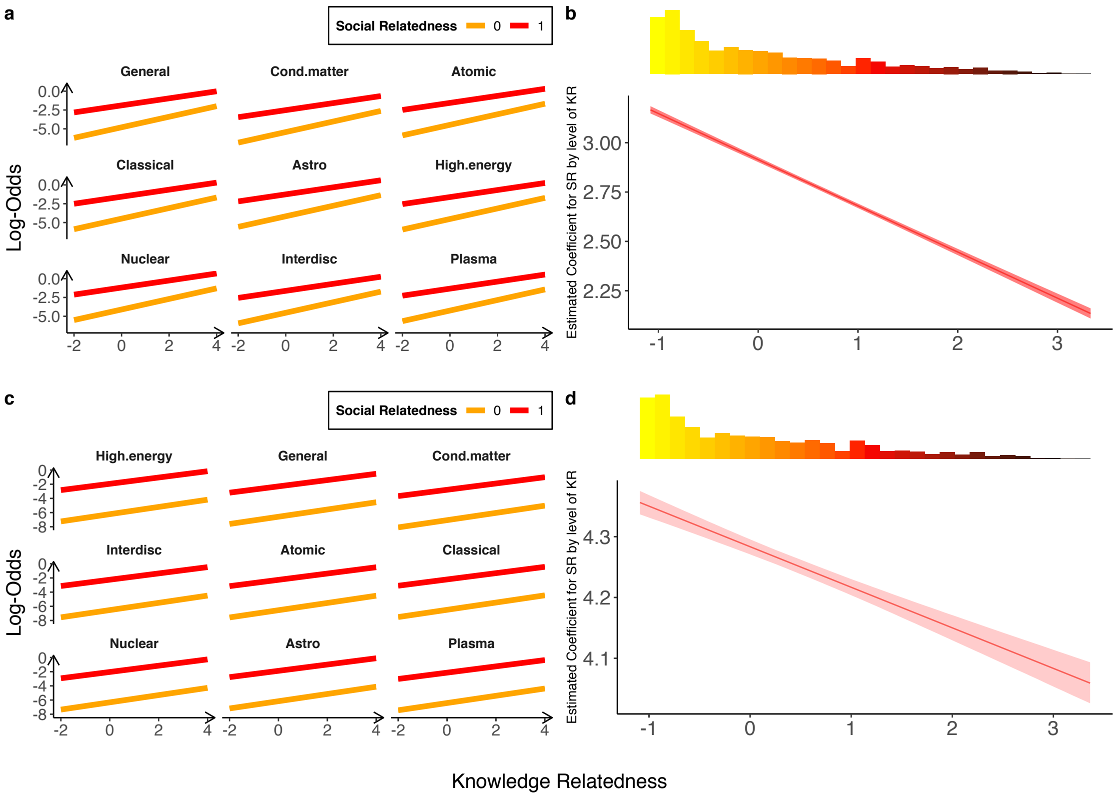


Figure B.3: Scientists' research portfolio diversification - (between field diversification) single and multiple specialization. **a**, Log-odds as a function of social and (standardized) knowledge relatedness, controlling for all the confounding variables - specification (v). **b**, Estimated coefficient for social relatedness conditional on (standardized) knowledge relatedness - specification (v). **c**, Log-odds as a function of social and (standardized) knowledge relatedness, controlling for all the confounding variables - specification (vi). **d**, Estimated coefficient for social relatedness conditional on (standardized) knowledge relatedness - specification (vi). **b** and **d** include the distribution of the conditional variable (i.e., knowledge relatedness). The color palette is in accordance with the similarity matrix (Figure 3.3-**c**).

Table B.8: (v) Single specialization - between field diversification

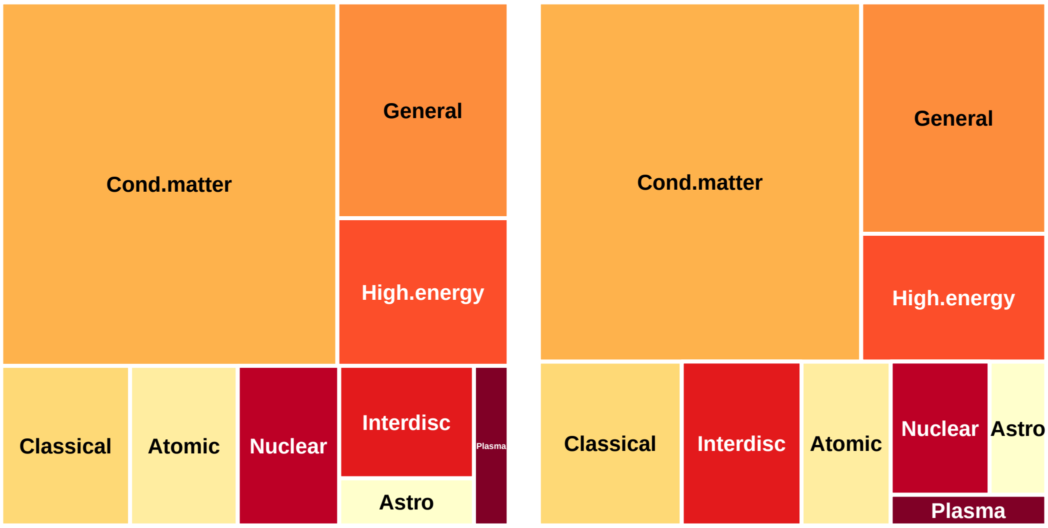
	Dependent variable:		
	Y		
	Baseline	Interactions	Robust SE
	(1)	(2)	(3)
Knowledge Relatedness	0.622***	0.702***	0.702***
	(0.002)	(0.003)	(0.003)
Social Relatedness	2.768***	2.914***	2.914***
	(0.006)	(0.006)	(0.008)
field core-Atomic	−0.292***	−0.303***	−0.303***
	(0.010)	(0.010)	(0.010)
field core-Classical	−0.304***	−0.313***	−0.313***
	(0.010)	(0.010)	(0.010)
field core-Cond.matter	−1.294***	−1.263***	−1.263***
	(0.013)	(0.013)	(0.017)
field core-General	−0.628***	−0.632***	−0.632***
	(0.011)	(0.011)	(0.012)
field core-High.energy	−0.352***	−0.360***	−0.360***
	(0.013)	(0.014)	(0.013)
field core-Interdisc	−0.356***	−0.365***	−0.365***
	(0.010)	(0.010)	(0.011)
field core-Nuclear	0.060***	0.068***	0.068***
	(0.011)	(0.011)	(0.011)
field core-Plasma	−0.062***	−0.074***	−0.074***
	(0.014)	(0.014)	(0.015)
# of PACS	1.010***	1.003***	1.003***
	(0.004)	(0.004)	(0.004)
# of papers	−0.050***	−0.032***	−0.032***
	(0.004)	(0.004)	(0.004)
PACS target popularity	1.114***	1.108***	1.108***
	(0.003)	(0.003)	(0.003)
Δ crowd	0.322***	0.320***	0.320***
	(0.003)	(0.003)	(0.003)
# of co-authors	−0.461***	−0.488***	−0.488***
	(0.003)	(0.003)	(0.003)
Δ PACS citations	−0.375***	−0.369***	−0.369***
	(0.004)	(0.004)	(0.004)
Δ field citations	−0.209***	−0.196***	−0.196***
	(0.004)	(0.004)	(0.006)
KR:SR		−0.234***	−0.234***
		(0.004)	(0.005)
Constant	−4.115***	−4.168***	−4.168***
	(0.009)	(0.009)	(0.010)
Observations	6,072,156	6,072,156	6,072,156
Log Likelihood	−717,839.000	−716,398.900	−716,398.900
Akaike Inf. Crit.	1,435,714.000	1,432,836.000	1,432,836.000
Note:			*p<0.1; **p<0.05; ***p<0.01

Table B.9: (vi) Multiple specialization - between field diversification

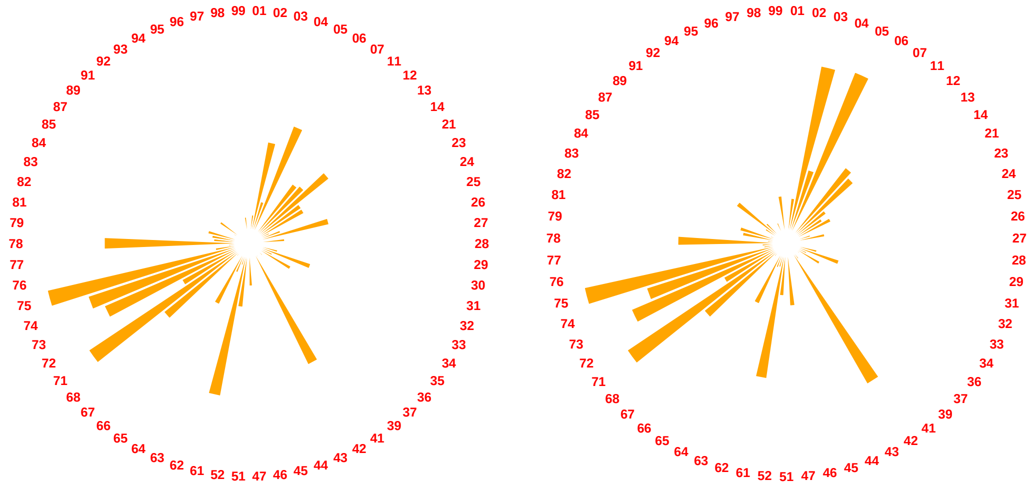
	Dependent variable:		
	Y		
	Baseline	Interactions	Robust SE
	(1)	(2)	(3)
cos	0.446***	0.511***	0.511***
	(0.001)	(0.005)	(0.011)
Social Relatedness	4.261***	4.284***	4.284***
	(0.006)	(0.006)	(0.021)
field core-Atomic	−0.384***	−0.385***	−0.385***
	(0.005)	(0.005)	(0.008)
field core-Classical	−0.328***	−0.328***	−0.328***
	(0.005)	(0.005)	(0.008)
field core-Cond.matter	−0.904***	−0.903***	−0.903***
	(0.005)	(0.005)	(0.013)
field core-General	−0.421***	−0.422***	−0.422***
	(0.005)	(0.005)	(0.008)
field core-High.energy	−0.060***	−0.060***	−0.060***
	(0.005)	(0.005)	(0.008)
field core-Interdisc	−0.367***	−0.367***	−0.367***
	(0.005)	(0.005)	(0.008)
field core-Nuclear	−0.160***	−0.161***	−0.161***
	(0.005)	(0.005)	(0.009)
field core-Plasma	−0.255***	−0.256***	−0.256***
	(0.007)	(0.007)	(0.009)
# of PACS	0.944***	0.944***	0.944***
	(0.001)	(0.001)	(0.004)
# of papers	0.050***	0.050***	0.050***
	(0.001)	(0.001)	(0.004)
PACS target popularity	0.559***	0.559***	0.559***
	(0.001)	(0.001)	(0.002)
Δ crowd	0.392***	0.393***	0.393***
	(0.002)	(0.002)	(0.003)
# of co-authors	−0.444***	−0.444***	−0.444***
	(0.001)	(0.001)	(0.006)
Δ PACS citations	−0.354***	−0.354***	−0.354***
	(0.002)	(0.002)	(0.003)
Δ field citations	−0.143***	−0.143***	−0.143***
	(0.002)	(0.002)	(0.005)
KR:SR		−0.067***	−0.067***
		(0.005)	(0.011)
Constant	−6.142***	−6.165***	−6.165***
	(0.007)	(0.008)	(0.022)
Observations	30,154,990	30,154,990	30,154,990
Log Likelihood	−4,971,576.000	−4,971,497.000	−4,971,497.000
Akaike Inf. Crit.	9,943,188.000	9,943,033.000	9,943,033.000
Note:			*p<0.1; **p<0.05; ***p<0.01

B.4 Temporal evolution of the physics knowledge space

The structure of the knowledge space can evolve over time, and sharp differences might undermine our strategy. To check whether such changes are significant, we split our initial dataset into three subsets, one for each decade: 1980-1989, 1990-1990, 2000-2009. We compare the structure of the physics knowledge space in the last decade of our sample with the one referring to the entire period. Figure B.4 compares popularity of one- and two-digit PACS in the last decade with the one for the full sample. Figure B.5 shows how the network and, as a consequence, the cosine similarity matrix have changed in the last ten years. Figure B.6 shows the popularity of one- and two-digit PACS in the three decades. Data confirm the rise of interdisciplinary physics within an otherwise stable distribution of interests, as much as it was observed in previous works (Pan et al., 2012).

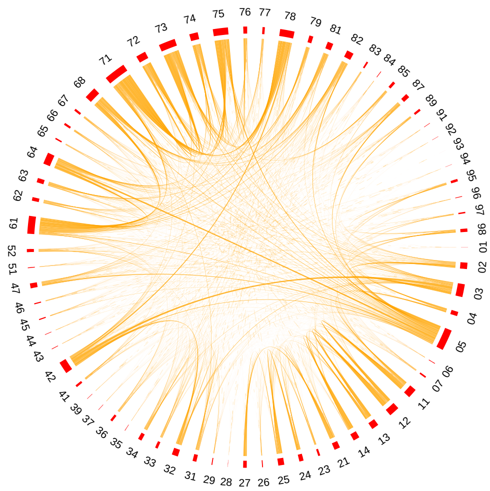


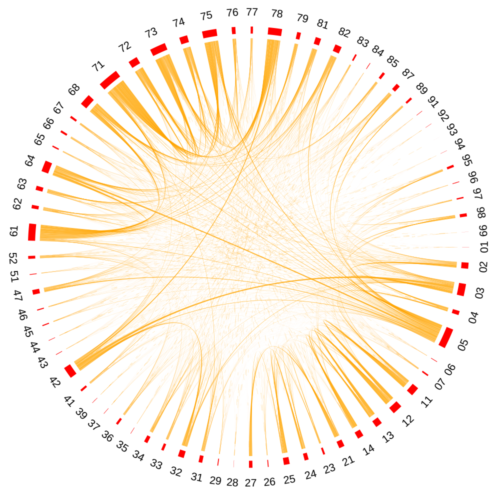
(a) Full data (b) Last 10 years



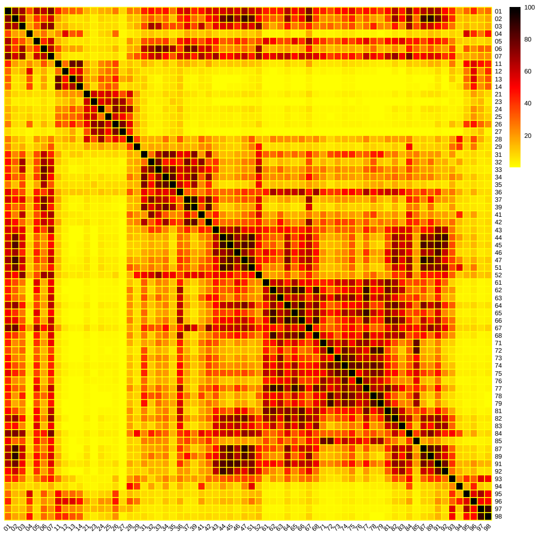
(c) Full data (d) Last 10 years

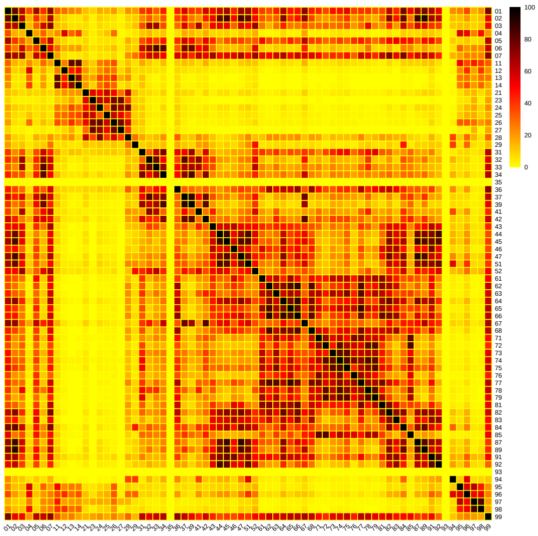
Figure B.4: Popularity of fields and sub-fields over time. We focus on a subset including articles published from 2000 to 2009 (last 10 years in our data) to compare the popularity of physics fields and sub-fields over time (i.e., number of articles assigned to a given field/sub-field). The distribution of topics remains fairly stable, except for the rise of interdisciplinary physics.





(a) Full Data **(b)** Last 10 Years

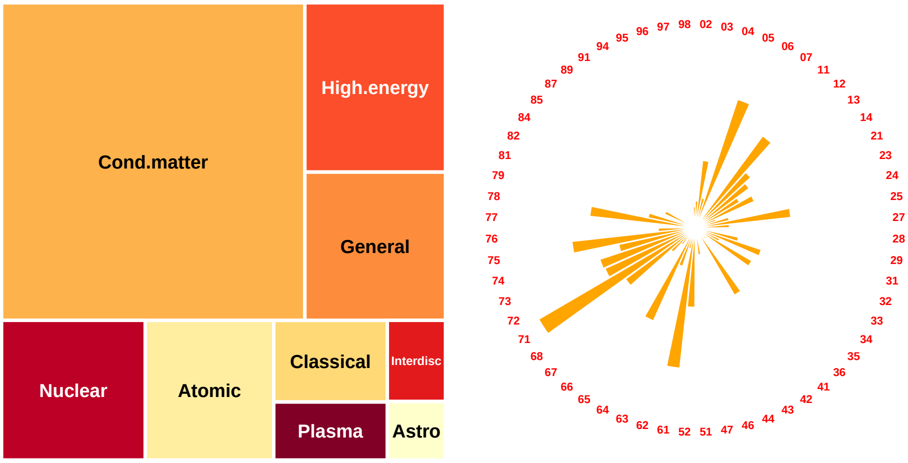




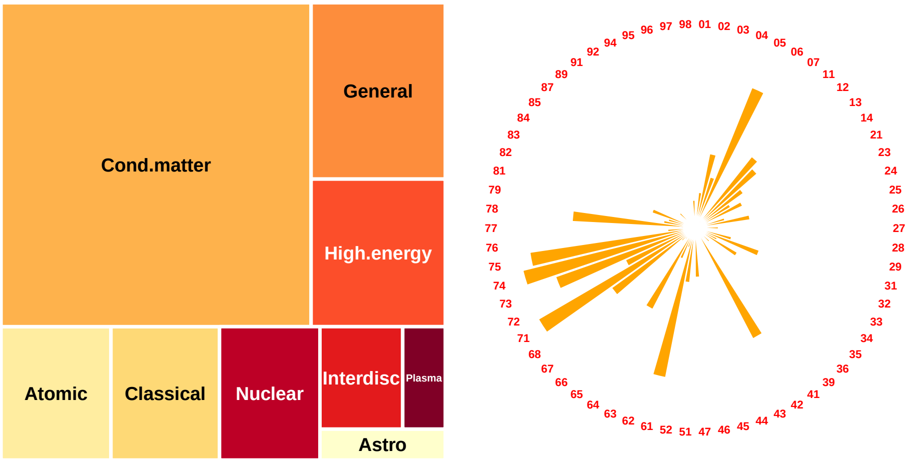
(c) Full Data **(d)** Last 10 Years

Figure B.5: Knowledge relatedness over time. We focus on a subset including articles published from 2000 to 2009 (last 10 years in our data) to evaluate the evolution of the physics knowledge space over time. Despite a slightly general increase of interdisciplinarity, subject proximity indicates a stable structure among sub-fields.

