Doctoral School of Social Sciences

Doctoral program in Development Economics and Local Systems

Curriculum in Local Systems

Proximity and Innovation:

Analyzing the path through topic modeling and business model design

a dissertation submitted in partial fulfillment of the requirements for the Doctoral degree (Ph.D.)

doctoral program in Development Economics and Local Systems

Matteo

Devigili

XXXIII cycle, 2019/2020

Supervisor Prof. Tommaso Pucci

Doctoral Committee

Prof. Nicola Bellini

Prof. Rune Fitjar

Prof. Lorenzo Zanni

Abstract

This thesis aims to deepen the relationship between the different forms of proximity that emerge between economic actors and the consequent influence on their innovative capacity. Over the years, this topic has generated a great deal of attention in conference proceedings and scientific publications. The first step to deepen the understanding of this amount of knowledge was to identify a suitable methodology. In so doing, the recent advances of the Machine Learning community – particularly Natural Language Processing academics - have offered interesting insights. In particular, "Topic Modeling" was identified as a suitable methodology to bring out latent semantic structures. Therefore, the first chapter tries to study how this methodology has been implemented in the social sciences and, in particular, in management. The contribution offered is a rationalization of the achievable goals and their relationship with evaluation practices. Once clarified how to use this algorithm, the second chapter studied the relationship between proximity and innovation. Using an unsupervised machine learning technique, the research attempts to identify thematic management cores in a multifocal literature such as proximity. Together with a qualitative analysis, the study attempts to bring out the theoretical and empirical contributions offered to the management community. Once the theoretical and empirical expectations have been clarified, the third chapter introduces a strategic theme, namely the business model. This section proposes a mediating effect of the business model concerning the central relationship between proximity and innovation. After a theoretical introduction, the conceptual model is studied with an exploratory approach. Without any presumption of generalizability and completeness, a novel analytical key is offered to open further debate in the community of proximity.

Keywords: proximity, innovation, business model, topic modeling.

Acknowledgments

This thesis is the result of a beautiful stormy journey that began in Siena. Lorenzo has accompanied me throughout my growth path, encouraging me to do better and better, letting me make mistakes, and correct them together. His energy and curiosity showed me the way I am taking today. On this path, Tommaso accompanied me step by step, putting all his strength and power into transmitting his passion and dedication to this profession. A few words, but many facts that have taught me what stubbornness and willpower mean. A journey that brings me to London, where Simone is helping me push my limits every day a step further and teaching me the beauty of knowledge. To prof Pucci, Santoni, and Zanni, I owe more than I could ever repay.

I wish to thank professor Lazzeretti for her passion and dedication to Ph.D. students. Everything she did, she did it for good. I will never forget her desire to get into new scientific challenges, knowing how to start from scratch "a testa alta". She inspired all her colleagues, which I thank for the shown support across the years.

I wish to thank my friends, Andrea, David, Francesco, Francesca, Giuseppe, Livia, Sara, and Tommaso. Because, even if scattered all over the world, you have given me the energy to face every difficulty knowing that I could always count on your support.

Without my family, this would not have been possible. Through death and love, you have taught me to build day after day. From wooden boxes to the Olympics.

Similarly, I must thank Sveva which is more than my partner but my best friend. For her enormous patience in listening to my endless monologues.

This page is intentionally left blank.

Contents

List of Tables

List of Figures

This page is intentionally left blank.

Introduction

The rise of globalization questions the relevance of local and distant debate for actors' cooperation and coordination. However, even in online settings, the actors' location influences b2b relations (Lanzolla and Frankort, 2016), cultural production suffers from physical limitations (Davidson and Poor, 2019), social ties are geographically bounded (Takhteyev, Gruzd, and Wellman, 2012), and even selling drugs in online drug crypto-markets appears to suffer from geographical constraints (Norbutas, 2018). In the end, one may ask if the world is really 'flat' as some authors expected (e.g., Friedman, 2006).