(a) 1980 - 1989



(b) 1990 - 1999



(c) 2000 - 2009

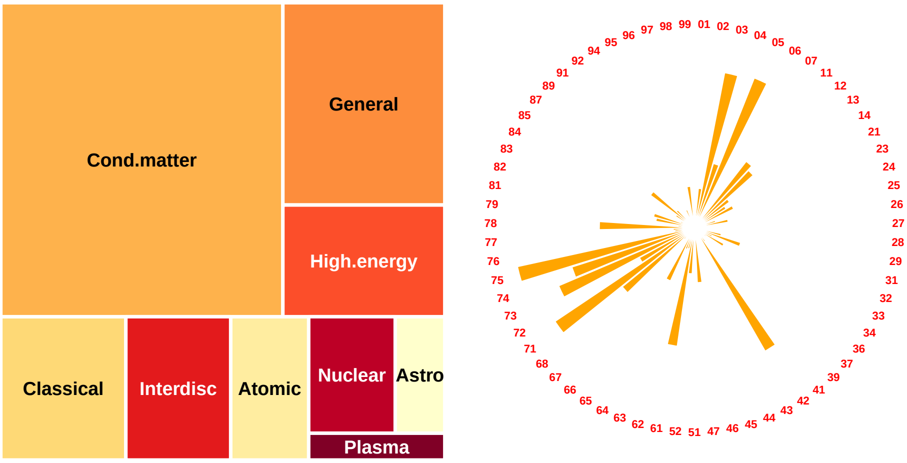
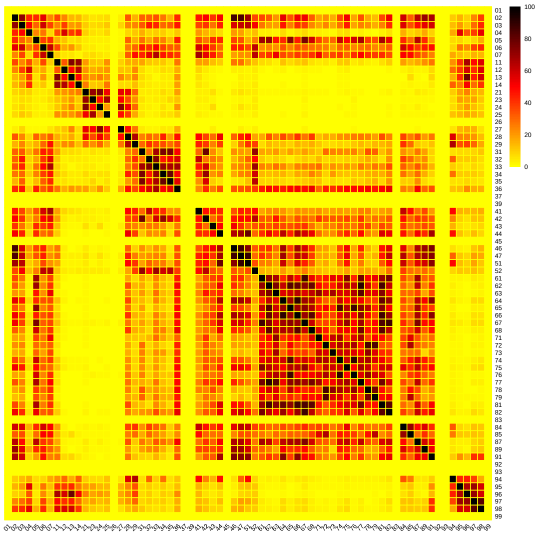
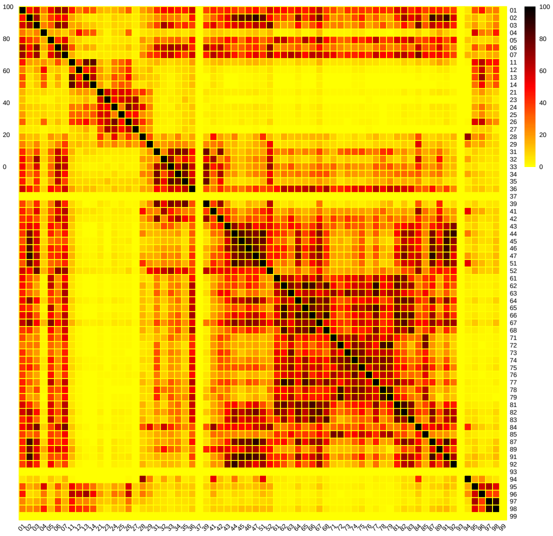


Figure B.6: Popularity of fields and sub-fields through decades. The plots compare the popularity of physics fields and sub-fields over time (i.e., number of articles assigned to a given field/sub-field).

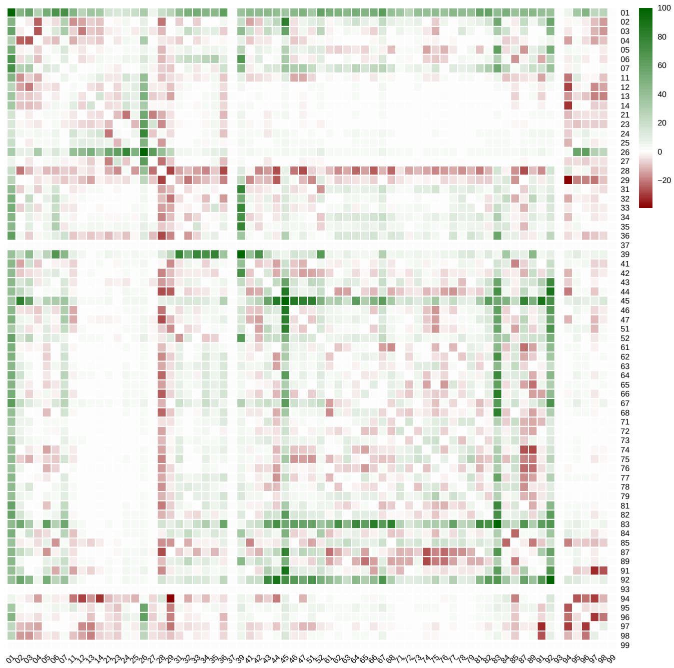
Since our measure of knowledge relatedness depends on PACS co-occurrences in research articles, we provide a more robust quantitative test to check whether the relationships among sub-fields have changed significantly over time. To do so, we first construct the difference between the cosine similarity matrix in two decades (see Figure B.7 and B.8). Then we validate the resulting difference matrices against the null of zero difference by sampling with replacement and generating 1,000 additional of such matrices. Finally, we compute the confidence interval ($\alpha = 0.05$) for each element of the difference matrix to assess its statistical significance, taking into account multiple hypothesis testing issues (Bonferroni correction). Figure B.9 shows the results of the bootstrap validation procedure (statistically significant pairs in black). In general, the number of significant element is not large, especially for consecutive decades, indicating a fairly stable structure of the physics knowledge space. More importantly, the analysis discussed in Section B.5, where past knowledge space is used in the regression, shows that changes in knowledge relatedness do not affect the main conclusions on the drivers of research portfolio diversification.



(a) 1980-1989

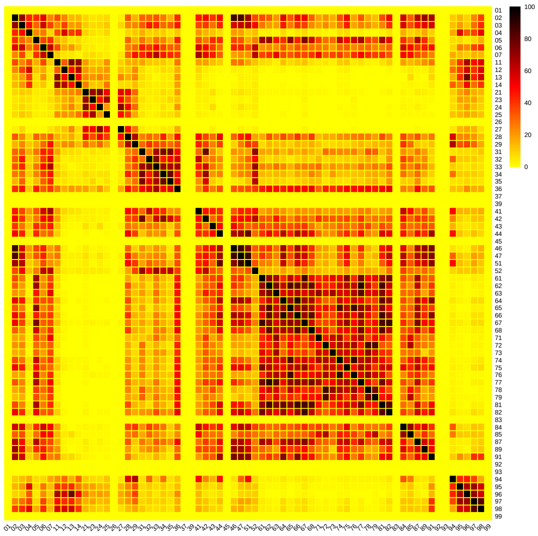


(b) 1990-1999

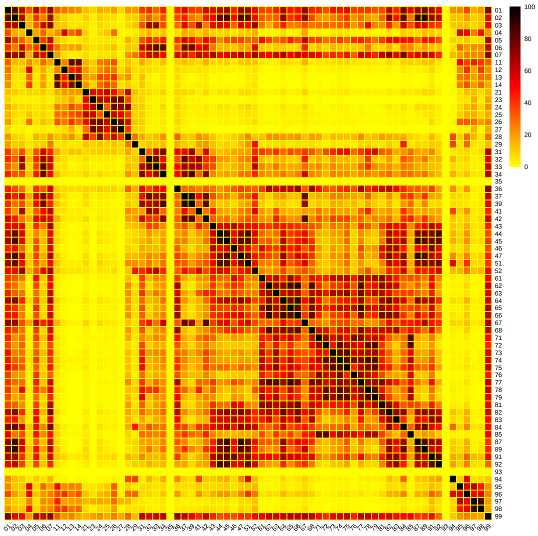


(c) Difference

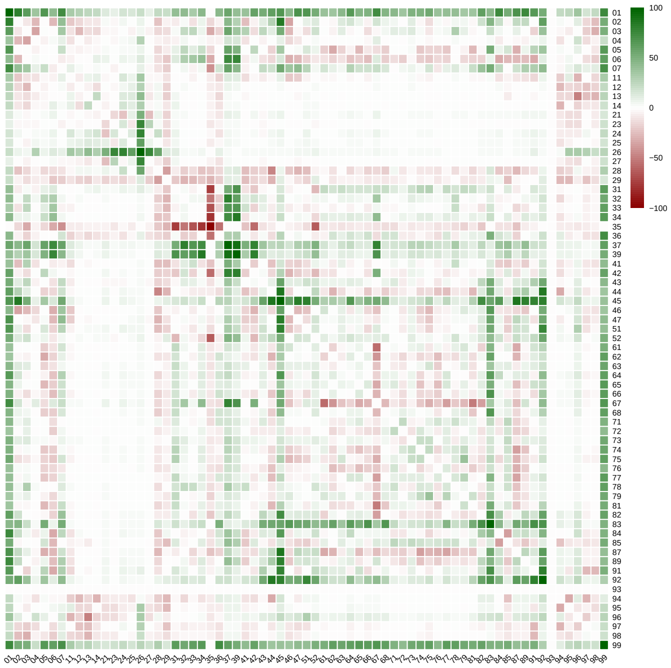
Figure B.7: Knowledge relatedness evolution over the first two decades. The top panels show the cosine similarity matrix between two-digit PACS in two decades, while the bottom panel shows their difference.



(a) 1980-1989

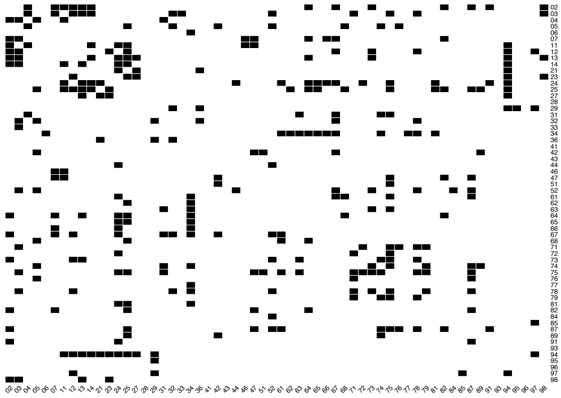


(b) 2000-2009



(c) Difference

Figure B.8: Knowledge relatedness evolution over three decades. The top panels show the cosine similarity matrix between two-digit PACS in two decades, while the bottom panel shows their difference.



(a) 90/99-80/89

(b) 00/09-90/99

(c) 00/09-80/89

Figure B.9: Bootstrap validation. Bootstrap validation of the difference matrices computed over decades by sampling with replacement and generating 1,000 additional difference matrices (showing only PACS codes present in each decade). The confidence interval ($\alpha = 0.05$) for each element of the matrix assesses the statistical significance (elements in black), taking into account multiple hypothesis testing correction (Bonferroni correction).

B.5 Alternative estimation strategies

Multidisciplinarity

Keeping track of diversification patterns for truly multidisciplinary scientists is a non-trivial task. Indeed, some scientists might have several core specializations leading to a positive bias in the previous estimates. To take into account this issue, we present an additional robustness check to validate further our empirical strategy: we assign each scientist to a single specialization - the one corresponding to the maximum value of RSA - but we constrain the choices of each scientists by eliminating from the regression the possibility to diversify in any of the PACS for which $RSA > 0$. In other words, we take into account only truly unexplored sub-fields. Figure B.10 confirms that scientists research portfolio diversification depends on social and knowledge relatedness, and the two measures interact with each other. Table B.10 summarizes the results.

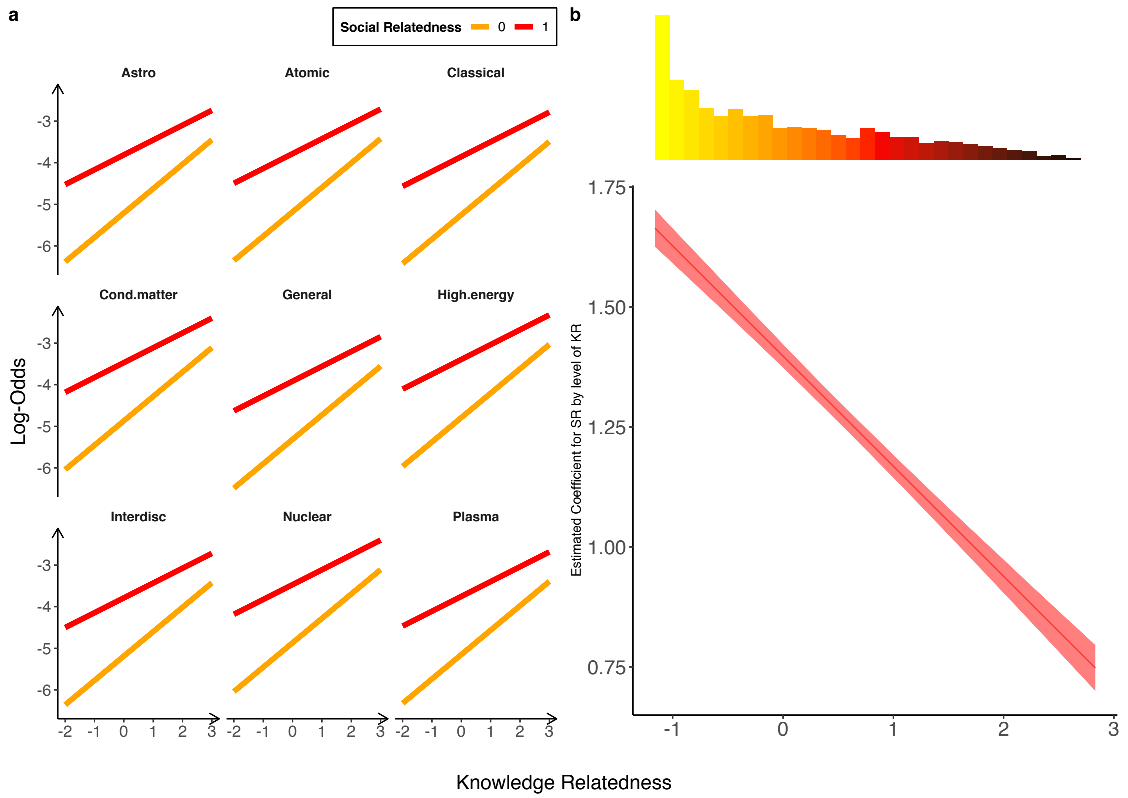


Figure B.10: Scientists' research portfolio diversification - constrained diversification. (a) Log-odds as a function of (binary) social relatedness and (standardized) knowledge relatedness, accounting for multiple control variables. (i). (b) Estimated coefficient for social relatedness conditional on knowledge relatedness, and distribution of knowledge relatedness. The analysis is performed considering only truly unexplored subfields (see text).

Time dimension

The temporal dimension is of paramount importance when evaluating scientific activities, especially to disentangle the direction of causality. Over time, our measures of knowledge and social relatedness might be affected by scientists' research diversification themselves. We tackle this issue by running an additional robustness check to isolate the effect of our measures on scientists' diversification strategies. First, we split our dataset into three time periods (i.e., three decades: 1980-1989, 1990-1990, 2000-2009) and we identify 15,466 scientists active in all periods. Then, we compute our measures of knowledge and social relatedness for each period to predict authors' diversification in a given decade using relatedness measures of a past decade. As before, we use a logistic regression where our dependent variable is a binary one (being

Table B.10: Constrained diversification.

	Dependent variable:		
	Y		
	Baseline	Interactions	Robust SE
	(1)	(2)	(3)
Knowledge Relatedness	0.507***	0.586***	0.586***
	(0.005)	(0.006)	(0.007)
Social Relatedness	1.268***	1.398***	1.398***
	(0.012)	(0.013)	(0.014)
field core-Atomic	0.036	0.029	0.029
	(0.025)	(0.025)	(0.023)
field core-Classical	−0.034	−0.043*	−0.043*
	(0.026)	(0.026)	(0.024)
field core-Cond.matter	0.341***	0.342***	0.342***
	(0.027)	(0.027)	(0.028)
field core-General	−0.092***	−0.105***	−0.105***
	(0.027)	(0.027)	(0.026)
field core-High.energy	0.426***	0.418***	0.418***
	(0.030)	(0.030)	(0.027)
field core-Interdisc	0.040	0.023	0.023
	(0.026)	(0.026)	(0.024)
Nuclear	0.326***	0.341***	0.341***
	(0.027)	(0.027)	(0.024)
field core-Plasma	0.063*	0.058	0.058*
	(0.036)	(0.036)	(0.032)
# of PACS	0.761***	0.753***	0.753***
	(0.007)	(0.007)	(0.006)
# of papers	0.374***	0.389***	0.389***
	(0.007)	(0.007)	(0.006)
PACS target popularity	1.478***	1.473***	1.473***
	(0.006)	(0.006)	(0.006)
Δ crowd	0.193***	0.192***	0.192***
	(0.006)	(0.006)	(0.006)
# of co-authors	−0.099***	−0.116***	−0.116***
	(0.006)	(0.006)	(0.006)
Δ PACS citations	−0.380***	−0.381***	−0.381***
	(0.007)	(0.007)	(0.006)
Δ field citations	0.309***	0.320***	0.320***
	(0.008)	(0.008)	(0.009)
KR:SR		−0.230***	−0.230***
		(0.009)	(0.010)
Constant	−5.179***	−5.209***	−5.209***
	(0.024)	(0.024)	(0.024)
Observations	1,503,010	1,503,010	1,503,010
Log Likelihood	−165,560.600	−165,263.900	−165,263.900
Akaike Inf. Crit.	331,157.100	330,565.900	330,565.900
Note:			*p<0.1; **p<0.05; ***p<0.01

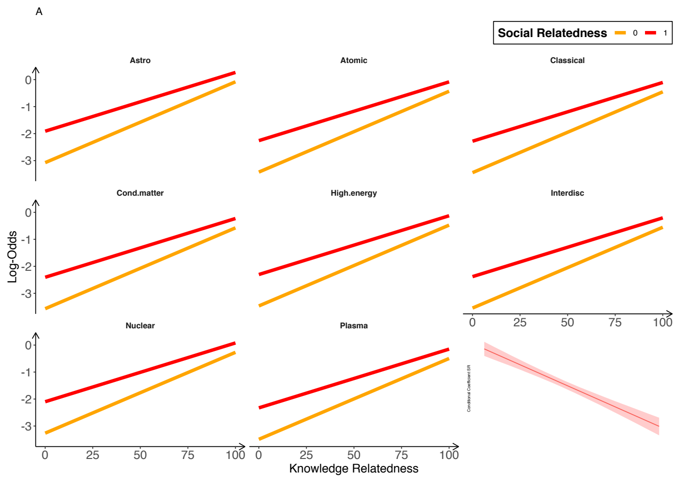
active in a sub-field different from specialization), but this time we use knowledge and social relatedness computed at time $t-1$ and $t-2$. Formally, we use three econometric specifications:

$$Y_{t-1} = \alpha + \beta KR_{t-2} + \gamma SR_{t-2} \ + \zeta(KR_{t-2} \times SR_{t-2}) + \delta field\ core \tag{B.1}$$

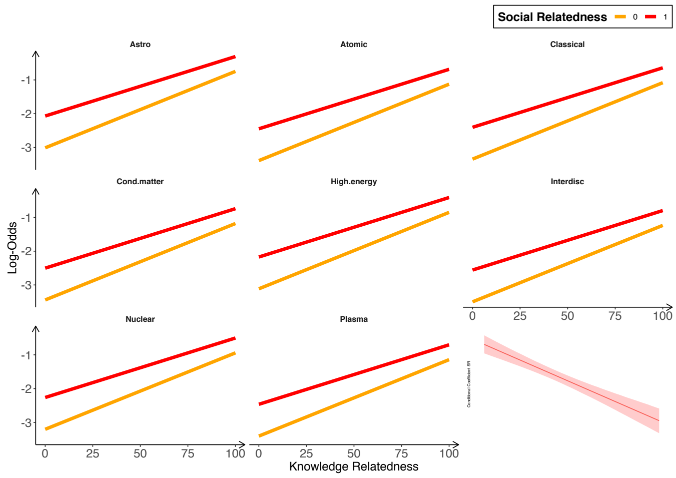
$$Y_{t} = \alpha + \beta KR_{t-1} + \gamma SR_{t-1} \ + \zeta(KR_{t-1} \times SR_{t-1}) + \delta field\ core \tag{B.2}$$

$$Y_{t} = \alpha + \beta KR_{t-2} + \gamma SR_{t-2} \ + \zeta(KR_{t-2} \times SR_{t-2}) + \delta field\ core \tag{B.3}$$

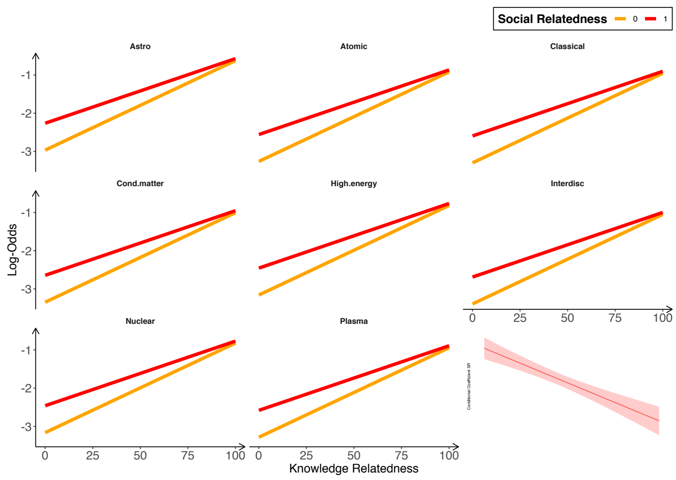
where t indicates the last decade (2000-2009). Such additional tests provide indication of the direction of causality since we take in account social and cognitive proximity prior to the scientists' choice to diversify. Moreover, we only consider sub-fields never explored before by each author so to approximate a quasi-experimental setting. Results confirm the role played by knowledge and social relatedness as well as the negative interaction between our two measures (see Figure B.11 and Table B.11).



a $Y_{t-1} = \alpha + \beta KR_{t-2} + \gamma SR_{t-2} \ + \zeta(KR_{t-2} \times SR_{t-2}) + \delta field\ core$



b $Y_{t} = \alpha + \beta KR_{t-1} + \gamma SR_{t-1} \ + \zeta(KR_{t-1} \times SR_{t-1}) + \delta field\ core$



c $Y_{t} = \alpha + \beta KR_{t-2} + \gamma SR_{t-2} \ + \zeta(KR_{t-2} \times SR_{t-2}) + \delta field\ core$

Figure B.11: Models with lagged variables. Log-odds as function of social and (standardized) knowledge relatedness and (bottom right panel), estimated coefficient for social relatedness conditional on (standardized) knowledge relatedness.

Table B.11: Diversification (lag)

	Dependent variable:		
	Y_{t-1}	Y_t	
	lag1	lag1	lag2
	(1)	(2)	(3)
KR_{t-2}	0.030***		0.023***
	(0.0003)		(0.0003)
SR_{t-2}	1.165***		0.703***
	(0.037)		(0.047)
KR_{t-1}		0.023***	
		(0.0003)	
SR_{t-1}		0.941***	
		(0.035)	
field core-Atomic	−0.350***	−0.378***	−0.296***
	(0.043)	(0.053)	(0.054)
field core-Classical	−0.371***	−0.334***	−0.333***
	(0.047)	(0.058)	(0.059)
field core-Cond.matter	−0.495***	−0.434***	−0.383***
	(0.039)	(0.049)	(0.049)
field core-High.energy	−0.393***	−0.104*	−0.194***
	(0.046)	(0.057)	(0.057)
field core-Interdisc	−0.471***	−0.490***	−0.428***
	(0.046)	(0.060)	(0.061)
field core-Nuclear	−0.188***	−0.194***	−0.198***
	(0.042)	(0.053)	(0.054)
field core-Plasma	−0.416***	−0.395***	−0.319***
	(0.050)	(0.062)	(0.063)
$KR_{t-2} : SR_{t-2}$	−0.008***		−0.007***
	(0.001)		(0.001)
$KR_{t-1} : SR_{t-1}$		−0.005***	
		(0.001)	
Constant	−3.075***	−3.010***	−2.964***
	(0.038)	(0.047)	(0.048)
Observations	766,519	618,352	618,352
Log Likelihood	−180,229.100	−142,158.800	−143,114.400
Akaike Inf. Crit.	360,480.300	284,339.500	286,250.900
Note:			*p<0.1; **p<0.05; ***p<0.01

Appendix C

Supplementary Information Chapter4

C.1 Data

This work relies on three primary sources of data: Web of Science, Reliance on Science, and Altmetric. The identification of all NETs research articles follows a standard search strategy used in the literature (Minx et al., 2017, 2018). Using WoS, we retrieve 3301 articles for the time window 1998-2017. We consider eight different NETs, using keywords and patterns in titles and abstracts. The articles that match NETs keywords but do not include such words or patterns in titles or abstracts are listed in a residual category (i.e., General). The **query NETs WoS** in this section and Table C.1 includes all the details and a brief summary description. Some articles might belong to more than one category. Accordingly, we count such articles in each potential category. For details, see the matrix of overlap in Figure C.1. Working on quantifying knowledge spillovers, we are interested in keeping track of citation flows (in several dimensions). Therefore, we only include articles that received at least one academic citation for the analysis and for which the DOI was retrievable. Around 62% of the original WoS sample is also included in Altmetric (see Table C.2 for a quick comparison, taking into account also the relative share of articles cited by patents or policy documents).

Query NETs WoS

(TS = (biochar* AND ((carbon OR CO2) NEAR/3 (sequest* OR storage OR stock OR accumulat* OR capture))) OR TS = (ocean NEAR/5 iron NEAR/5 (fertili*ation OR enrichment) NOT natural NOT ice* NOT glaci*) OR TS = ((soil NEAR/3 (carbon OR CO2) NEAR/3 (sequest* OR storage)) AND ("climate change" OR "global warm*") AND (manag* OR practice* OR restoration OR land-use)) OR TS = ((afforestation OR reforestation) AND ((carbon OR CO2) NEAR/3 (sequest* OR storage))) OR (TS = (("ocean liming") AND (removal OR storage) AND (CO2 OR carbon*)) OR TS = ((geoengineer*) AND (silicate OR olivine OR albite OR CACO3)) OR TS = ((silicate OR olivine OR albite OR CACO) AND (mitigat* NEAR/3 ("climate change" OR "global warming"))) OR TS = (("ocean alkalini*") AND (remov* OR storage OR mitigat* OR sequest*) AND (CO2 OR carbon*)) OR TS = (((enhance* OR artificial*) NEAR/2 weathering) AND ((carbon OR CO2 OR "climate change" OR "global warming") NEAR/3 (remov* OR sequest* OR storage OR sink OR mitigat* OR reduc*)))) NOT TS = (glaci* OR ice* OR ordovic* OR Aptian OR Cenozo* OR Paleo* OR Mezoso*) OR (TS = (((capture OR extraction OR absorbtion) NEAR/3 (air OR atmosph*)) AND (ambient OR "atmosph* pressure*") AND (CO2 OR carbon)) OR TS = (((captur* OR extract) NEAR/3 (direct* OR "carbon dioxide") NEAR/3 (air OR atmosph*)) AND (CO2 OR carbon)) OR TS = ((*sorbent OR amine) AND capture AND (carbon OR CO2) AND ("ambient air")) OR TS = ((captur* NEAR/3 CO2 NEAR/3 (air OR atmosph*)) AND solar)) NOT TS = (phenolic OR PCB* OR particulate OR NOx OR isotope OR "heat pump" OR polycyclic OR *bacteria* OR lignin OR sink OR pollution OR photosynth* OR biofuel* OR sugar) OR TS = (BECCS OR ((biomass OR bioenerg*) AND ("CCS" OR "Carbon capture and Storage" OR "Carbon dioxide capture and Storage" OR "CO2 capture and storage")) NOT "co-fir*" NOT "co-generat*" NOT cogeneration NOT coal) OR TS = ((seagrass OR mangrove* OR macroalgae OR "blue carbon") AND ((carbon OR CO2) NEAR/3 (sequest* OR accumulat* OR storage OR capture)) AND (defor-est* OR afforest* OR conserv* OR restor* OR manag*)) OR (TS = ((CDR AND

(CO2 OR carbon*)) OR ”negative carbon dioxide emission*” OR ”negative CO2 emission*” OR ”negative GHG emission*” OR ”negative greenhouse gas emission*” OR ”carbon-negative emission*” OR (”negative emission*” NEAR/10 carbon) OR (”negative emission*” NEAR/10 CO2)) OR TS = (geoengineering AND ((carbon OR CO2) NEAR/3 (sequest* OR accumulat* OR storage OR capture))) OR TS = ((”geoengineering” OR ”climate engineering”) AND CDR)) NOT TS = (N2O OR nitrogen OR NOX)) NOT TS = (”bioactive equivalent combinatorial components” OR ”bandwidth-efficient-channel-coding-scheme” OR ”bronchial epithelial cell cultures” OR ”california current system” OR comet OR mars OR exoplanet* OR ”competition chambers” OR gastric OR (mercury NEAR/3 capture) OR (image NEAR/3 capture) OR ”canary current system” OR ”heavy metal” OR eicosanoid OR ”companion cells” OR ”calcium carbonate sand” OR ”copper chaperone” OR ”commercial cane sugar” OR ”Cindoxin reductase” OR ”coupled dissolution reprecipitation” OR ”carbon dioxide reforming” OR rats OR ”complementarity determining regions” OR deoxycytidine)

Table C.1: Data description

NET	Code	Description	N	%
Afforestation/Reforestantion	AR	Forest management and restoration programs increase the CO2 captured from the atmosphere and stored in living biomass.	677 (603)	21 (21)
Bio-energy with Cabon Capture and Storage	BECCS	Biomass is grown and used to power as a source of thermal energy. The CO2 produced is captured and stored in geological reservoirs.	247 (206)	7 (7)
Biochar	Biochar	The pyrolysis of biomass produce charcoal (i.e., biochar). It can be used as soil additive, with positive effect in terms of carbon capture and stored in soil.	555 (478)	17 (17)
Blue Carbon	BC	Blue carbon refers to carbon captured by the world's ocean and coastal ecosystems, such as sea grasses, mangroves, or salt marshes.	121 (96)	4 (3)
Direct Air Capture	DAC	CO2 is absorbed directly from the atmosphere through chemicals and stored.	245 (214)	7 (8)
Enhanced Weathering	EW	Minerals that can absorb CO2 are grinded and spread in lands or oceans.	71 (63)	2 (2)
Ocean Fertilization	OF	Nutrients (such as iron) can stimulate the growth of phytoplankton. Consequently, the absorbed CO2 is naturally sequestered in the ocean.	113 (102)	3 (4)
Soil Carbon Sequestration	SCS	More efficient agricultural practices enhance soils carbon absorption potential.	410 (354)	12 (12)
General	General	NETs scientific articles with no specific keywords in titles or abstract.	1059 (906)	
Total (with citation info WoS)		The total only includes unique articles (they might belong to more than one category).	3301 (2850)	

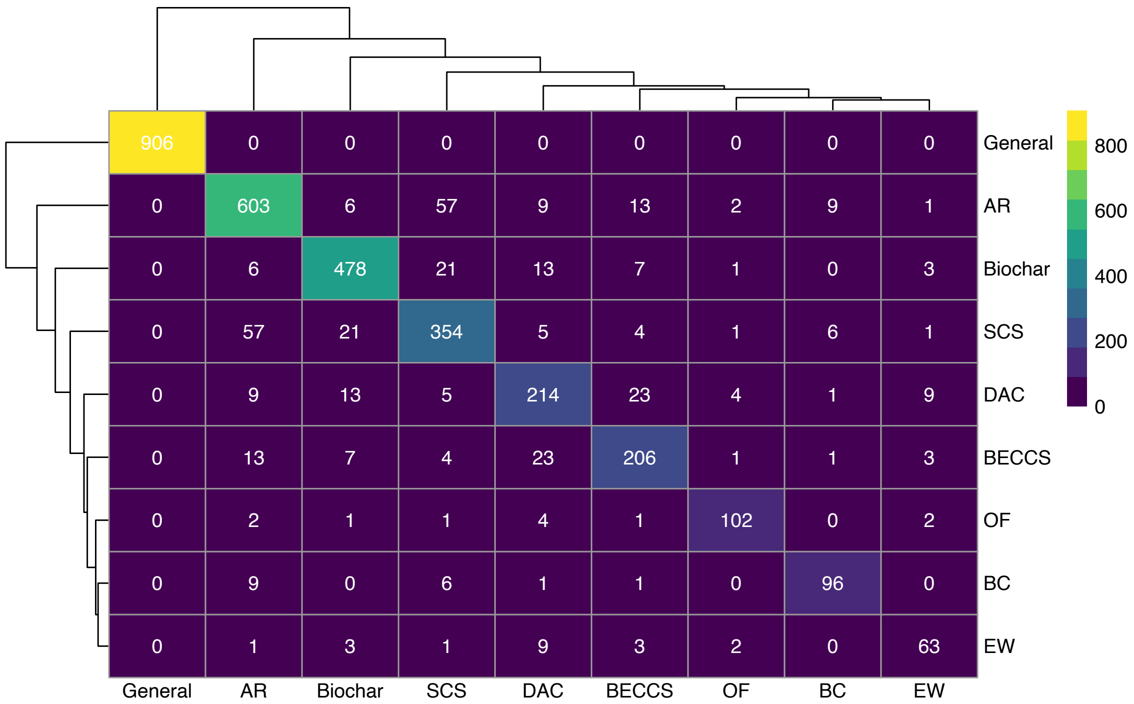


Figure C.1: NETs overlap in articles. Each entry of the matrix shows the number of articles including more than one NET category. Along the diagonal the total sum for every NET.

.

Table C.2: Share of NETs articles linked to technology and policy – Altmetric vs. (WoS–RoS)

NET	N	% with policy citation	% with patent citation
General	556 (906)	39	3 (3)
AR	354 (603)	43	0.2 (1)
BECCS	135 (206)	49	5 (3)
Biochar	293 (478)	20	8 (9)
BC	78 (96)	44	0 (0)
DAC	143 (214)	32	27 (23)
EW	54 (63)	44	11 (9)
OF	78 (102)	42	3 (8)
SCS	247 (354)	47	0.8 (2)

The total number of articles retrieved via Altmetric is 2040. The number of articles for which we have both WoS and Altmentic data, coupled with citations data and DOI information in both dataset is 1800. Articles might belong to more than one NET category as shown in C.1. The information collected through Reliance in Science (RoS) allow us to compare the coverage in terms of patent citations for a larger sub-sample.

C.2 Matching

Our empirical analysis relies on quantitative comparisons. Therefore, we rely on a matching strategy to avoid reaching misleading interpretations based on potentially biased estimates. The first step of our matching scheme is to construct a control group by collecting – through WoS – up to 10 articles (with replacement) published in the same year and the same journal. Consequently, we obtain up to 10 twin articles for each NETs paper of interest. As far as the regression analyses are concerned, we further enhance our matching strategy. In practice, we generate 30 one-to-one matched sub-samples (without replacement) to control our estimates' stability. Figure C.2 depicts the steps of our matching scheme. Furthermore, we use the same strategy to construct a second set of control groups specifically linked to the climate-related literature. We retrieve climate-specific articles following the **Query climate control WoS** (listed below and already validated in the literature (Grieneisen and Zhang, 2011)). Finally, Figure C.3 shows a treemap of the most popular venues that publish NETs articles. We list venues that appear at least 10 times in our sample.

Query climate control WoS

SO=(Climate Alert OR Climate Dynamics OR Climate Policy OR Climatic Change OR Global and Planetary Change OR Global Change Biology OR International Journal of Greenhouse Gas Control OR Mitigation and Adaptation Strategies for Global Change) OR TS=(((CO2 OR "carbon dioxide" OR methane OR CH4 OR "carbon cycle" OR "carbon cycles" OR "carbon cycling" OR "carbon budget*" OR "carbon flux*" OR "carbon mitigation") AND (climat*)) OR (("carbon cycle" OR "carbon cycles" OR "carbon cycling" OR "carbon budget*" OR "carbon flux*" OR "carbon mitigation") AND (atmospher*))) OR TS=("carbon emission*" OR "sequestration of carbon" OR "sequester* carbon" OR "sequestration of CO2" OR "sequester* CO2" OR "carbon tax*" OR "CO2 abatement" OR "CO2 capture" OR "CO2 storage" OR "CO2 sequester*" OR "CO2 sequestration" OR "CO2 sink*" OR "anthropogenic carbon" OR "captur* of carbon dioxide" OR "captur* of CO2" OR "climat* variability" OR "climat* dynamic*" OR "chang* in climat*" OR "climat* proxies" OR "climat* proxy" OR "climat* sensitivity" OR "climat* shift*" OR "coupled ocean-climat*" OR "early climat*" OR "future climat*" OR "past climat*" OR "shift* climat*" OR "shift in climat*") OR TS=("atmospheric carbon dioxide" OR "atmospheric CH4" OR "atmospheric CO2" OR "atmospheric methane" OR "atmospheric N2O" OR "atmospheric nitrous oxide" OR "carbon dioxide emission*" OR "carbon sink*" OR "CH4 emission*" OR "climat* policies" OR "climat* policy" OR "CO2 emission*" OR dendroclimatolog* OR ("emission* of carbon dioxide" NOT nanotube*) OR "emission* of CH4" OR "emission* of CO2" OR "emission* of methane" OR "emission* of N2O" OR "emission* of nitrous oxide" OR "historical climat*" OR IPCC OR "methane emission*" OR "N2O emission*" OR "nitrous oxide emission*") OR TS=("climat* change*" OR "global warming" OR "greenhouse effect" OR "greenhouse gas*" OR "Kyoto Protocol" OR "warming climat*" OR "cap and trade" OR "carbon capture" OR "carbon footprint*" OR "carbon neutral" OR "carbon offset" OR "carbon sequestration" OR "carbon storage" OR "carbon trad*" OR "changing climat*" OR "climat* warming")

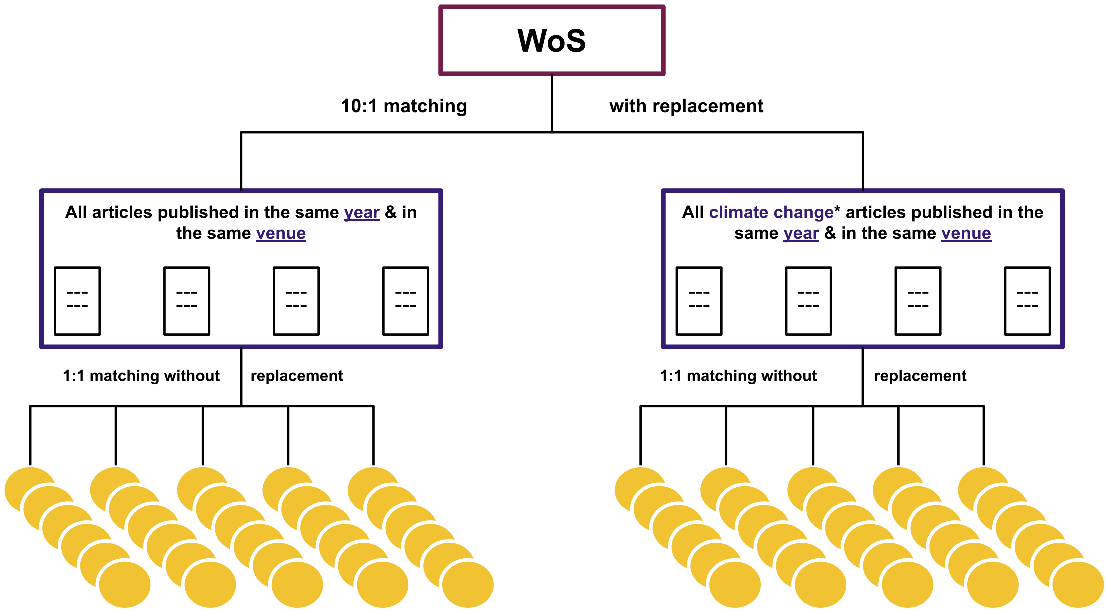


Figure C.2: Matching scheme. A schematic representation of the matching procedure used in the empirical analysis. The right end side refer to the construction of the baseline control group; while the left end side refers to the climate control.