Simultaneously, the dilemma of how and why actors organize their economic activities can not be bounded solely on location. On this line, the Proximity school was fully aware that being close is not only a matter of geography. Hence, from the seminal contribution of Torre and Gilly (2000) to the structured and critical studies of Boschma (2005) and Moodysson and Jonsson (2007), more dimensions of closeness were included in the economic, management, and sociology debate of coordination and collaboration. This thesis starts from here and proposes a management-flavored analysis of actor closeness. In particular, I focus on the *influence of proximity dimensions on innovation*. As it will become apparent, scholars mostly accept this influence, while less clear is how and why that relation exists. Missing an unambiguous answer to those questions, this study's original aim has been decomposed into smaller chunks and limited its attention to one literature stream, namely management. In so doing, Chapter One introduces topic modeling, a novel methodology offered by scholars of natural language processing. Inspired by the work of DiMaggio, Nag, and Blei (2013) and in the light of Hannigan et al. (2019), the review of possible applications of this methodology to management research guides the author to analyze the enormous variety of conversations nested under the proximity and innovation debate. Hence, the latter is the focus of Chapter Two, aiming to uncover the managerial conversations in which proximity scholars took part, collect theoretical mechanisms justifying proximity influence over innovation, and unveil any empirical consistency concerning effects. Critically summarizing the proximity literature and placing it within the broader management stream lays

the foundation for Chapter Three. In this exploratory study, the author proposes the business model as the set of activities able to translate proximity into innovation.

Reading my work and the articles reviewed through the lens of Delanty and Strydom (2003) and Van de Ven and Poole (2005), both ontological and epistemological choices can be articulated. The conversation addressed by the three chapters touches an audience with a more or less clear affiliation to the positivist or logical empiricist paradigm. Within this conversation, I started my journey through social science believing in an objective reality that can be understood through science. In so doing, I applied reductionism, determinism, and I preferred collecting empirical data through structured surveys or secondary sources. From an ontological perspective, the following chapters (and mainly the third) still follow the mainstream approach to understanding change. Hence, I studied innovation as "made of things in which processes represent change in things" (Van de Ven and Poole, 2005, p. 1379). In so doing, the contribution of Chapter Three can be ascribed to variance theory more than process. While writing the thesis conclusions, I soon realized that it is the absence of diversified ontological and epistemological stances that detriment the most proximity and innovation field. As it is clear, this thesis has not dealt with these issues at all.

From these ontological and epistemological stances, methodological directions have consequently emerged. In particular, Chapter One starts with a qualitative coding based on previous knowledge on the topic, to then test emergent categories against a random sample, and finally proposing normative practice to follow. Chapter Two applies a quantitative analysis of qualitative data (textual), informed by a deductive logic in retrieving topic meanings (i.e., using keywords from the Academy of Management). Here, a great deal of attention is devoted to analysing constructs, their definition, operationalization, and the consequent p-values, or more generally effects, observed. Similarly, Chapter Three applies an exploratory quantitative analysis, informed by prior theories, once again in a deduction fashion.

More in-depth, Chapter One is inspired by the development of natural language processing in making sense of enormous corpora of unstructured textual data. In particular, topic modeling is widely recognized to offer excellent management (Hannigan et al., 2019) and social sciences opportunities in general (DiMaggio, Nag, and Blei, 2013). Reviewing recent contributions from 18 top-tier management journals, this chapter tries to promote convergence of evaluation practices and elaborate on the interdependence between authors' theoretical expectations and model evaluation choices. Clarifying these aspects, this chapter's findings offer a solid foundation for what comes next in this thesis.

Given the multivocality of proximity literature and the need to locate this research stream within the

broader management literature, topic modeling offers a great opportunity. In particular, following several outstanding exemplars such as Doldor, Wyatt, and Silvester (2019) and Sieweke and Santoni (2020), this chapter focuses on topic meanings to uncover latent managerial conversations. To do that, the topic modeling algorithm has been trained on more than 12,000 abstracts from top-tier management journals. The topics learned through this phase are then projected on a sample of selected proximity and innovation articles. Hence, this offered the chance to locate each proximity and innovation contribution within its managerial conversation. Furthermore, qualitative coding has been build to collect theoretical claims and empirical findings (when applicable). Hence, sampled articles are manually grouped based on the 'outcome' variable investigated. This process enables me to clarify both the scopes and contributions emerging.