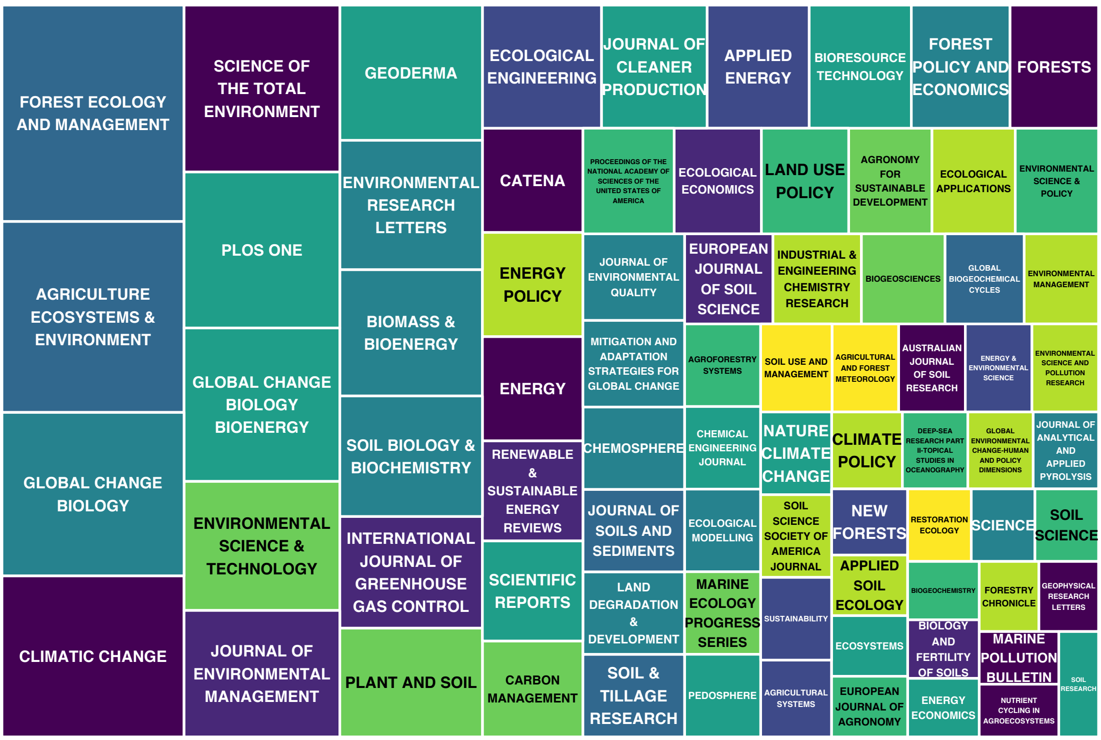


Figure C.3: Top NETs Venues. Treemap listing the most representative venues for NETs articles. The map include academic journals that published at least 10 paper related to NETs.

C.3 Policy sentiment analysis

Citations coming from policy documents are collected through Altmetric (see section 4.7.1. While academic and patent citations have historically been used to keep track of knowledge flows Fortunato et al. (2018a); Jaffe and De Rassenfosse (2019), there is little evidence that policy citations capture positive mentions for scientific results.

To partly tackle this issue, we explore the sentiment of a subset of policy documents that cite our focal articles. First, we select 208 (English) documents and then analyze their entire text using NLP methods. Although working on the entire text is subject to potential measurement errors, the overall sentiment ratio of each document gives us a first indication of the orientation. We define the sentiment ratio by simply counting the number of positive words over the total number of words (see Figure C.4).

To derive our measure, we use a dictionary-based approach, that is, a list of general-purpose lexicons collected for text analysis studies and freely available through the R package *tidytext* (Hu and Liu, 2004).

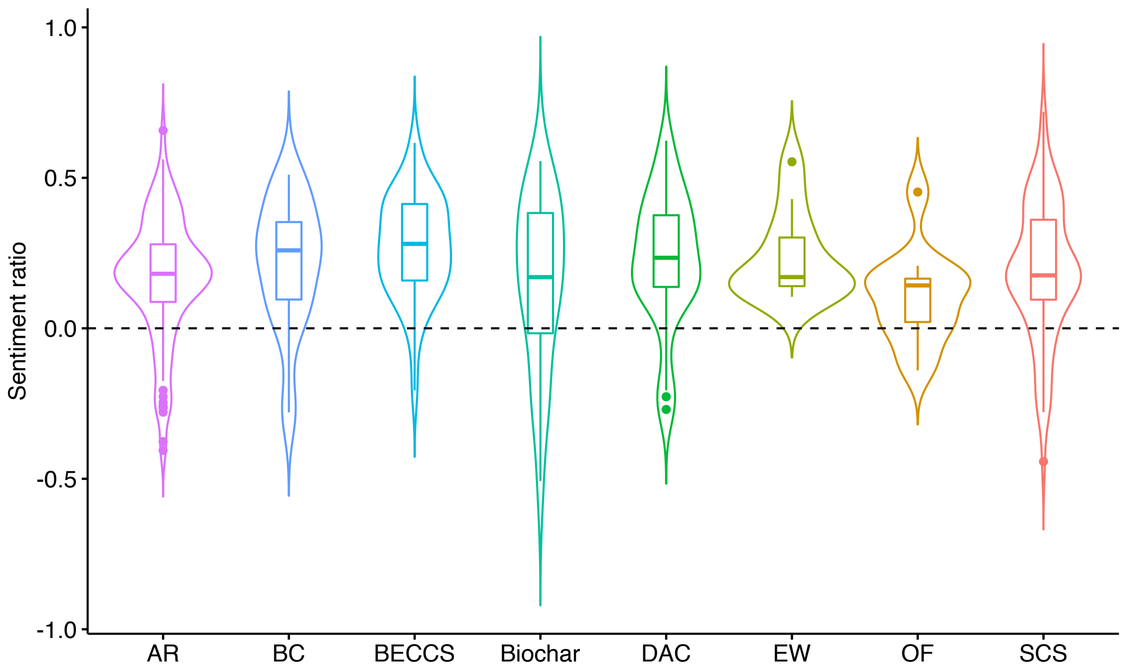


Figure C.4: Sentiment of policy documents. Violin plots showing the sentiment ratio of policy documents citing scientific articles related to NETs.

C.4 Knowledge flows

As mentioned in Section 4.3, we collect all backward and forward citations (through WoS) for all the NETs articles in our sample. The purpose is to identify the knowledge base and the potential direction of scientific spillovers. The flow diagrams depicted in Figure C.5 highlight the differences between nature-based and technology-based negative emissions options. We consider the 10 largest sub-fields to clarify the scientific linkages concerning different NETs better. The 10 most important fields account for more than 75% of the total citations. Figure C.6 shows the exact distribution for all our NETs options. To summarize the main trends: forestry, ecology, and soil science dominate in the nature-based NETs, while advances in chemistry and chemical engineering shape DAC and BECCS developments. EW and Biochar can be placed between these two groups, sharing some nature-based knowledge components and technical features. Not surprisingly, OF and BC disproportionally link to oceanography and marine biology.

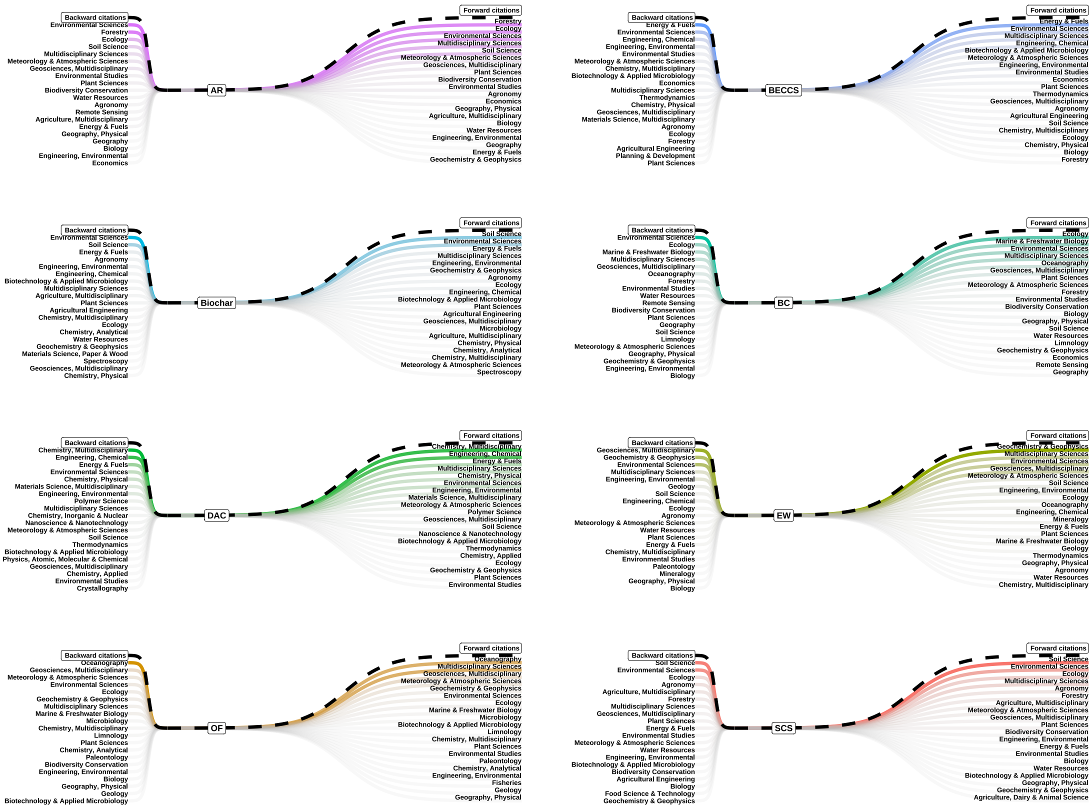


Figure C.5: Knowledge flows. Top 10 WoS subjects that affect (backward citations) and are affected (forward citations) by NETs research. **(AR)** Afforestation and Reforestation. **(BECCS)** Bio-energy with Carbon Capture ans Storage. **(Biochar)** Biochar. **(BC)** Blue Carbon. **(DAC)** Direct Air Capture. **(EW)** Enhanced Weathering. **(OF)** Ocean Fertilization. **(SCS)** Soil Carbon Sequestration.

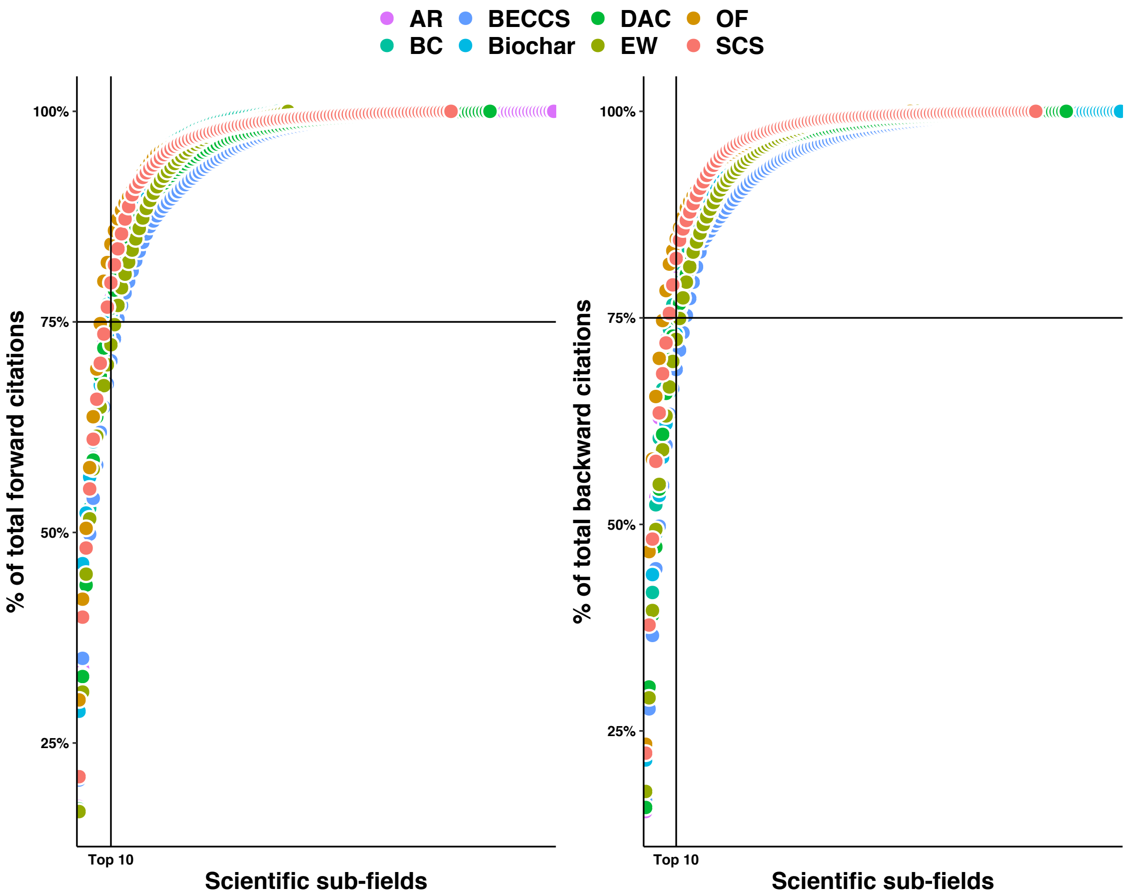


Figure C.6: Pareto plot. Cumulative percentage of the total number of forward and backward scientific citations by NETs. The horizontal reference line marks the 75% of total citations. The vertical reference line indicates the 10 largest scientific sub-fields.

C.5 Radar - climate control

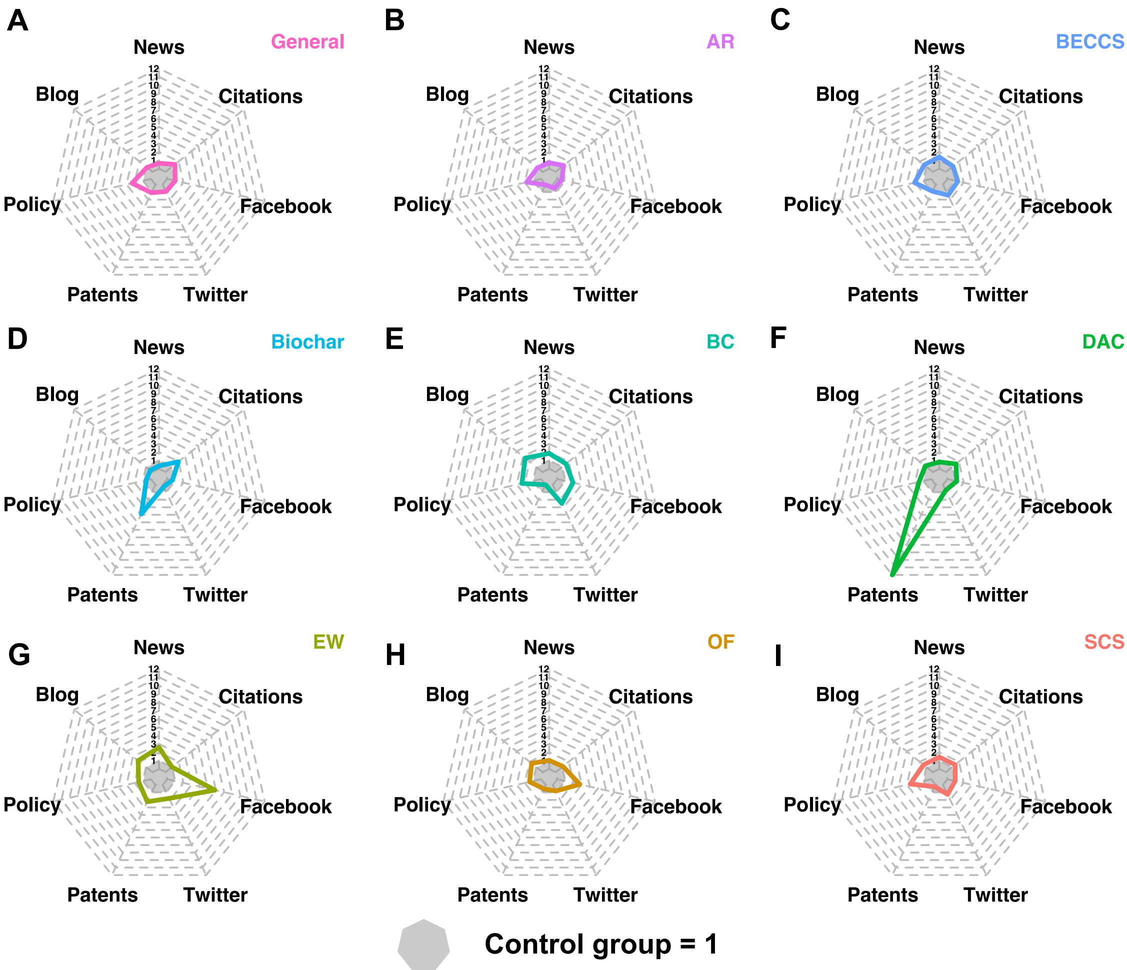


Figure C.7: NETs multidimensional impact – Climate control (A–I) Radar charts for each NET, showing multidimensional spillovers (*climate* control group fixed at 1). **(A)** General. **(B)** Afforestation and reforestation – AR. **(C)** Bio–energy with Carbon Capture and Storage – BECCS. **(D)** Biochar. **(E)** Blue Carbon – BC. **(F)** Direct Air Capture – DAC. **(G)** Enhanced weathering – EW. **(H)** Ocean fertilization – OF. **(I)** Ocean fertilization – OF.

C.6 Regressions - Robustness

We perform a long series of robustness checks to validate our results. First, as mentioned in Section 4.7.2, we estimate our baseline models 30 times, with varying control groups. The boxes depicted in Figure 4.4 show the average point estimate β_k^* and the average confidence intervals $\langle C.I. \rangle$ across the 30 runs of our statistical model (see Table C.3 for more details). Second, we run the analysis using different control groups, focusing, for instance, on the climate change literature (Section C.2 covers details on the construction of the climate control). Figure C.4 and Table C.4 summarize the result of our analysis with the climate control groups as reference.

In addition, we repeat our analysis using a linear model instead of GLMs. Formally, we used the following specification:

$$log(S_{ikt}) = \alpha + \sum_k \beta_k NET_{ik} + \sum_t \gamma_t T_{it} + \epsilon \tag{C.1}$$

where S_{ikt} is the number of forward citations in the science, technology, or policy dimensions, NET_{ik} refers to the corresponding NET and T_{it} represent a year dummy, as in Section 4.4. Results are summarized in Figure C.9 and Table C.5 with the baseline control, and in Figure C.15 and Table C.6 with the climate control.

To further evaluate the consistency of our results and control for potential differences in coverage between WoS and Altmetric, we repeat the analysis, comparing the quantitative trends highlighted so far in terms of scientific and technological spillovers. Following the empirical strategy of Section 4.4, we first use WoS – instead of Altmetric – to quantify scientific spillovers, namely: citations and scope (i.e., # of different fields that cite a given article). Figure C.11,C.12 and Table C.7, C.12 summarize the results concerning both the baseline and the climate control groups. Then, we also use RoS to keep track of the science-technology links. We repeat the analysis using logistic regressions as in Section 4.4. Figure C.13 and Table C.9 confirm the overall distance of NETs from the technological frontier, and the relative advantage of DAC. We also perform an additional robustness check by estimating the models of Section 4.4 including two potentially relevant control variables: a field (or combinations of fields) indicator extracted via Altmetric and whether the article is open access. A categorical field variable (F_{if}) allows us to control for disciplinary differences in citation patterns within and beyond science. A dummy that captures whether articles are open access (OA_i) controls for the possibility of broader/more accessible diffusion of knowledge.

More formally, we employ the following specification:

$$g(E(S_{ikt}|\dots)) = \alpha + \sum_k \beta_k NET_{ik} + \sum_t \gamma_t T_{it} + \sum_f \delta_f F_{if} + \mu OA_i \tag{C.2}$$

Figure C.14,C.15 and Table C.10,C.11 include all details. Some coefficients shirk vis-a-vis our baseline model of Section 4.4, as the field control is sufficiently strong to clean out the disciplinary heterogeneity that distinguishes, for instance, engineering-based articles from marine biology or generally less cited sub-fields.

Lastly, we finally check the robustness of our results by running individual regressions for some NETs with NET-specific control groups. In detail, Figure C.16 show the estimates for AR, BECCS, DAC, and SCS. The outcomes confirm that only DAC has a significant association with technological developments. Overall, the many specifications we have explored corroborate our main results, with coefficients' values ranging across specific specifications and control groups.