The in-depth analysis of topics and theories at the basis of proximity literature paves the way for asking — again — how this closeness translates into value for firms. Bringing a strategy topic into the table, Chapter Three argues that firms' actions are the means through which closeness to actors may lead to innovation. In particular, the Business Model is introduced as a mediator of the proximity and innovation relation. Leveraging on crucial strategy contributions (Amit and Zott, 2015; Baden-Fuller and Haefliger, 2013; McDonald and Eisenhardt, 2020), the author disentangles possible phases of proximity and business model interactions, and consequent influences over innovation. The theoretical claim is then empirically explored on a regional sample of firms. Without any request for completeness, causality, or generalization, the research findings seem likely to open up to further academic debate on an omitted relation, that of proximity and strategy.

To make sense of the following chapters, I retrospectively reconstruct the steps I went through while working on my main research question. In particular, this thesis aims to unpack the relationship between proximity and innovation, claiming the business model as the mean through which this effect unfolds. The main novelty and contribution of this work are situated in the mediator introduced. However, two steps were to be made in order to sustain this research proposition. The first was to clarify what mechanisms have been previously proposed and studied, and with what consequences for managers and firms. This required reframing the proximity and innovation debate from the lens of management, collocating both the proximity field and its constituent articles within that broader literature landscape. The latter task necessitated the analysis of an enormous amount of scientific publications that have been made possible only by textual analysis. Approaching topic modeling, a methodology that is still emerging in our field, called for a reflection on how to set-up each step coherently with the final research goal.

Reading through this introduction, the reader may doubt of business interest toward proximity and innovation debate. While proximity literature has mainly focused on providing guidelines for local policies (at different scales, e.g., regional, national), the theoretical and empirical evidence of Chapters Two and Three makes several elements of regard emerge. In particular, through the lens of transaction cost theory or absorptive capacity, each dimension of closeness gets into the managerial conversation. Still, it also adds other critical viewpoints such as knowledge-spillovers, embeddedness, or serendipity coming from other literature streams. Once businesses fully appreciate the competitive advantage — or disadvantages — arising from these proximity endowments, closeness gains its place within the business interest. Lastly, the reader may also ask why focusing so broadly on offline settings while not considering digital advances and online platforms. While this is a significant omission here, this thesis should be seen as a work in progress directed there. Studying one side of the story opens up to great questions to be further explored.

Chapter 1

Topic modeling in management research: emerging practices and evaluation guidelines

Abstract

This literature review focuses on the use of topic modeling in management research. In particular, it aims to investigate the relationship between researchers' goals and evaluation practices. In so doing, the paper analyses 38 research articles from top-tier management journals through a structured coding scheme. Hence, studies are categorized based on their semantic interest, distinguishing four units of analysis: classification, qualitative variables, individual topics, and topology. Additionally, the author collected the evaluation practices used by each article. These practices are grouped into six larger buckets: heuristic, statistic, eyeballing, semantic, external, and assessment of the statistical model. Analyzing the relationship between the units of analysis and the evaluation techniques, the chapter highlights stratified practices by research goal. This qualitative finding is then tested against a simulated random sample with 10,000 replications. In conclusion, topic modeling guidelines are introduced and schematically summarized through a flow chart.

Keywords: topic modeling, management, research goals, evaluation practices.

1.1 Introduction

As the need to make sense of large corpora of unstructured textual data propagates, topic modeling has attracted significant attention in the organization and management community. For example, scholars have used topic modeling to investigate subjects such as diversity (Corritore, Goldberg, and Srivastava, 2020), legitimization (Croidieu and Kim, 2018), leadership development (Doldor, Wyatt, and Silvester, 2019), identity (Geva, Oestreicher-Singer, and Saar-Tsechansky, 2019), and novelty in markets (Haans, 2019; Kaplan and Vakili, 2015). The attention is also documented by literature review and methodological articles, highlighting both advantages and problems hidden behind topic detection (Hannigan et al., 2019; Kobayashi et al., 2018; Schmiedel, Müller, and Brocke, 2019). Among difficulties, the results' valuation is one of the greatest. Indeed, being an unsupervised learning method, the validation of topic modeling outcomes is less direct and requires researchers to combine different pieces of evidence (Grimmer and Stewart, 2013).

This paper takes stock of topic modeling practices emerging in the management literature, with a particular emphasis on ties between researchers' scopes and evaluation strategies. Drawing on a review of 38 articles from top tier management journals, the author highlights a series of factors that call for further attention on the part of authors and reviewers. Overall, it emerges a lack of clarity, hampering the integration of research aims, epistemological approaches, and methodological applications. This disconnectedness is likely to hinder the whole validation process, thus findings reliability and reproducibility. Moreover, a clear representation of the set of evaluation models at scholars' disposal seems missing. To cope with this complexity and lack of clarity, this research set up a qualitative coding and a simulation design to uncover emerging research scopes and evaluation practices, offering a set of guidelines to implement topic modeling.

This chapter aims to leverage management scholars' attention toward different pathways of topic modeling evaluation. Thus, three distinct contributions are provided. Firstly, I aim to promote convergence of evaluation practices — in so doing, I also aim at clarifying the expectations authors and reviewers should have about what constitutes a 'good' topic modeling. Secondly, I elaborate on the interdependencies between the theoretical expectations of authors and model evaluation choices. Thirdly, I create momentum around the reproducibility of topic modeling in organization and management research.

The chapter is structured as follows. In the next sections, a concise review of topic modeling and evaluation strategies briefly introduces the reader to the problem at hand. Then, the methodology is discussed, providing insights behind the articles' search and selection. Consequently, emerging practices are identified and thoroughly analyzed. Next, I introduce a decision tree linking research scopes to findings evaluation.

Finally, I recap the key messages and conclude.

1.2 Literature Review

Born in the community of computer science, topic modeling algorithms were designed to infer hidden topical structures from unstructured textual data (Blei, Ng, and Jordan, 2003). In so doing, Topic Modeling enables humans to appreciate and organizing large corpus of text (Blei, Ng, and Jordan, 2003). Such cross-disciplinary applicability appealed scholars from different fields to apply topic modeling with different goals (e.g., Hall, Jurafsky, and Manning, 2008; Quinn et al., 2010). Once the researchers are provided with topic modeling results, so probabilities linking words to topics and topics to documents, they can "use this information to address the analytic questions that motivated the research" (DiMaggio, Nag, and Blei, 2013, p. 586). However, authors need to validate topic modeling solutions to sustain any further claim (Grimmer and Stewart, 2013). The subtle link between researchers' goals and evaluation of the solutions is particularly evident in studies applying topic modeling for classification purposes. Indeed, for some "extrinsic tasks, such as information retrieval or document classification" (Wallach, Murray, Salakhutdinov, and Mimno, 2009, p. 1), authors can directly assess the performance looking at the ability of topic modeling to perform the task at hand, like classifying Twitter users and their posts in different categories of interest (Hong and Davison, 2010, p. 5). In other cases, in which the aim is to build some qualitative variables, such as diversity or novelty (Haans, 2019; Kaplan and Vakili, 2015), or to describe meanings of topics (Banks et al., 2019; Doldor, Wyatt, and Silvester, 2019), the validation is not so straightforward. Researchers are required to piggyback on different strategies (DiMaggio, Nag, and Blei, 2013; Grimmer and Stewart, 2013; Quinn et al., 2010). The following paragraphs introduce the topic modeling technique and some common evaluation practices.