Table C.3: **Coefficients and C.I. Figure 4.4**

NET	Science exp β	Science $\langle C.I. \rangle$	Technology exp β	Technology $\langle C.I. \rangle$	Policy exp β	Policy $\langle C.I. \rangle$
General	1.68	[1.53,1.86]	0.70	[0.45,1.1]	2.85	[2.26,3.58]
AR	1.25	[1.11,1.41]	0.04	[0.01,0.18]	2.66	[2.03,3.47]
BECCS	1.88	[1.58,2.27]	1.29	[0.6,2.66]	5.65	[3.8,8.41]
Biochar	2.61	[2.3,2.97]	2.29	[1.48,3.53]	1.31	[0.93,1.82]
BC	1.88	[1.5,2.39]	✗		4.36	[2.62,7.18]
DAC	1.86	[1.57,2.23]	7.48	[5,11.34]	2.06	[1.36,3.09]
EW	1.39	[1.05,1.88]	2.47	[1.03,5.63]	4.70	[2.49,8.87]
OF	0.92	[0.72,1.21]	0.39	[0.09,1.32]	2.31	[1.32,4.01]
SCS	1.63	[1.43,1.88]	0.14	[0.03,0.44]	3.75	[2.76,5.1]
Year dummies	✓		✓		✓	
Matched samples	30		30		30	
# of obs.	3392		3392		3392	

Table C.4: **Coefficients and C.I. Figure C.8**

NET	Science exp β	Science $\langle C.I. \rangle$	Technology exp β	Technology $\langle C.I. \rangle$	Policy exp β	Policy $\langle C.I. \rangle$
General	1.40	[1.28,1.55]	0.99	[0.57,1.64]	1.93	[1.56,2.38]
AR	0.96	[0.85,1.08]	0.06	[0,0.3]	1.75	[1.36,2.25]
BECCS	1.55	[1.3,1.86]	1.43	[0.52,3.34]	3.74	[2.58,5.42]
Biochar	2.18	[1.93,2.47]	3.26	[1.97,5.23]	0.88	[0.64,1.19]
BC	1.78	[1.42,2.26]	✗		3.19	[1.98,5.13]
DAC	1.82	[1.53,2.18]	12.34	[7.61,19.68]	1.47	[0.98,2.16]
EW	1.19	[0.92,1.57]	4.02	[1.5,9.3]	2.50	[1.41,4.39]
OF	0.70	[0.55,0.9]	1.03	[0.26,2.92]	1.34	[0.79,2.25]
SCS	1.37	[1.2,1.58]	0.20	[0.04,0.66]	2.49	[1.87,3.31]
Year dummies	✓		✓		✓	
Matched samples	30		30		30	
# of obs.	3716		3716		3716	

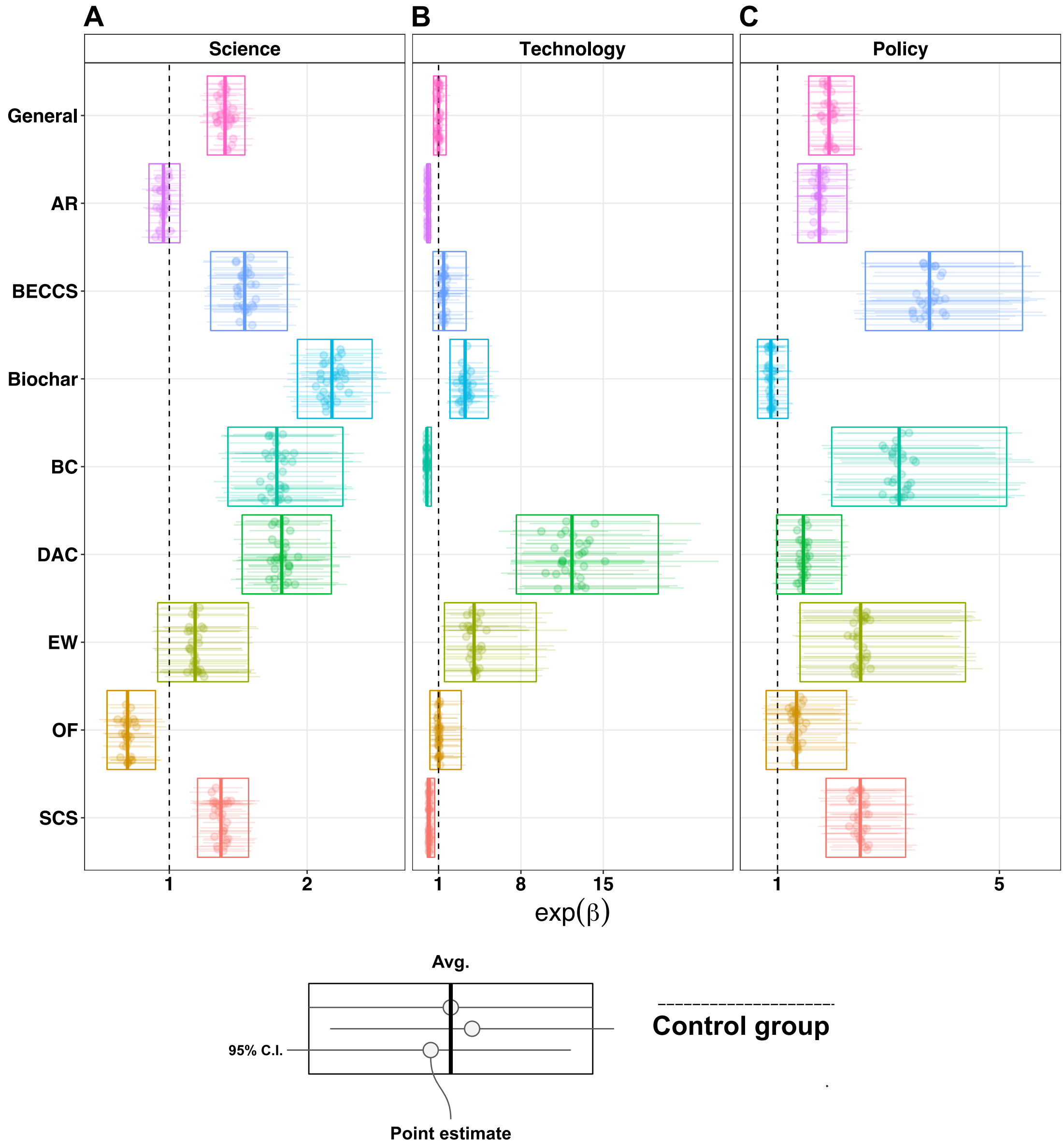


Figure C.8: NETs spillovers to science, technology and policy (climate). Coefficients of the regression models of Eq. (3). Results are obtained by fitting 30 negative binomial regressions (**A**) and 30 logistic regressions (**B–C**) on one-to-one matched samples with year dummies. **(A)** Estimated coefficients (exponentiated) for each NET on the number of scientific citations. **(B)** Estimated coefficients (exponentiated) for each NET on the probability of being cited by a patent (BC estimates set to zero since there is no patent documents citing BC papers). **(C)** Estimated coefficients (exponentiated) for each NET on the probability of being cited by a policy document.

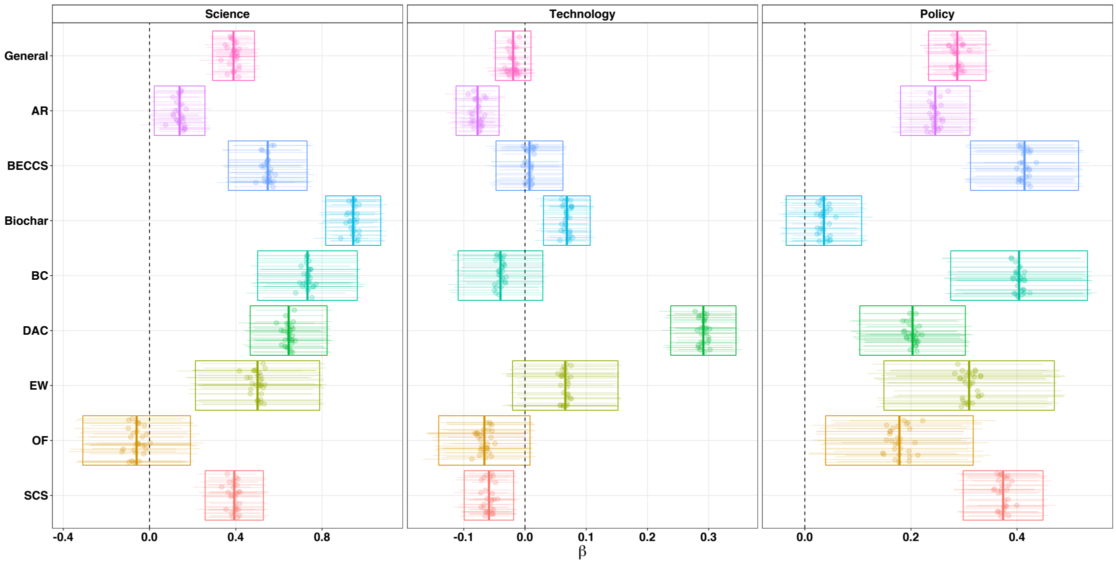


Figure C.9: NETs spillovers to science, technology and policy - OLS

Table C.5: Coefficients and C.I. Figure C.9

	Science		Technology		Policy	
NET	β	$\langle C.I. \rangle$	β	$\langle C.I. \rangle$	β	$\langle C.I. \rangle$
General	0.39	[0.29,0.49]	-0.02	[-0.05,0.01]	0.28	[0.22,0.33]
AR	0.12	[0,0.24]	-0.08	[-0.12,-0.05]	0.24	[0.18,0.31]
BECCS	0.54	[0.35,0.72]	0.01	[-0.05,0.06]	0.42	[0.32,0.52]
Biochar	0.92	[0.79,1.05]	0.06	[0.02,0.1]	0.05	[-0.02,0.12]
BC	0.62	[0.39,0.86]	-0.04	[-0.11,0.03]	0.34	[0.21,0.47]
DAC	0.63	[0.46,0.81]	0.29	[0.24,0.35]	0.21	[0.11,0.3]
EW	0.42	[0.13,0.72]	0.07	[-0.02,0.16]	0.33	[0.17,0.49]
OF	-0.07	[-0.33,0.19]	-0.07	[-0.15,0.01]	0.17	[0.03,0.31]
SCS	0.37	[0.24,0.51]	-0.06	[-0.1,-0.02]	0.34	[0.26,0.42]
Year dummies	✓		✓		✓	
Matched samples	30		30		30	
# of obs.	3392		3392		3392	

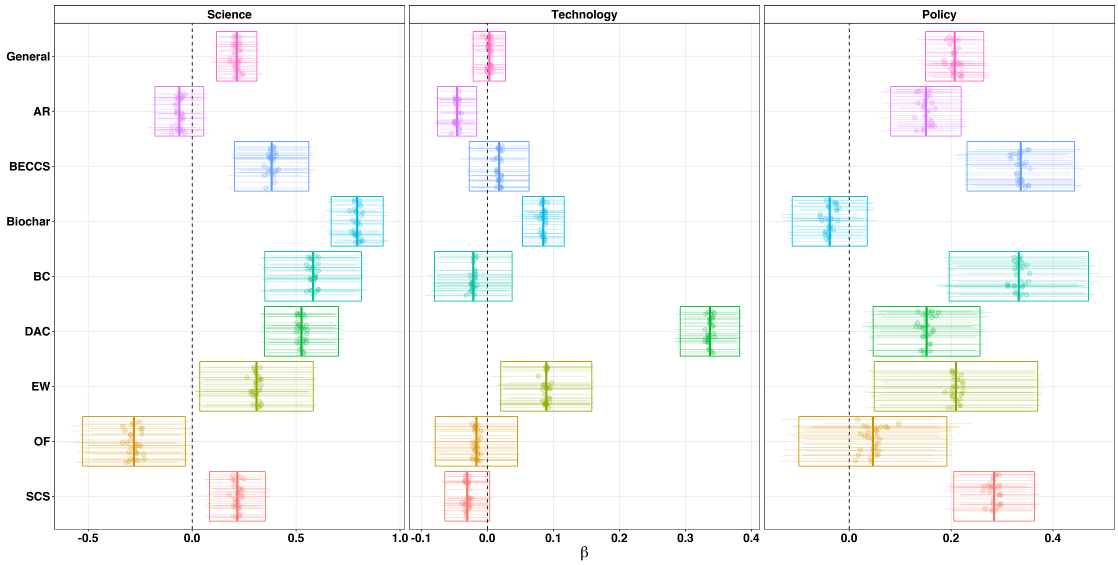


Figure C.10: NETs spillovers to science, technology and policy - OLS (climate)

Table C.6: Coefficients and C.I. Figure C.10

	Science		Technology		Policy	
NET	β	$\langle C.I. \rangle$	β	$\langle C.I. \rangle$	β	$\langle C.I. \rangle$
General	0.21	[0.12,0.31]	0.00	[-0.02,0.03]	0.21	[0.15,0.26]
AR	-0.06	[-0.18,0.06]	-0.05	[-0.08,-0.02]	0.15	[0.08,0.22]
BECCS	0.38	[0.2,0.56]	0.02	[-0.03,0.06]	0.34	[0.23,0.44]
Biochar	0.79	[0.67,0.92]	0.08	[0.05,0.12]	-0.04	[-0.11,0.04]
BC	0.58	[0.35,0.81]	-0.02	[-0.08,0.04]	0.33	[0.2,0.47]
DAC	0.53	[0.35,0.7]	0.34	[0.29,0.38]	0.15	[0.05,0.26]
EW	0.31	[0.04,0.58]	0.09	[0.02,0.16]	0.21	[0.05,0.37]
OF	-0.28	[-0.53,-0.03]	-0.02	[-0.08,0.05]	0.05	[-0.1,0.19]
SCS	0.22	[0.08,0.35]	-0.03	[-0.06,0]	0.28	[0.21,0.36]
Year dummies	✓		✓		✓	
Matched samples	30		30		30	
# of obs.	3716		3716		3716	

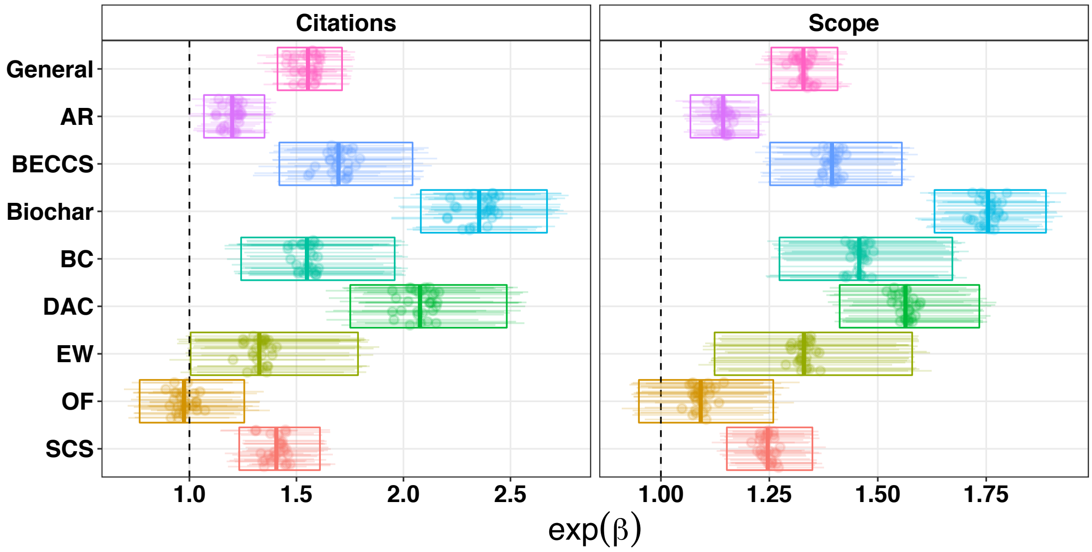


Figure C.11: NETs spillovers to science - WoS citations

Table C.7: Coefficients and C.I. Figure C.11

	Citations		Scope	
term	$\exp\beta$	$\langle C.I.\rangle$	$\exp\beta$	$\langle C.I.\rangle$
General	1.55	[1.4,1.71]	1.32	[1.25,1.4]
AR	1.22	[1.09,1.38]	1.13	[1.06,1.22]
BECCS	1.75	[1.46,2.11]	1.39	[1.25,1.55]
Biochar	2.37	[2.1,2.69]	1.75	[1.63,1.89]
BC	1.36	[1.09,1.73]	1.37	[1.19,1.58]
DAC	2.12	[1.79,2.53]	1.56	[1.41,1.72]
EW	1.26	[0.95,1.71]	1.30	[1.09,1.55]
OF	0.99	[0.78,1.29]	1.08	[0.94,1.25]
SCS	1.40	[1.22,1.61]	1.23	[1.14,1.34]
Year dummies	✓		✓	
Matched samples	30		30	
# of obs.	3392		3392	

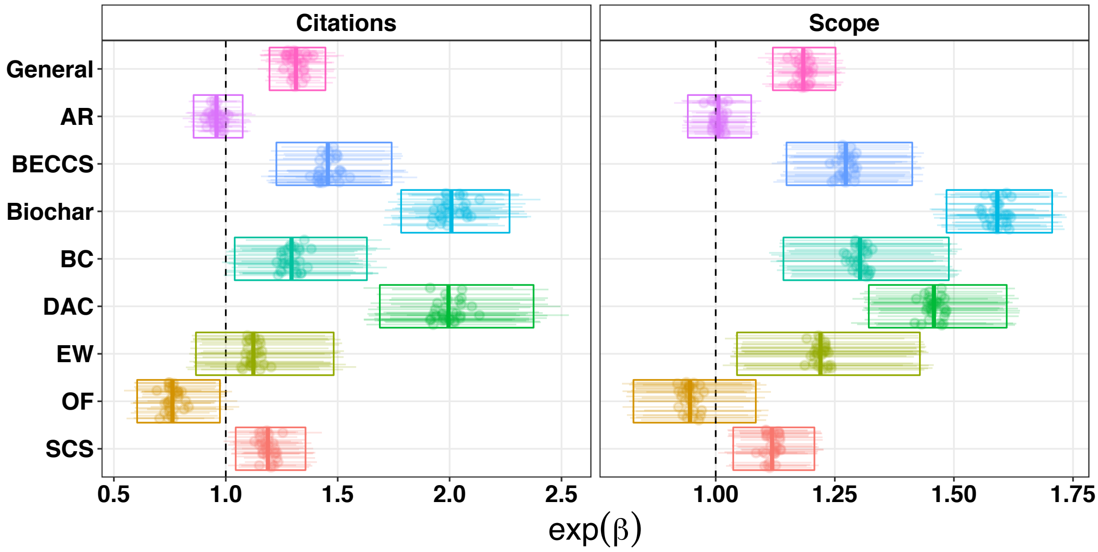


Figure C.12: NETs spillovers to science - WoS sample (climate) control

Table C.8: Coefficients and C.I. Figure C.12

	Citations		Scope	
term	$\exp\beta$	$\langle C.I.\rangle$	$\exp\beta$	$\langle C.I.\rangle$
General	1.31	[1.2,1.45]	1.18	[1.12,1.25]
AR	0.96	[0.86,1.08]	1.01	[0.94,1.07]
BECCS	1.46	[1.23,1.74]	1.27	[1.15,1.41]
Biochar	2.01	[1.78,2.27]	1.59	[1.48,1.71]
BC	1.29	[1.04,1.63]	1.30	[1.14,1.49]
DAC	1.99	[1.69,2.38]	1.46	[1.32,1.61]
EW	1.12	[0.87,1.48]	1.22	[1.04,1.43]
OF	0.76	[0.6,0.97]	0.95	[0.83,1.08]
SCS	1.19	[1.04,1.36]	1.12	[1.04,1.21]
Year dummies	✓		✓	
Matched samples	30		30	
# of obs.	3716		3716	