1.2.1 What topic modeling is

This paragraph introduces topic modeling's main features, referring the reader to previous studies for more detailed descriptions (e.g. Blei, 2012; Blei, Ng, and Jordan, 2003; DiMaggio, Nag, and Blei, 2013; Hannigan et al., 2019). Topic modeling consists of a series of algorithms (e.g., Latent Dirichlet Allocation, Hierarchical Dirichlet Process) to deduce hidden semantic structures (i.e., meanings). These algorithms are particularly useful because they allow the researcher to analyze vast amounts of textual data, which is impossible for a human coder. In short, topic modeling describes each document (e.g., an abstract, article, tweet) as a set of N topics, and each topic as a set of K words. Probabilities describe topic-document and

word-topic relationships. The higher the probability, the more critical the topic (word) is in defining the document (topic). In this review, the focus is on the algorithm called Latent Dirichlet Allocation, a simple but also versatile one (Blei, 2012).

More specifically, the LDA algorithm assumes that the set of documents to be analyzed (corpus) is made up of a group of topics. Also, each topic is defined by a distribution of words [1]. The first difficulty that the researcher encounters lies in deciding "a priori" the number of topics to be searched (e.g., 50, 100, 500). Once you have selected which number to consider, the algorithm generates two results. The first consists in linking each topic (none excluded) to each document through posterior probabilities. The topic(s) with the highest probability is (are) also the one that best describes the text's content. The second links each topic to the distribution of words, again in the form of posterior probabilities. The words with the highest probability best describe the topic's meaning (usually the first ten are used). This whole process helps the researcher to code the texts he/she is analyzing.

Topic modeling is gaining popularity in the social sciences, but its use requires the reconciliation of two different perspectives, such as computer and social scientists (DiMaggio, 2015). The social sciences are attracted by the idea of having a mechanism that, acting as a "super partes", can justify the researchers' theoretical claims. On the contrary, computer scientists, more doubtful in machines' ability to replicate human analysis, test such models against human coders. Divergent positions which are nonetheless reconciled by the concerns related to the evaluation of results obtained through an Unsupervised Learning algorithm such as LDA. A need to validate findigns that is even stronger when is the researcher to establish the number of topics to be considered, a priori. Despite the intricate path toward validation, topic modeling remains an attractive technique given it being explicit, automated, inductive, and able to recognize the relationality of meanings (DiMaggio, Nag, and Blei, 2013).

As documented by Hannigan et al. (2019), topic modeling is always gaining more attention within management literature. Analyzing papers using this technique, the authors identified management literature subsets to which topic modeling has been applied and consequently described the process from unstructured textual data to theorizing. In particular, they distinguished among rendering corpora, rendering topics, and rendering theoretical artifacts offering a map of the conceptual and practical steps the researcher should go through. In this chapter, I focus on what connects rendering topics, id est the identification of *appropriate*

[1] To simplify the discussion, all the tasks necessary to prepare the corpus for later analysis are omitted. In general, through other natural language processing algorithms, the text is subjected to tokenization, lemmatization (or stemming), and stop-words removal.

topics, and rendering theoretical artifacts, that is the topics' relation with theory. In particular, I claim that appropriateness is contextual to the aims guiding the researcher to apply topic modeling.

1.2.2 Evaluation Practices

The evaluation of topic modeling has been a matter of great interest among scholars from different research fields (DiMaggio, Nag, and Blei, 2013; Grimmer and Stewart, 2013; Wallach, Murray, Salakhutdinov, and Mimno, 2009). Many evaluation strategies are at scholars' disposal, and none of them is exclusive but complementary. Indeed, the validation of an unsupervised method, such as topic modeling, requires researchers to "combine experimental, substantive, and statistical evidence" (Grimmer and Stewart, 2013, p. 271) to sustain their claims. The following discussion takes stock of some key contributions influencing management scholars on the crucial task of topic modeling evaluation. Thus, five practices are discussed: heuristic, statistical, eyeballing, semantic, external, and assessment of the statistical model.

Following a heuristic approach, several authors are inclined to train the model on a standard topic threshold, such as 100 (e.g., Blei and Lafferty, 2007; Hall, Jurafsky, and Manning, 2008). Some researchers rationalize this approach as a convenience practice or as a direct consequence of their research interest. However, most times, the 100 threshold is similar to a golden rule that nobody questions.

Statistical evaluation is a practice that aims to validate and lower the complexity of model selection (Wallach, Murray, Salakhutdinov, and Mimno, 2009), practically speaking, of setting the number of topics. For example, Blei, Ng, and Jordan (2003) initially assessed topic modeling performance through a conventional language modeling indicator, the perplexity score. More systematically, Wallach, Murray, Salakhutdinov, and Mimno (2009) examined several statistical methods for topic modeling evaluation in the search for a universal routine. The interest in statistical evaluation caused the rise of several complementary indicators (e.g. Arun, Suresh, Madhavan, and Narasimha Murty, 2010; Cao et al., 2009; Mimno, Wallach, Talley, Leenders, and Mccallum, 2011), see Appendix A.2 for further details. Applying statistical evaluation, scholars can follow different paths, employing a single statistical measure or comparing the outcome of more than one (e.g Lappas, Sabnis, and Valkanas, 2016).

For what concerns eyeballing practices, we can distinguish among keywords per topic inspection and visualization. In both cases, the aim is to assess the quality of the topic model (e.g Chuang, Manning, and Heer, 2012) and to "aid the user in interpreting individual topics" (Sievert and Shirley, 2014, p. 52). Experts go through a list of high probability words (keywords) for each topic in the first case (e.g. Atkins et al., 2012;

Mimno, Wallach, Talley, Leenders, and Mccallum, 2011). In the second case, visualization techniques such as Termite (Chuang, Manning, and Heer, 2012) or LDAvis (Sievert and Shirley, 2014) offer the analyst some more information such as terms salience or relevance and intuitive visualization of the semantic space.

For what concerns semantic evaluation, the aim is to ensure topics to discriminate between different meanings, thus interpretability (DiMaggio, Nag, and Blei, 2013; Grimmer and Stewart, 2013; Quinn et al., 2010). From a quantitative fashion, Chang et al. (2009) describes word and topic intrusion as measures of semantic meaning. In both cases, a human evaluator is required to select the 'intruder' word (topic) among a list composed of high probability words (topics) and a randomly included 'intruder' for a specific topic (document). The task results are then used to generate quantitative indicators (see Appendix A.2). From a more qualitative perspective, other evaluative paths are at the researcher's disposal. For example, digging into terms' meanings, the algorithm assignment of a single word to different topics should be humanly comprehensible. Thus, the analyst may try to appreciate the polysemy of words affiliated with multiple topics (DiMaggio, Nag, and Blei, 2013). Furthermore, researchers can verify topics quality inspecting those documents in which a specific topic shows a high probability. Authors may also require a human coder (often an expert) to assess the goodness of topic modeling results.

A further evaluation practice concerns comparing variation in topic usage to some exogenous event, without any claim on the relationship's direction (Quinn et al., 2010). Indeed, some external event (not related to the measurement process) should be associated with an increase in some topic attention more than others (DiMaggio, Nag, and Blei, 2013; Grimmer and Stewart, 2013). For example, Quinn et al. (2010) showed the increase of words included in those topics concerning symbolic speeches in support of the military and other public servants immediately after 9/11.

Lastly, topic modeling can also be evaluated assessing the statistical model's ability to perform some extrinsic tasks, such as classification (Wallach, Murray, Salakhutdinov, and Mimno, 2009). For example, Hong and Davison (2010) set up two classification tasks, the first aiming to classify messages among tweets and retweets, the second to group users and messages into topical categories. In the first case, they employed three evaluation metrics, namely, Precision, Recall, and F-measure. In the second case, authors rely simply on an accuracy metric. These and other classification metrics are discussed in Appendix A.2.