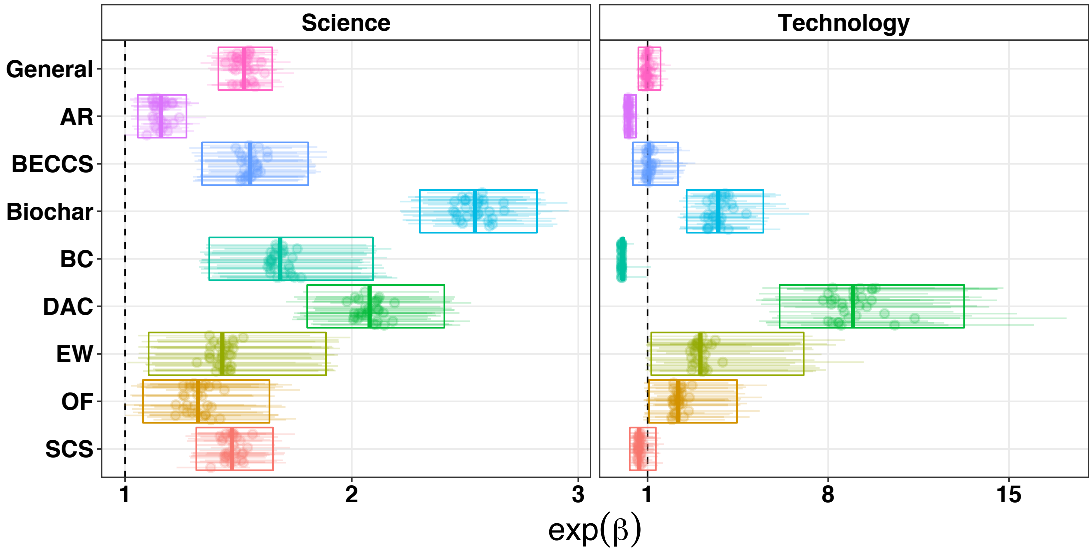


Figure C.13: NETs spillovers to science & technology - WoS/RoS sample

Table C.9: Coefficients and C.I. Figure C.13

	Science	Technology
term	$\exp\beta$ $\langle C.I. \rangle$	$\exp\beta$ $\langle C.I. \rangle$
General	1.53 [1.41,1.65]	0.99 [0.65,1.5]
AR	1.16 [1.06,1.27]	0.27 [0.11,0.55]
BECCS	1.55 [1.34,1.81]	1.05 [0.43,2.18]
Biochar	2.54 [2.3,2.82]	3.74 [2.52,5.49]
BC	1.68 [1.37,2.09]	✗
DAC	2.08 [1.8,2.41]	8.96 [6.12,13.27]
EW	1.43 [1.1,1.89]	3.05 [1.14,7.05]
OF	1.32 [1.08,1.64]	2.19 [1.05,4.47]
SCS	1.47 [1.31,1.65]	0.68 [0.31,1.32]
Year dummies	✓	✓
Matched samples	30	30
# of obs.	5822	5822

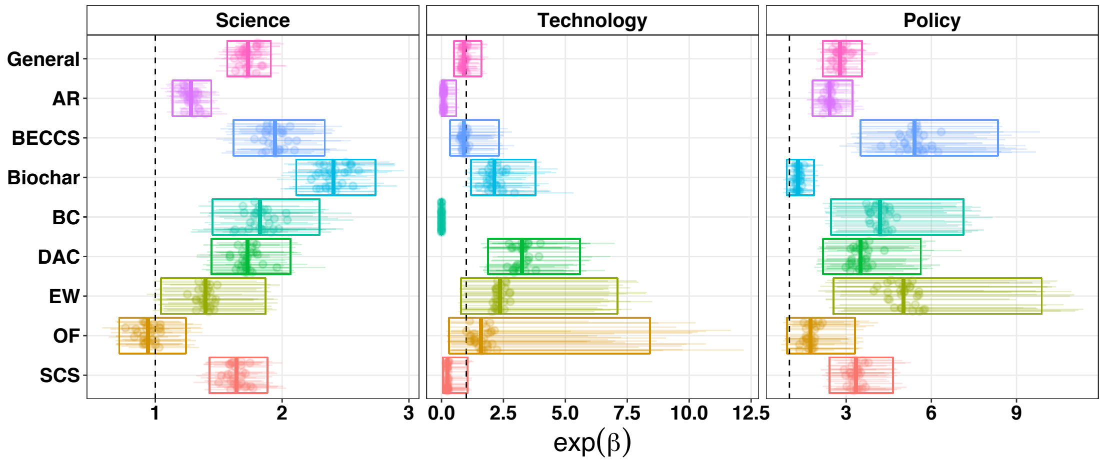


Figure C.14: NETs spillovers to science, technology and policy - Additional controls

Table C.10: Coefficients and C.I. Figure C.14

	Science		Technology		Policy	
NET	$\exp\beta$	$\langle C.I.\rangle$	$\exp\beta$	$\langle C.I.\rangle$	$\exp\beta$	$\langle C.I.\rangle$
General	1.73	[1.57,1.91]	0.90	[0.51,1.61]	2.78	[2.18,3.55]
AR	1.28	[1.14,1.44]	0.08	[0.01,0.6]	2.42	[1.82,3.22]
BECCS	1.94	[1.62,2.34]	0.90	[0.35,2.32]	5.41	[3.5,8.35]
Biochar	2.40	[2.11,2.74]	2.13	[1.19,3.79]	1.30	[0.91,1.86]
BC	1.83	[1.45,2.3]	✗		4.19	[2.46,7.13]
DAC	1.73	[1.44,2.07]	3.24	[1.88,5.59]	3.51	[2.19,5.63]
EW	1.40	[1.04,1.87]	2.35	[0.78,7.1]	5.02	[2.55,9.89]
OF	0.94	[0.72,1.24]	1.60	[0.3,8.42]	1.74	[0.91,3.32]
SCS	1.64	[1.43,1.88]	0.24	[0.06,1.06]	3.35	[2.41,4.65]
Year dummies	✓		✓		✓	
Controls	✓		✓		✓	
Matched samples	30		30		30	
# of obs.	3378		3378		3378	

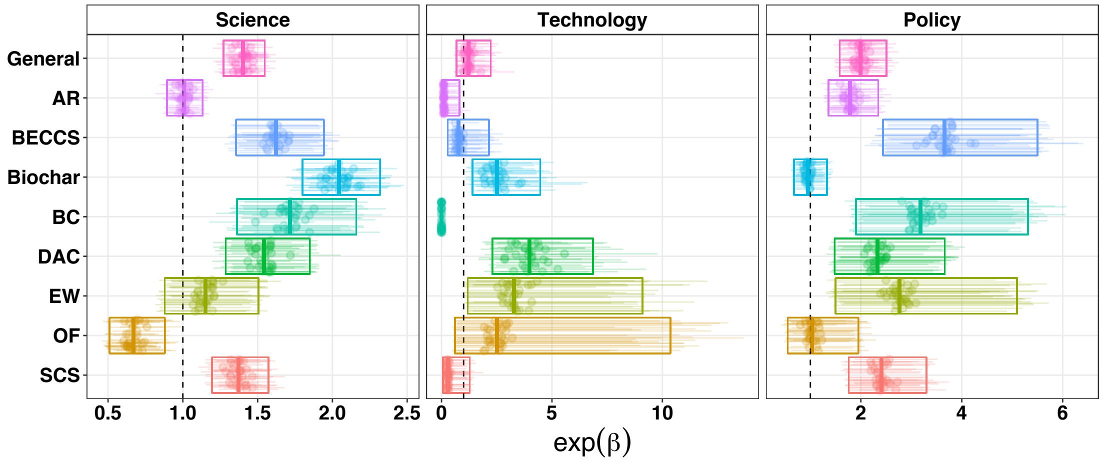


Figure C.15: NETs spillovers to science, technology and policy - Additional controls (climate)

Table C.11: Coefficients and C.I. Figure C.15

	Science		Technology		Policy	
NET	$\exp\beta$	$\langle C.I. \rangle$	$\exp\beta$	$\langle C.I. \rangle$	$\exp\beta$	$\langle C.I. \rangle$
General	1.40	[1.27,1.55]	1.22	[0.67,2.23]	1.99	[1.58,2.51]
AR	1.01	[0.89,1.13]	0.10	[0.01,0.81]	1.79	[1.36,2.34]
BECCS	1.62	[1.36,1.94]	0.78	[0.28,2.14]	3.66	[2.44,5.5]
Biochar	2.04	[1.8,2.32]	2.50	[1.4,4.47]	0.95	[0.68,1.33]
BC	1.72	[1.36,2.16]	✗		3.18	[1.91,5.31]
DAC	1.54	[1.29,1.85]	3.97	[2.3,6.86]	2.33	[1.48,3.66]
EW	1.15	[0.88,1.51]	3.29	[1.19,9.11]	2.77	[1.5,5.1]
OF	0.67	[0.51,0.88]	2.50	[0.6,10.36]	1.04	[0.55,1.95]
SCS	1.37	[1.2,1.57]	0.29	[0.07,1.28]	2.41	[1.76,3.3]
Year dummies	✓		✓		✓	
Controls	✓		✓		✓	
Matched samples	30		30		30	
# of obs.	3518		3518		3518	

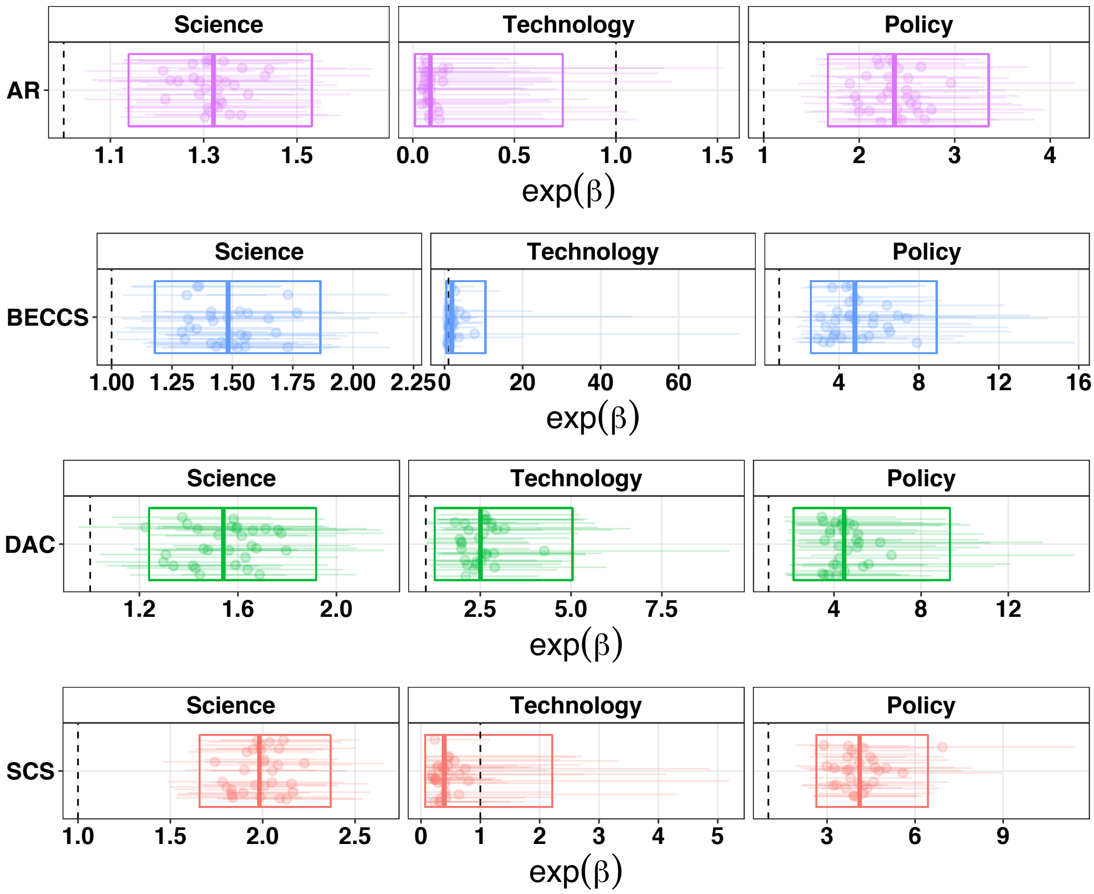


Figure C.16: NETs spillovers to science, technology and policy - Separate matching control (AR) Afforestation and Reforestation. **(BECCS)** Bio-energy with Carbon Capture ans Storage. **(DAC)** Direct Air Capture. **(SCS)** Soil Carbon Sequestration.

C.7 Geography

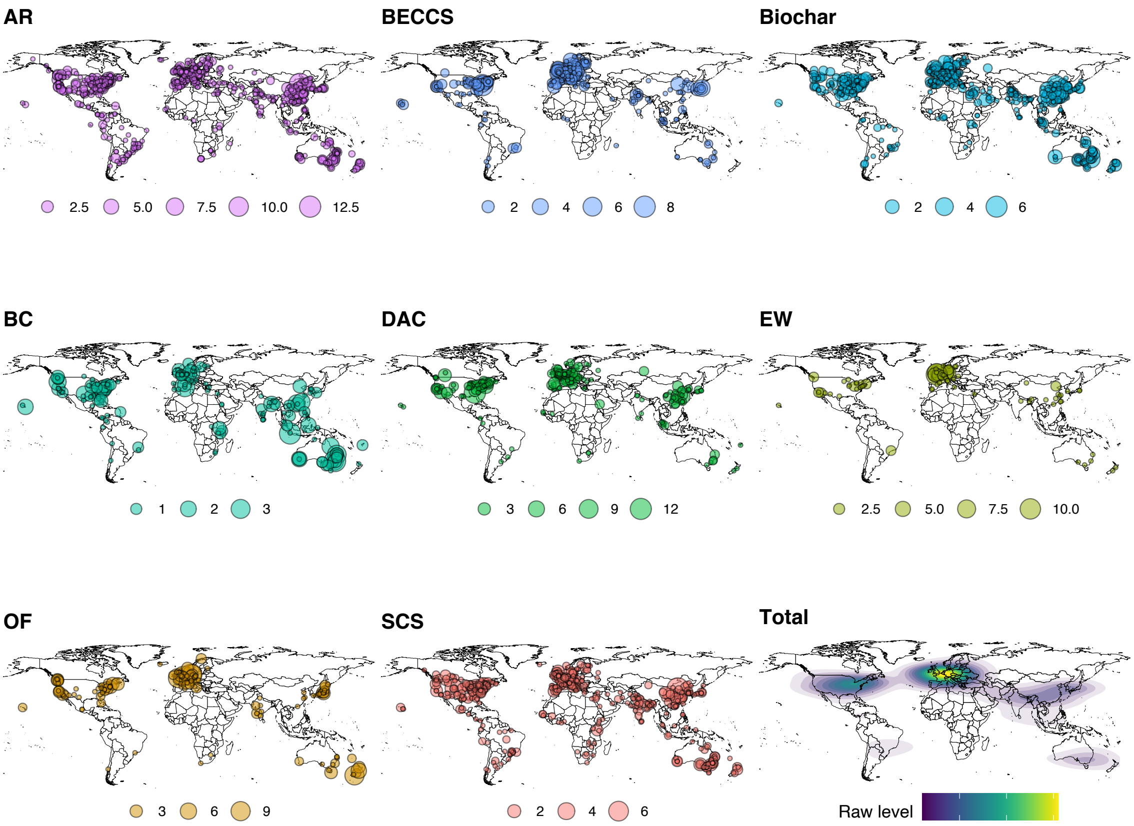


Figure C.17: Geographical distribution of NETs research. Geographical distribution of negative emissions articles at city-level by category (% values). The total map depicts the aggregate unweighted density of cities where NETs research is performed. Geo-localized data are described in Section 4.7.3.

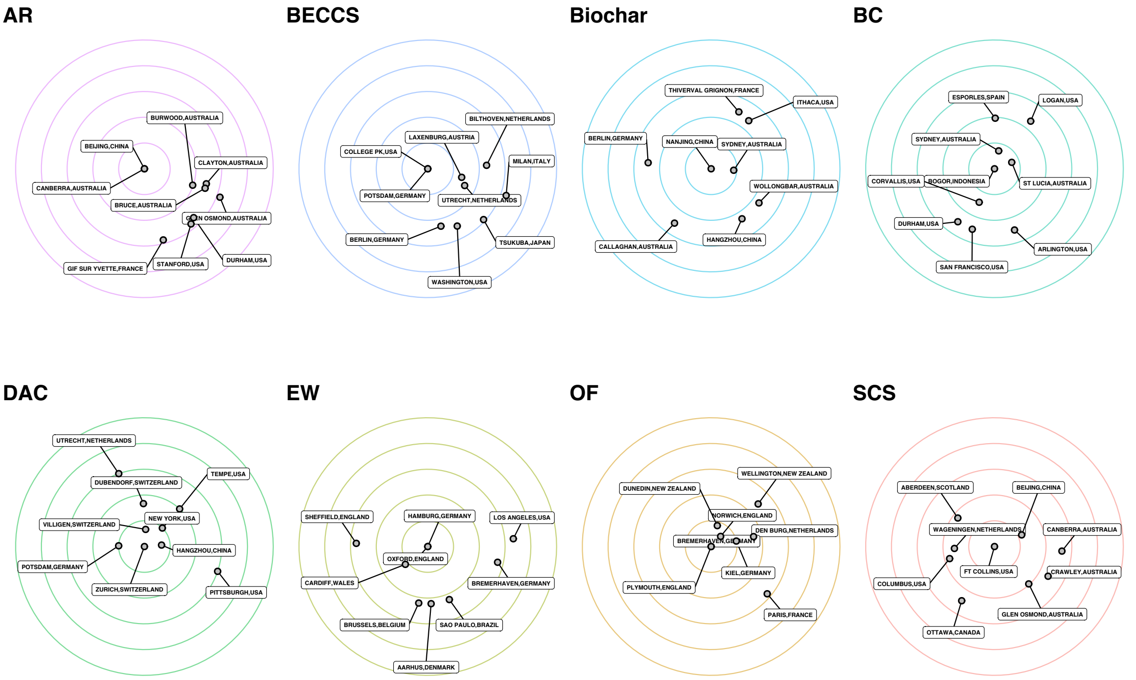


Figure C.18: Most central cities. The most central cities for each NET-specific collaboration network. Centrality is measured by computing nodes' strength in the collaboration network based on affiliation data (see Section 4.7.3).

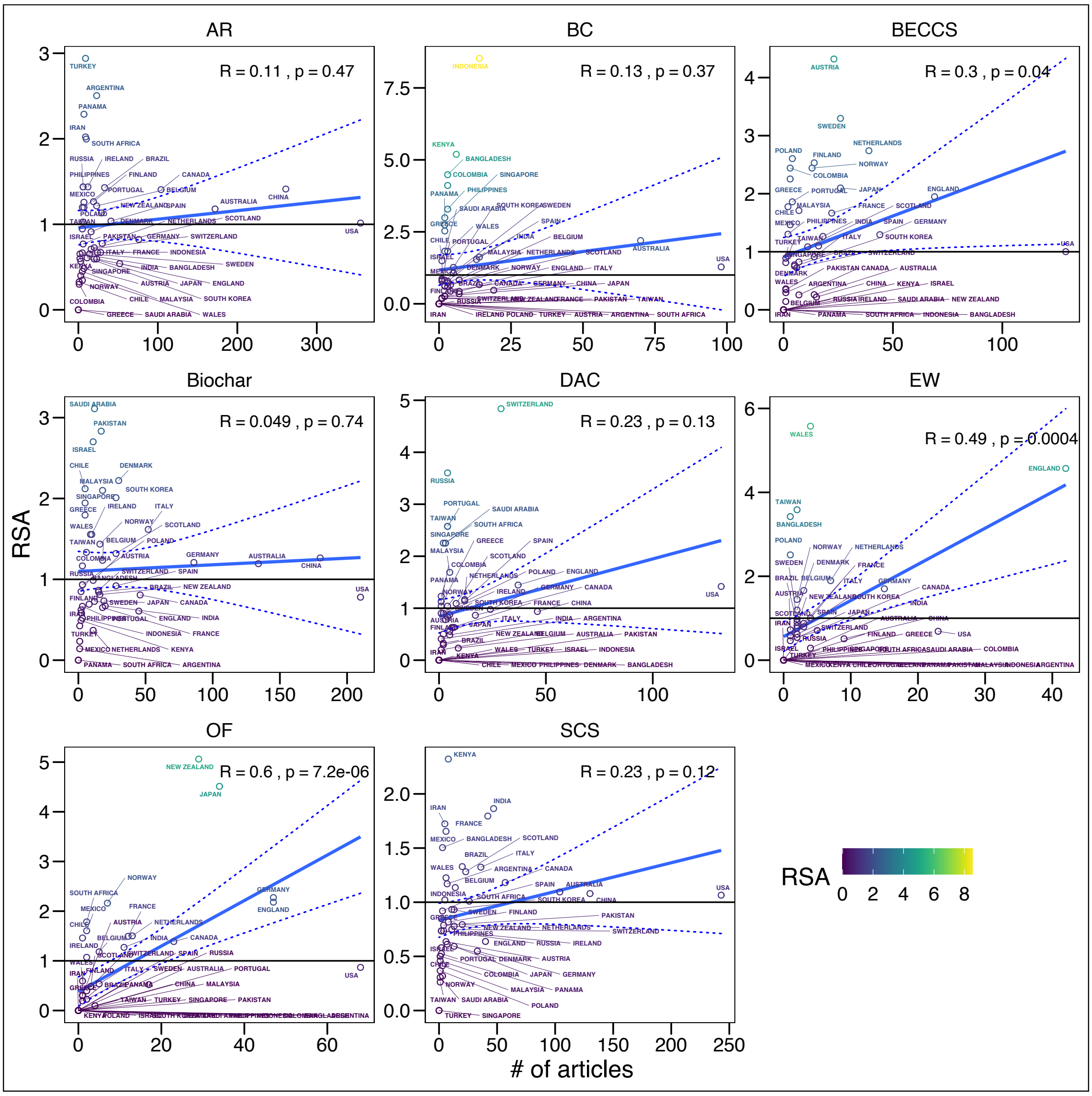


Figure C.19: Correlation between RSA and total number of articles. Scatter plots for a subset of countries (i.e., countries with at least 10 articles in NETs) for each NET category. The RSA horizontal reference line fixed at 1 indicates relative advantage.