1.3 Methodology

Our literature review is built on contributions from 12 top management journals. The articles' retrieval process has been performed within Scopus and ISI web of knowledge, and further complemented and validated within Google Scholar and each Journal website. Firstly, I selected a list of keywords and a cluster of 18 journals [2]. Then, I performed two queries: (i) one within Scopus looking at articles' title, abstract, and keywords; (ii) the other within ISI web of knowledge among all fields [3]. In so doing, I obtained 33 unique research articles. Looking at articles' content, I excluded studies just citing topic modeling as a matter of reference (Moe and Schweidel, 2017) or not related to topic modeling (Johnson, Safadi, and Faraj, 2015). Thus, I discounted nine studies. Besides, I performed a Google Scholar and Journals' websites search with identical parameters but considering the whole articles' text. I discarded articles briefly mentioning topic modeling (Miric, Boudreau, and Jeppesen, 2019) or citing it as reference or future research advance (Ocasio, Laamanen, and Vaara, 2018). This led to the inclusion of fifteen articles. During the analysis phase, I discarded a further study not clear on its topic modeling aim. The final set of 38 articles has a timespan ranging from 2013 to 2019, with two articles still in press (Corritore, Goldberg, and Srivastava, 2020; Sieweke and Santoni, 2020). Figure 1.1 describes the temporal distribution of retrieved articles. Overall, growing attention comes to light. The most active journals on this topic are MIS Quarterly and Information Systems Research, but awareness is rising among all management streams (see Table A.1 in the Appendix).

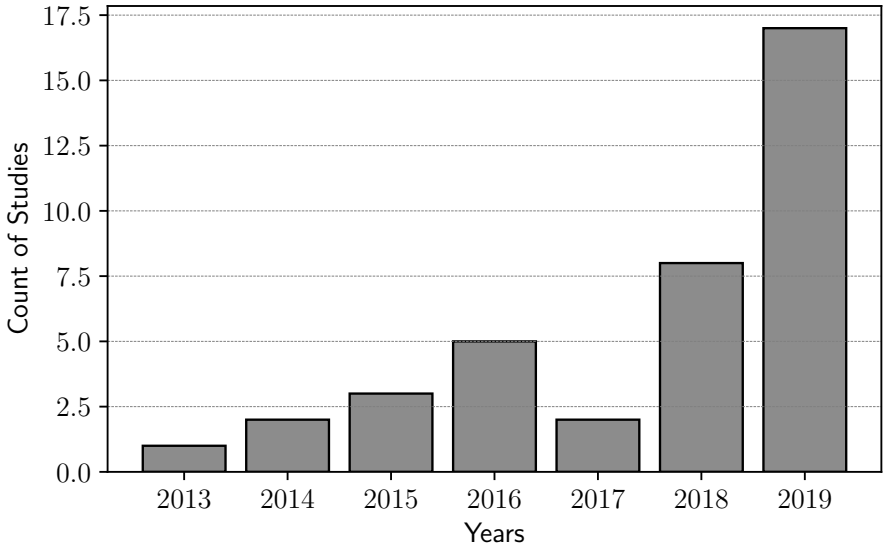
1.3.1 Coding Scheme

Through an iterative process, a coding scheme was outlined to analyze the retrieved articles. Firstly, the author went through five articles and generated meticulous annotations. Information was collected concerning study features, dataset, textual pre-processing, estimation, and evaluation practices in this phase. Next, an in-depth analysis of the dimensions and variables worth to be investigated has been performed. This led to an updated coding structure that has been again tested on a set of five articles to check for robustness. Consequently, the whole set of retrieved articles were codified. In this phase, new variables' items were detected and considered, thus enlarging the initial items' collection. The whole corpus examination leads to a final coding scheme where two main domains have been identified: goals and evaluation practices

[2] Keywords: "topic modeling", "topic model*", "natural language processing", "nlp", "latent dirichlet", "LDA". Journals: Academy of Management Journal, Administrative Science Quarterly, Entrepreneurship Theory and Practice, Industrial and Corporate Change, Information Systems Research, Journal of Business Venturing, Journal of Management, Journal of Management Studies, Journal of Product Innovation Management, Leadership Quarterly, Management Science, MIS Quarterly, Organization Science, Organization Studies, Research Policy, Strategic Entrepreneurship Journal, Strategic Management Journal, Strategic Organization.