Bibliography

Adams, J. D. and Jaffe, A. B. (1996). Bounding the effects of R & D: An investigation using matched establishment-firm data. *The RAND Journal of Economics*, 27(4):700–721.

Adams, P., Fontana, R., and Malerba, F. (2016). User-industry spinouts: Downstream industry knowledge as a source of new firm entry and survival. *Organization Science*, 27(1):18–35.

Adams, P., Fontana, R., and Malerba, F. (2018). Vertical spinouts: Linking vertically related industries. *Industrial and Corporate Change*, forthcoming.

Ahmadpoor, M. and Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science*, 357(6351):583–587.

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, 45(3):425–455.

Alcacer, J. and Gittelman, M. (2006). Patent citations as a measure of knowledge flows: The influence of examiner citations. *The Review of Economics and Statistics*, 88(4):774–779.

Aleta, A., Meloni, S., Perra, N., and Moreno, Y. (2019). Explore with caution: mapping the evolution of scientific interest in physics. *EPJ Data Science*, 8(1):27.

Allen, M., Antwi-Agyei, P., Aragon-Durand, F., Babiker, M., Bertoldi, P., Bind, M., Brown, S., Buckeridge, M., Camilloni, I., Cartwright, A., Cramer, W., Dasgupta, P., Diedhiou, A., Djalante, R., Dong, W., Ebi, K., Engelbrecht, F., Fifita, S., Ford, J., Fuß, S., Hayward, B., Hourcade, J.-C., Ginzburg, V., Guiot, J., Handa, C., Hijioka, Y., Humphreys, S., Kainuma, M., Kala, J., Kanninen, M., Kheshgi, H., Kobayashi, S., Kriegler, E., Ley, D., Liverman, D., Mahowald, N., Mechler, R., Mehrotra, S., Mulugetta, Y., Mundaca, L., Newman, P., Okereke, C., Payne, A., Perez, R., Pinho, P., Revokatova, A., Riahi, K., Schultz, S., Seferian, R., Seneviratne, S., Steg, L., Rogriguez, A., Sugiyama, T., Thonas, A., Vilarino, M., Wairiu, M., Warren, R., Zhou, G., and Zickfeld, K. (2019). Technical summary: Global warming of 1.5?c. an ipcc special report on the impacts of global warming of 1.5?c above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.

Almeida, P. and Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. *Management science*, 45(7):905–917.

Alshamsi, A., Pinheiro, F. L., and Hidalgo, C. A. (2018). Optimal diversification strategies in the networks of related products and of related research areas. *Nature communications*, 9(1):1328.

Amable, B. and Verspagen, B. (1995). The role of technology in market shares dynamics. *Applied Economics*, 27(2):197–204.

Amendola, G., Dosi, G., and Papagni, E. (1993). The dynamics of international competitiveness. *Weltwirtschaftliches Archiv*, 129(3):451–471.

Andersen, K. V. (2013). The problem of embeddedness revisited: Collaboration and market types. *Research Policy*, 42(1):139 – 148.

Anderson, K. and Peters, G. (2016). The trouble with negative emissions. *Science*, 354(6309):182–183.

Antonelli, C., Crespi, F., Ospina, C. A. M., and Scellato, G. (2017). Knowledge composition, jacobs externalities and innovation performance in european regions. *Regional Studies*, 51(11):1708–1720.

Arora, A., Belenzon, S., Patacconi, A., and Suh, J. (2020). The changing structure of american innovation: Some cautionary remarks for economic growth. *Innovation Policy and the Economy*, 20:39–93.

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity: Economic and social factors*, pages 609–626. Princeton University Press.

Arzaghi, M. and Henderson, J. V. (2008). Networking off madison avenue. *The Review of Economic Studies*, 75(4):1011.

Audretsch, D. B. and Feldman, M. P. (1996). R & D spillovers and the geography of innovation and production. *The American Economic Review*, 86(3):630–640.

Baes, C. F., Beall, S. E., Lee, D. W., and Marland, G. (1980). *The Collection, Disposal, and Storage of Carbon Dioxide*, pages 495–519. Springer Netherlands, Dordrecht.

Balassa, B. (1965). Trade liberalisation and "revealed" comparative advantage 1. *The manchester school*, 33(2):99–123.

Balland, P.-A., Boschma, R., and Frenken, K. (2015). Proximity and innovation: From statics to dynamics. *Regional Studies*, 49(6):907–920.

Balland, P.-A., Jara-Figueroa, C., Petralia, S. G., Steijn, M. P., Rigby, D. L., and Hidalgo, C. A. (2020). Complex economic activities concentrate in large cities. *Nature Human Behaviour*, 4(3):248–254.

Barrat, A., Barthélemy, M., Pastor-Satorras, R., and Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11):3747–3752.

Basalla, G. (1988). *The Evolution of Technology*. Cambridge Studies in the History of Science. Cambridge University Press.

Bathelt, H., Malmberg, A., and Maskell, P. (2004). Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, 28(1):31–56.

Battiston, F., Musciotto, F., Wang, D., Barabási, A.-L., Szell, M., and Sinatra, R. (2019). Taking census of physics. *Nature Reviews Physics*, 1(1):89.

Becker, W. and Peters, J. (2000). Technological opportunities, absorptive capacities, and innovation. Technical report, Volkswirtschaftliche Diskussionsreihe, Institut für Volkswirtschaftslehre der Universität Augsburg.

Bertram, C., Quaas, M., Reusch, T. B., Vafeidis, A. T., Wolff, C., and Rickels, W. (2021). The blue carbon wealth of nations. *Nature Climate Change*, 11(8):704–709.

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*, 110(4):1104–44.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182.

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, 27(1):55 – 71.

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

Boschma, R., Balland, P. A., and Kogler, D. F. (2014). Relatedness and technological change in cities: the rise and fall of technological knowledge in us metropolitan areas from 1981 to 2010. *Industrial and Corporate Change*.

Boschma, R. A. and ter Wal, A. L. J. (2007). Knowledge networks and innovative performance in an industrial district: The case of a footwear district in the south of italy. *Industry and Innovation*, 14(2):177–199.

Bowen, Y. and Jianxi, L. (2016). Measuring technological distance for patent mapping. *Journal of the Association for Information Science and Technology*, 68(2):423–437.

Breschi, S. and Lissoni, F. (2001). Knowledge spillovers and local innovation systems: A critical survey. *Industrial and Corporate Change*, 10(4):975–1005.

Breschi, S. and Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of economic geography*, 9(4):439–468.

Breschi, S., Lissoni, F., and Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. *Research Policy*, 32(1):69 – 87.

Breschi, S. and Malerba, F. (2005). *Clusters, Networks and Innovation*. OUP Oxford.

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology*, 110(2):349–399.

Campos-Sánchez, R., Cremona, M. A., Pini, A., Chiaromonte, F., and Makova, K. D. (2016). Integration and fixation preferences of human and mouse endogenous retroviruses uncovered with functional data analysis. *PLOS Computational Biology*, 12(6):1–41.

Cantwell, J. and Andersen, B. (1996). A statistical analysis of corporate technological leadership historically. *Economics of Innovation and New Technology*, 4(3):211–234.

Carlin, W., Glyn, A., and Van Reenen, J. (2001). Export market performance of OECD countries: An empirical examination of the role of cost competitiveness. *The Economic Journal*, 111(468):128–162.

Carlino, G. and Kerr, W. R. (2015). Chapter 6 - Agglomeration and Innovation. In Gilles Duranton, J. V. H. and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*, pages 349 – 404. Elsevier.

Carlino, G. A., Chatterjee, S., and Hunt, R. M. (2007). Urban density and the rate of invention. *Journal of Urban Economics*, 61(3):389–419.

Castellacci, F. (2008). Innovation and the competitiveness of industries: Comparing the mainstream and the evolutionary approaches. *Technological Forecasting and Social Change*, 75(7):984 – 1006.

Catalini, C. (2018). Microgeography and the direction of inventive activity. *Management Science*, 64(9):4348–4364.

Chinazzi, M., Gonçalves, B., Zhang, Q., and Vespignani, A. (2019). Mapping the physics research space: a machine learning approach. *EPJ Data Science*, 8(1):33.

Christensen, C., McDermott, C., Davenport, T., Hargadon, A. B., Kotnour, T. G., Reeves, R., Mallak, L. A., Cardullo, M. W., Anderson, P., Iansiti, M., Betz, F., Greenstein, S. M., Tidd, J., and Pavitt, K. (1998). Innovation and change. In Dorf, R. C. ., editor, *The Technology Management Handbook*. CRC Press.

Cimoli, M., Dosi, G., and Stiglitz, J. E. (2009). The political economy of capabilities accumulation: the past and future of policies for industrial development. In *Industrial Policy and Development: The Political Economy of Capabilities Accumulation*. Oxford University Press.

Clarke, L., Jiang, K., Akimoto, K., Babiker, M., Blanford, G., Fisher-Vanden, K., Hourcade, J.-C., Krey, V., Kriegler, E., Löschel, A., McCollum, D., Paltsev, S., Rose, S., Shukla, P., Tavoni, M., van der Zwaan, B., , and van Vuuren, D. (2014). Assessing transformation pathways. in: Climate change 2014: Mitigation of climate change. contribution of working group iii to the fifth assessment report of the intergovernmental panel on climate change.

Cockburn, I. M., Kortum, S., and Stern, S. (2002). Are all patent examiners equal? the impact of examiner characteristics. Working Paper 8980, National Bureau of Economic Research.

Coffano, M. and Tarasconi, G. (2014). Crios - patstat database: Sources, contents and access rules. center for research on innovation, organization and strategy, crios working paper n.1.

Cohen, W. M. and Levinthal, D. A. (1989). Innovation and learning: The two faces of R&D. *Economic Journal*, 99(397):569–96.

Cohen, W. M. and Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Sciences Quarterly*, 35:128–152.

Cohen, W. M. and Malerba, F. (2001). Is the tendency to variation a chief cause of progress? *Industrial and Corporate Change*, 10(3):587–608.

Criscuolo, P. and Verspagen, B. (2008). Does it matter where patent citations come from? inventor vs. examiner citations in european patents. *Research Policy*, 37(10):1892 – 1908. Special Section Knowledge Dynamics out of Balance: Knowledge Biased, Skewed and Unmatched.

de Coninck, H., Revi, A., Babiker, M., Bertoldi, P., Buckeridge, M., Cartwright, A., Dong, W., Ford, J., Fuss, S., Hourcade, J.-C., Ley, D., Mechler, R., Newman, P., Revokatova, A., Schultz, S., Steg, L., Sugiyama, T., Araos, M., Bakker, S., Bazaz, A., Belfer, E., Benton, T., Camiloni, I., Connors, S., Dasgupta, D., Ebi, K., den Elzen, M., Pinho, P., Forster, P., Fuglestvedt, J., Ghersi, F., Ginzburg, V., Grandis, A., Hayward, B., Haughey, E., Guldberg, O. H., Jiang, K., Kala, J., Klein, R., de Kleijne, K., Liverman, D., Dominguez, M., Mehrotra, S., Mundaca, L., Opio, C., Payne, A., Plazzotta, M., Pereira, J., Reisinger, A., Rhiney, K., Roberts, T., Rogelj, J., van Rooij, A., Seferian, R., Shindell, D., Singh, C., Slade, R., Sparovek, G., Suarez, A., Seneviratne, S., Sillmann, J., Solecki, W., Suarez, A., Taylor, M., Thomas, A., Trutnevyte, E., van Valkengoed, A., and Wollenberg, L. (2018). Chapter 4: Strengthening and implementing the global response. In *Global Warming of 1.5 ?C an IPCC special report on the impacts of global warming of 1.5 ?C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change*. Intergovernmental Panel on Climate Change.

De Noni, I., Ganzaroli, A., and Orsi, L. (2017). The impact of intra-and inter-regional knowledge collaboration and technological variety on the knowledge productivity of european regions. *Technological Forecasting and Social Change*, 117:108–118.

Dechezleprêtre, A., Martin, R., and Mohnen, M. (2013). *Knowledge spillovers from clean and dirty technologies: A patent citation analysis*. Grantham Research Institute on Climate Change and the Environment.

Diodato, D., Malerba, F., and Morrison, A. (2018). The made-in effect and leapfrogging: A model of leadership change for products with country-of-origin bias. *European Economic Review*, 101:297–329.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research policy*, 11(3):147–162.

Dosi, G. (1988). Sources, procedures, and microeconomic effects of innovation. *Journal of Economic Literature*, 26(3):1120–1171.

Dosi, G., Faillo, M., and Marengo, L. (2008). Organizational capabilities, patterns of knowledge accumulation and governance structures in business firms: An introduction. *Organization Studies*, 29(8-9):1165–1185.

Dosi, G., Grazzi, M., and Moschella, D. (2015). Technology and costs in international competitiveness: From countries and sectors to firms. *Research Policy*, 44(10):1795 – 1814.

Dosi, G. and Nelson, R. R. (2010). Technical change and industrial dynamics as evolutionary processes. In *Handbook of the Economics of Innovation*, volume 1, pages 51–127. Elsevier.

Dosi, G., Pavitt, K., and Soete, L. (1990). *The Economics of Technical Change and International Trade*. New York University Press, Washington Square, New York.

Duernecker, G., Meyer, M., and Redondo, F. V. (2015). Being close to grow faster: A network-based empirical analysis of economic globalization. Working paper, World Bank,University of Mannheim and Bocconi Univesity.

Duernecker, G. and Vega-Redondo, F. (2017). Social networks and the process of globalization. *The Review of Economic Studies*, page rdx054.

Dyson, F. J. (1977). Can we control the carbon dioxide in the atmosphere? *Energy*, 2(3):287–291.

Engelsman, E. and van Raan, A. (1994). A patent-based cartography of technology. *Research Policy*, 23(1):1 – 26.

Fagerberg, J. (1988). International competitiveness. *The Economic Journal*, 98(391):355–374.

Fagerberg, J. (1992). The home market hypothesis re-examined: The impact of domestic user-producer interaction on export specialisation. In Lundvall, B. , editor, *National Systems of Innovation. Towards a Theory of Innovation and Interactive Learning*. Pinter Publishers, London.

Fagerberg, J. (1995). User - producer interaction, learning and comparative advantage. *Cambridge Journal of Economics*, 19(1):243–256.

Fagerberg, J. (1996). Technology and competitiveness. *Oxford Review of Economic Policy*, 12(3):39–51.

Fagerberg, J. (1997). Competitiveness, scale and R&D. In Fagerberg, J., Hansson, P., Lundberg, L., and Melchior, A., editors, *Technology and international trade*. Cheltenham: Edward Elgar.

Fagerberg, J., Mowery, D. C., Powell, W. W., and Grodal, S. (2006). Networks of innovators.

Ferguson, J. P. and Carnabuci, G. (2017). Risky recombinations: Institutional gatekeeping in the innovation process. *Organization Science*, 28(1):133–151.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science*, 47(1):117–132.

Fleming, L. and Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. *Research Policy*, 30(7):1019 – 1039.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., et al. (2018a). Science of science. *Science*, 359(6379):eaao0185.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., and Barabási, A.-L. (2018b). Science of science. *Science*, 359(6379).

Foster, J. G., Rzhetsky, A., and Evans, J. A. (2015). Tradition and innovation in scientists' research strategies. *American Sociological Review*, 80(5):875–908.

Freeman, C. (1974). *The Economics of Industrial Innovation*. Harmondsworth: Penguin Books.

Freeman, C. (1987). *Technology, policy, and economic performance: Lessons from japan*. Pinter, London and New York.

Freeman, C., Clark, J., and Soete, L. (1982). *Unemployment and technical innovation: a study of long waves and economic development*. Burns & Oates.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215 – 239.

Fuhrman, J., McJeon, H., Patel, P., Doney, S. C., Shobe, W. M., and Clarens, A. F. (2020). Food–energy–water implications of negative emissions technologies in a+ 1.5 c future. *Nature Climate Change*, 10(10):920–927.

Fuss, S., Canadell, J. G., Peters, G. P., Tavoni, M., Andrew, R. M., Ciais, P., Jackson, R. B., Jones, C. D., Kraxner, F., Nakicenovic, N., et al. (2014). Betting on negative emissions. *Nature climate change*, 4(10):850–853.

Fuss, S., Jones, C. D., Kraxner, F., Peters, G. P., Smith, P., Tavoni, M., van Vuuren, D. P., Canadell, J. G., Jackson, R. B., Milne, J., et al. (2016). Research priorities for negative emissions. *Environmental Research Letters*, 11(11):115007.

Fuss, S., Lamb, W. F., Callaghan, M. W., Hilaire, J., Creutzig, F., Amann, T., Beringer, T., de Oliveira Garcia, W., Hartmann, J., Khanna, T., et al. (2018). Negative emissions—part 2: Costs, potentials and side effects. *Environmental Research Letters*, 13(6):063002.

Gallaud, D. and Torre, A. (2005). Geographical proximity and the diffusion of knowledge. In *Rethinking Regional Innovation and Change*, pages 127–146. Springer.

Galunic, D. C. and Rodan, S. (1998). Resource recombinations in the firm: Knowledge structures and the potential for schumpeterian innovation. *Strategic management journal*, pages 1193–1201.

Garcia-Vega, M. (2006). Does technological diversification promote innovation? An empirical analysis for European firms. *Research Policy*, 35(2):230–246.

Gittelman, M. (2006). National institutions, public–private knowledge flows, and innovation performance: A comparative study of the biotechnology industry in the us and france. *Research Policy*, 35(7):1052 – 1068.

Greenhalgh, C. (1990). Innovation and trade performance in the united kingdom. *The Economic Journal*, 100(400):105–118.

Grieneisen, M. L. and Zhang, M. (2011). The current status of climate change research. *Nature Climate Change*, 1(2):72–73.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics*, 10(1):92–116.

Griliches, Z. (1998). *R&D and Productivity: the Econometric Evidence*. University of Chicago Press.

Gross, D. P. and Sampat, B. N. (2021a). Crisis innovation policy from world war ii to covid-19. Technical report, National Bureau of Economic Research.

Gross, D. P. and Sampat, B. N. (2021b). The economics of crisis innovation policy: A historical perspective. In *AEA Papers and Proceedings*, volume 111, pages 346–50.

Grossman, G. and Helpman, E. (1991). *Innovation and Growth in the Global Economy.* The MIT Press, Cambridge, MA.

Grossman, G. and Helpman, E. (1995). Technology and trade. In Grossman, G. M. and Rogoff, K., editors, *Handbook of International Economics*, volume 3, chapter 25, pages 1279–1337. Elsevier, 1 edition.

Grubler, A., Wilson, C., Bento, N., Boza-Kiss, B., Krey, V., McCollum, D. L., Rao, N. D., Riahi, K., Rogelj, J., De Stercke, S., et al. (2018). A low energy demand scenario for meeting the 1.5 c target and sustainable development goals without negative emission technologies. *Nature energy*, 3(6):515–527.

Guarascio, D. and Pianta, M. (2017). The gains from technology: new products, exports and profits. *Economics of Innovation and New Technology*, 26(8):779–804.

Guarascio, D., Pianta, M., and Bogliacino, F. (2016). Export, r&d and new products. a model and a test on european industries. *Journal of Evolutionary Economics*, 26(4):869–905.

Guarascio, D., Pianta, M., Lucchese, M., and Bogliacino, F. (2015). Business cycles, technology and exports. *Economia Politica*, 32(2):167–200.

Hagedoorn, J. (2002). Inter-firm r&d partnerships: an overview of major trends and patterns since 1960. *Research policy*, 31(4):477–492.

Hanna, R., Abdulla, A., Xu, Y., and Victor, D. G. (2021). Emergency deployment of direct air capture as a response to the climate crisis. *Nature communications*, 12(1):1–13.

Hargadon, A. and Sutton, R. I. (1997). Technology brokering and innovation in a product development firm. *Administrative Science Quarterly*, 42(4):716–749.

Hargadon, A. B. (2002). Brokering knowledge: Linking learning and innovation. *Research in Organizational Behavior*, 24:41 – 85.

Helfat, C. E. (2015). Vertical firm structure and industry evolution. *Industrial and Corporate Change*, 24(4):803–818.

Helfat, C. E. and Campo-Rembado, M. A. (2016). Integrative capabilities, vertical integration, and innovation over successive technology lifecycles. *Organization Science*, 27(2):249–264.

Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics*, pages 1–22.

Hidalgo, C. A., Klinger, B., Barabási, A.-L., and Hausmann, R. (2007). The product space conditions the development of nations. *Science*, 317(5837):482–487.

Hinze, S., Reiß, T., and Schmoch, U. (1997). Statistical analysis on the distance between fields of technology : Report for research project "innovation systems and european integration (isi)" funded by the european commission (dg xii). subproject 3.2.5: Technological entry: Diversification vs. new innovators.

Hirschman, A. (1958). *The Strategy of Economic Development*. Number v. 10 in The strategy of economic development. Yale University Press.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.

Iansiti, M. (1995). Technology integration: Managing technological evolution in a complex environment. *Research Policy*, 24(4):521 – 542.

Jaffe, A. B. (1986). Technological opportunity and spillovers of R & D: Evidence from firms' patents, profits and market value. *The American Economic Review*, 76(5):984–1001.

Jaffe, A. B. and De Rassenfosse, G. (2019). Patent citation data in social science research: Overview and best practices. *Research handbook on the economics of intellectual property law.*

Jaffe, A. B. and Trajtenberg, M. (1999). International knowledge flows: Evidence from patent citations. *Economics of Innovation and New Technology*, 8(1-2):105–136.

Jaffe, A. B. and Trajtenberg, M. (2002). *Patents, Citations, and Innovations. A Window on the Knowledge Economy.* MIT press, Cambridge, MA.

Jaffe, A. B., Trajtenberg, M., and Fogarty, M. S. (2000). Knowledge spillovers and patent citations: Evidence from a survey of inventors. *American Economic Review*, 90(2):215–218.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3):577–598.

Jia, T., Wang, D., and Szymanski, B. K. (2017). Quantifying patterns of research-interest evolution. *Nature Human Behaviour*, 1(4):0078.

Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *The Review of Economic Studies*, 76(1):283–317.

Jones, B. F. (2021). Science and innovation: The under-fueled engine of prosperity. *Rebuilding the Post-Pandemic Economy, Aspen Institute Economic Strategy Group.*

Joo, S. H. and Kim, Y. (2009). Measuring relatedness between technological fields. *Scientometrics*, 83(2):435–454.

Joppa, L., Luers, A., Willmott, E., Friedmann, S. J., Hamburg, S. P., and Broze, R. (2021). Microsoft's million-tonne co2-removal purchase — lessons for net zero. *Nature*, 597(7878):629–632.

Kaldor, N. (1978). The effect of devaluations on trade in manufactures. In *Further Essays on Applied Economics*, pages 99–118. Duckworth, London.

Klevorick, A. K., Levin, R. C., Nelson, R. R., and Winter, S. G. (1995). On the sources and significance of interindustry differences in technological opportunities. *Research Policy*, 24(2):185 – 205.

Kodama, F. (1986). Technological diversification of japanese industry. *Science*, 233(4761):291–296.

Kogler, D. F., Rigby, D. L., and Tucker, I. (2013). Mapping knowledge space and technological relatedness in us cities. *European Planning Studies*, 21(9):1374–1391.

Kremer, M., Levin, J., and Snyder, C. M. (2020). Advance market commitments: insights from theory and experience. In *AEA Papers and Proceedings*, volume 110, pages 269–73.

Krugman, P. (1980). Scale Economies, Product Differentiation, and the Pattern of Trade. *American Economic Review*, 70(5):950–959.

Krugman, P. (1992). *Geography and trade.* MIT press.

Kuhn, T. S. (1977). *The essential tension.* Chicago: The University of Chicago Press,.

Lamperti, F., Malerba, F., Mavilia, R., and Tripodi, G. (2019). Does the position in the inter-sectoral knowledge space affect the international competitiveness of industries? *Economics of Innovation and New Technology*, pages 1–48.

Lamperti, F., Malerba, F., Mavilia, R., and Tripodi, G. (2020). Does the position in the inter-sectoral knowledge space affect the international competitiveness of industries? *Economics of Innovation and New Technology*, 29(5):441–488.

Lane, J., Greig, C., and Garnett, A. (2021). Uncertain storage prospects create a conundrum for carbon capture and storage ambitions. *Nature Climate Change*, 11(11):925–936.

Laursen, K. (1999). The impact of technological opportunity on the dynamics of trade performance. *Structural Change and Economic Dynamics*, 10(3–4):341 – 357.

Laursen, K. and Drejer, I. (1999). Do inter-sectoral linkages matter for international export specialisation? *Economics of Innovation and New Technology*, 8(4):311–330.

Laursen, K. and Meliciani, V. (2000). The importance of technology-based intersectoral linkages for market share dynamics. *Weltwirtschaftliches Archiv*, 136(4):702–723.

Laursen, K. and Meliciani, V. (2002). The relative importance of international vis-à-vis national technological spillovers for market share dynamics. *Industrial and Corporate Change*, 11(4):875–894.

Laursen, K. and Meliciani, V. (2010). The role of *ICT* knowledge flows for international market share dynamics. *Research Policy*, 39(5):687 – 697. Special Section on Government as Entrepreneur.

Leten, B., Belderbos, R., and Van Looy, B. (2007). Technological diversification, coherence, and performance of firms. *Journal of Product Innovation Management*, 24(6):567–579.

Leydesdorff, L. (2008). Patent classifications as indicators of intellectual organization. *J. Am. Soc. Inf. Sci. Technol.*, 59(10):1582–1597.

Li, D., Capone, G., and Malerba, F. (2018). The long march to catch up: a history friendly model of china's mobile communication industry. *Research policy*, forthcoming.

Linder, S. B. (1961). *An essay on trade and transformation.* Almqvist & Wiksell, Stockholm.

Lissoni, F. and Miguelez, E. (2014). Patents, innovation and economic geography. *WIPO Journal*, 6(5):37–49.

Lundvall, B.-Å. (1988). Innovation as an interactive process: From user producer interaction to national systems of innovation. In Dosi, G and Freeman, C and Nelson, R and Silverberg, G and Soete, L, editor, *Technical Change and Economic theory.* LEM Book Series.

Lundvall, B.-Å. (1992). *National systems of innovation: Toward a theory of innovation and interactive learning*, volume 2. Anthem press.

Lundvall, B.-Å. and Johnson, B. (1994). The learning economy. *Journal of industry studies*, 1(2):23–42.

Magnier, A. and Toujas-Bernate, J. (1994). Technology and trade: Empirical evidences for the major five industrialized countries. *Weltwirtschaftliches Archiv*, 130(3):494–520.

Malerba, F. (2002). Sectoral systems of innovation and production. *Research policy*, 31(2):247–264.

Malerba, F., Nelson, R. R., Orsenigo, L., and Winter, S. G. (2016). *Innovation and the Evolution of Industries: History-Friendly Models.* Cambridge University Press.

Malerba, F. and Orsenigo, L. (1997). Technological regimes and sectoral patterns of innovative activities. *Industrial and corporate change*, 6(1):83–118.

Marx, M. and Fuegi, A. (2020). Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management Journal*, 41(9):1572–1594.

Mazzucato, M. (2013). *Entrepreneurial State: Debunking Public Vs. Private Sector Myths.* Anthem Press.

Meckling, J. and Allan, B. B. (2020). The evolution of ideas in global climate policy. *Nature Climate Change*, 10(5):434–438.

Meckling, J. and Biber, E. (2021). A policy roadmap for negative emissions using direct air capture. *Nature Communications*, 12(1):2051.

Meliciani, V., R. S. (1998). Specialization in areas of strong technological opportunity and economic growth. In Eliasson, G., Green, C., and McCann, C., editors, *Microfoundations of Economic Growth. A Schumpeterian Perspective*. The University of Michigan Press.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *econometrica*, 71(6):1695–1725.

Meyer, M. (2000). What is special about patent citations? differences between scientific and patent citations. *Scientometrics*, 49(1):93–123.

Minx, J. C., Lamb, W. F., Callaghan, M. W., Bornmann, L., and Fuss, S. (2017). Fast growing research on negative emissions. *Environmental Research Letters*, 12(3):035007.

Minx, J. C., Lamb, W. F., Callaghan, M. W., Fuss, S., Hilaire, J., Creutzig, F., Amann, T., Beringer, T., de Oliveira Garcia, W., Hartmann, J., et al. (2018). Negative emissions—part 1: Research landscape and synthesis. *Environmental Research Letters*, 13(6):063001.

Mullainathan, S. and Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106.

Nature-editorial (2022). How researchers can help fight climate change in 2022 and beyond. *Nature*, 601(7891).

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

Nelson, R. and Winter, S. (1982). *An Evolutionary Theory of Economic Change*. Belknap Press Series. Belknap Press of Harvard University Press.

Nelson, R. R. (1959). The simple economics of basic scientific research. *Journal of political economy*, 67(3):297–306.

Nelson, R. R. and Winter, S. G. (1977). In search of useful theory of innovation. *Research Policy*, 6(1):36 – 76.

Nemet, G. F., Callaghan, M. W., Creutzig, F., Fuss, S., Hartmann, J., Hilaire, J., Lamb, W. F., Minx, J. C., Rogers, S., and Smith, P. (2018). Negative emissions—part 3: Innovation and upscaling. *Environmental Research Letters*, 13(6):063003.

Oltra, M. J. and Flor, M. (2003). The impact of technological opportunities and innovative capabilities on firms' output innovation. *Creativity and Innovation Management*, 12(3):137–144.

Operti, E. and Carnabuci, G. (2011). Public knowledge, private gain. *Journal of Management*, 40(4):1042–1074.

Opsahl, T., Agneessens, F., and Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245 – 251.

Owen-Smith, J. and Powell, W. W. (2004). Knowledge networks as channels and conduits: The effects of spillovers in the boston biotechnology community. *Organization Science*, 15(1):5–21.

Pan, R. K., Sinha, S., Kaski, K., and Saramäki, J. (2012). The evolution of interdisciplinarity in physics research. *Scientific reports*, 2:551.

Petralia, S., Balland, P.-A., and Morrison, A. (2017). Climbing the ladder of technological development. *Research Policy*, 46(5):956–969.

Popp, D. (2016). Economic analysis of scientific publications and implications for energy research and development. *Nature Energy*, 1(4):1–8. Number: 4 Publisher: Nature Publishing Group.

Posner, M. V. (1961). International trade and technical change. *Oxford Economic Papers*, 13(3):323–341.

Probst, B., Touboul, S., Glachant, M., and Dechezleprêtre, A. (2021). Global trends in the invention and diffusion of climate change mitigation technologies. *Nature Energy*, 6(11).

Putnam, J. and Evenson, R. E. (1994). Inter-sectoral technology flows: Estimates from a patent concordance with an application to italy. Technical report, Yale University,mimeo.

Riahi, K., Bertram, C., Huppmann, D., Rogelj, J., Bosetti, V., Cabardos, A.-M., Deppermann, A., Drouet, L., Frank, S., Fricko, O., Fujimori, S., Harmsen, M., Hasegawa, T., Krey, V., Luderer, G., Paroussos, L., Schaeffer, R., Weitzel, M., van der Zwaan, B., Vrontisi, Z., Longa, F. D., Després, J., Fosse, F., Fragkiadakis, K., Gusti, M., Humpenöder, F., Keramidas, K., Kishimoto, P., Kriegler, E., Meinshausen, M., Nogueira, L. P., Oshiro, K., Popp, A., Rochedo, P. R. R., Ünlü, G., van Ruijven, B., Takakura, J., Tavoni, M., van Vuuren, D., and Zakeri, B. (2021). Cost and attainability of meeting stringent climate targets without overshoot. *Nature Climate Change*, 11(12).

Rogelj, J., Luderer, G., Pietzcker, R. C., Kriegler, E., Schaeffer, M., Krey, V., and Riahi, K. (2015). Energy system transformations for limiting end-of-century warming to below 1.5 c. *Nature Climate Change*, 5(6):519–527.

Rosenberg, N., Landau, R., and Mowery, D. (1992). *Technology and the Wealth of Nations*. Business and economics. Stanford University Press.

Rosenthal, S. S. and Strange, W. C. (2003). Geography, industrial organization, and agglomeration. *The Review of Economics and Statistics*, 85(2):377–393.

Rueda, O., Mogollón, J., Tukker, A., and Scherer, L. (2021). Negative-emissions technology portfolios to meet the 1.5° c target. *Global Environmental Change*, 67:102238.

Saracco, F., Straka, M. J., Di Clemente, R., Gabrielli, A., Caldarelli, G., and Squartini, T. (2017). Inferring monopartite projections of bipartite networks: an entropy-based approach. *New Journal of Physics*, 19(5):053022.

Saxenian, A. (1990). Regional networks and the resurgence of silicon valley. *California management review*, 33(1):89–112.

Scherer, F. M. (1982). Inter-industry technology flows and productivity growth. *The Review of Economics and Statistics*, 64(4):627–634.

Schumpeter, J. A. (1934). *Theory economic development: : An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Harvard University Press.

Shukla, P., Skea, J., Slade, R., Khourdajie, A. A., van Diemen, R., McCollum, D., Pathak, M., Some, S., Vyas, P., Fradera, R., Belkacemi, M., Hasija, A., Lisboa, G., Luz, S., Malley, J., et al. (2022). Climate change 2022: Mitigation of climate change. contribution of working group iii to the sixth assessment report of the intergovernmental panel on climate change.

Sinatra, R., Wang, D., Deville, P., Song, C., and Barabási, A.-L. (2016). Quantifying the evolution of individual scientific impact. *Science*, 354(6312).

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management science*, 51(5):756–770.

Smith, P. (2016). Soil carbon sequestration and biochar as negative emission technologies. *Global change biology*, 22(3):1315–1324.

Smith, P., Davis, S. J., Creutzig, F., Fuss, S., Minx, J., Gabrielle, B., Kato, E., Jackson, R. B., Cowie, A., Kriegler, E., et al. (2016). Biophysical and economic limits to negative co 2 emissions. *Nature climate change*, 6(1):42–50.

Soete, L. (1987). The impact of technological innovation on international trade patterns: The evidence reconsidered. *Research Policy*, 16(2):101 – 130.

Soete, L. L. G. (1981). A general test of technological gap trade theory. *Weltwirtschaftliches Archiv*, 117(4):638–660.

Sonmez, Z. (2017). Inventor mobility and the geography of knowledge flows: evidence from the us biopharmaceutical industry. *Science and Public Policy*, 44(5):670–682.

Sorenson, O., Rivkin, J. W., and Fleming, L. (2006). Complexity, networks and knowledge flow. *Research Policy*, 35(7):994 – 1017.

Stiglitz, J. E. and Greenwald, B. C. (2014). *Creating a Learning Society: A New Approach to Growth, Development, and Social Progress*. Columbia University Press.

Strong, A., Chisholm, S., Miller, C., and Cullen, J. (2009). Ocean fertilization: time to move on. *Nature*, 461(7262):347–348.

Summers, L. H. and Jones, B. (2020). The social returns to innovation investments. In *Innovation and Public Policy*. University of Chicago Press.

Sun, Y. and Latora, V. (2020). The evolution of knowledge within and across fields in modern physics. *ArXiv*, abs/2001.07199.

Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., and Pietronero, L. (2012). A new metrics for countries' fitness and products' complexity. *Scientific reports*, 2:723.

Tavoni, M. and Socolow, R. (2013). Modeling meets science and technology: an introduction to a special issue on negative emissions. *Climatic Change*, 118(1):1–14.

Teece, D. J., Rumelt, R., Dosi, G., and Winter, S. (1994). Understanding corporate coherence: Theory and evidence. *Journal of Economic Behavior & Organization*, 23(1):1 – 30.

Tomasello, M. V., Tessone, C. J., and Schweitzer, F. (2016). A model of dynamic rewiring and knowledge exchange in r&d networks. *Advances in Complex Systems*, 19(1 - 2).

Torre, A. (2008). On the role played by temporary geographical proximity in knowledge transmission. *Regional Studies*, 42(6):869–889.

Tortoriello, M. (2015). The social underpinnings of absorptive capacity: The moderating effects of structural holes on innovation generation based on external knowledge. *Strategic Management Journal*, 36(4):586–597.

Tripodi, G., Chiaromonte, F., and Lillo, F. (2020). Knowledge and social relatedness shape research portfolio diversification. *Scientific reports*, 10(1):1–12.

Tripodi, G., Lamperti, F., Mavilia, R., Mina, A., Chiaromonte, F., and Lillo, F. (2022). Quantifying knowledge spillovers from advances in negative emissions technologies. Lem papers series 2022/16, laboratory of economics and management (lem), sant'anna school of advanced studies, pisa, italy.

Tumminello, M., Micciche, S., Lillo, F., Piilo, J., and Mantegna, R. N. (2011). Statistically validated networks in bipartite complex systems. *PloS one*, 6(3):e17994.

Vaccario, G., Tomasello, M. V., Tessone, C. J., and Schweitzer, F. (2017). Quantifying knowledge exchange in r&d networks: A data-driven model. *Journal of Evolutionary Economics (accepted). Available at ArXiv.*

Van Looy, B., Vereyen, C., and Schmoch, U. (2015). Patent statistics: Concordance ipc v8—nace rev. 2 (version 2.0). Technical report, Eurostat Working Paper, Luxembourg.

Van Vuuren, D. P., Hof, A. F., Van Sluisveld, M. A., and Riahi, K. (2017). Open discussion of negative emissions is urgently needed. *Nature energy*, 2(12):902–904.

Van Vuuren, D. P., Stehfest, E., Gernaat, D. E., Van Den Berg, M., Bijl, D. L., De Boer, H. S., Daioglou, V., Doelman, J. C., Edelenbosch, O. Y., Harmsen, M., et al. (2018). Alternative pathways to the 1.5 c target reduce the need for negative emission technologies. *Nature climate change*, 8(5):391–397.

Vaughan, N. E. and Gough, C. (2016). Expert assessment concludes negative emissions scenarios may not deliver. *Environmental research letters*, 11(9):095003.

Veech, J. A. (2013). A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, 22(2):252–260.

Verdolini, E. and Galeotti, M. (2011). At home and abroad: An empirical analysis of innovation and diffusion in energy technologies. *Journal of Environmental Economics and Management*, 61(2):119–134.

Vernon, R. (1966). International investment and international trade in the product cycle. *The Quarterly Journal of Economics*, 80(2):190–207.

Verspagen, B. (1997a). Estimating international technology spillovers using technology flow matrices. *Weltwirtschaftliches Archiv*, 133(2):226–248.

Verspagen, B. (1997b). Measuring intersectoral technology spillovers: Estimates from the european and us patent office databases. *Economic Systems Research*, 9(1):47–65.

Von Hippel, E. (1994). ”sticky information” and the locus of problem solving: implications for innovation. *Management science*, 40(4):429–439.

Wakelin, K. (1998). The role of innovation in bilateral oecd trade performance. *Applied Economics*, 30(10):1335–1346.

Wang, D. and Barabási, A.-L. (2021). *The science of science*. Cambridge University Press.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of ”small-world” networks. *Nature*, 393(6684):440–442.

Wuchty, S., Jones, B. F., and Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827):1036–1039.

Xiao, T., Makhija, M., and Karim, S. (2021). A knowledge recombination perspective of innovation: Review and new research directions. *Journal of Management*, page 01492063211055982.

Yin, Y., Dong, Y., Wang, K., Wang, D., and Jones, B. (2021a). Science as a public good: Public use and funding of science. Technical report, National Bureau of Economic Research.

Yin, Y., Gao, J., Jones, B. F., and Wang, D. (2021b). Coevolution of policy and science during the pandemic. *Science*, 371(6525):128–130.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Zeileis, A. (2004). Econometric computing with hc and hac covariance matrix estimators. *Journal of Statistical Software, Articles*, 11(10):1–17.

Zeng, A., Shen, Z., Zhou, J., Fan, Y., Di, Z., Wang, Y., Stanley, H. E., and Havlin, S. (2019). Increasing trend of scientists to switch between topics. *Nature communications*, 10(1):1–11.