[3] Search performed on January 2nd, 2020. See Appendix A.1 for further detail on search queries.

Figure 1.1: Number of Articles per Year



Notes. The year 2019 includes two articles In-press by January 2020. At this date, the total population of retrieved articles amounts to 38.

(see Table 1.1).

Goals concern the reasons moving researchers to adopt topic modeling. Here, a first distinction comes with 'substantial semantic interest' (value 1) or its absence (value 0). In the first case, the researcher investigates either the single topics (category 2) or the topology (category 3). While, authors without any 'substantial semantic interest' are usually concerned with classification tasks (category 0) or crafting qualitative variables (category 1).

For what concerns evaluation practices, the coding scheme distinguishes between heuristic, statistical, eyeballing, semantic, external, and assessment of the statistical model. Each of these dimensions and their subdimensions was operationalized as a dummy variable, where 0 applies to studies not employing a specific practice and 1 otherwise. No practice is exclusive with respect to others. I coded as heuristic those studies fixing an a priori number of topics (e.g., 100, 200, 500), usually relying on crystallized practices (Chang et al., 2009; Hall, Jurafsky, and Manning, 2008). Similarly, I categorized as statistical all research providing metrics to inform and justify the number of selected topics. This evaluation category is populated by several measures described in Appendix A.2. Furthermore, I searched for eye-balling practices such as scrutinizing words with high probability (Mimno, Wallach, Talley, Leenders, and Mccallum, 2011) or visualizing topics

Table 1.1: Coding Scheme

Domain	Variable	Synopsis
Goals	Substantial Semantic Interest	[0 = No; 1 = Yes]
	Unit of Analysis	[0 = Classification; 1 = Qualitative Variables; 2 = Individual Topics; 3 = Topology]
Evaluation Practices	Heuristic	[0=No; 1=Yes]
	Statistical	[0=No; 1=Yes] Arun et al. 2010 [0=No; 1=Yes] Cao et al. 2009 [0=No; 1=Yes] Deveud et al. 2014 [0=No; 1=Yes] Dispersion of Residuals [0=No; 1=Yes] Document-completion Held-out likelihood [0 = No; 1 = Yes] Frequency and Exclusivity - FREX [0 = No; 1 = Yes] Griffiths and Steyvers 2004 [0=No; 1=Yes] Perplexity [0=No; 1=Yes] Semantic Coherence [0=No; 1=Yes] Silhoutte Coefficient [0=No; 1=Yes]
	Eyeballing	[0=No; 1=Yes] Keywords Inspection [0=No; 1=Yes] Visual Inspection [0=No; 1=Yes]
	Semantic	[0 = No; 1 = Yes] Word Intrusion [0=No; 1=Yes] Topic Intrusion [0=No; 1=Yes] Polysemy Inspection [0=No; 1=Yes] Topic to Document Inspection [0=No; 1=Yes] Human Coder Agreement [0=No; 1=Yes]
	External Assessment of the statistical model	[0=No; 1=Yes] [0 = No; 1 = Yes] Accuracy [0=No; 1=Yes] Area Under the ROC Curve [0=No; 1=Yes] F-measure (or F_1-score) [0 = No; 1 = Yes] Precision [0 = No; 1 = Yes] Recall [0 = No; 1 = Yes]

and words to topics distributions (Chuang, Manning, and Heer, 2012; Sievert and Shirley, 2014). Hence, I coded as semantic those studies approaching this evaluation practice both quantitatively or qualitatively: word and topic intrusion offer quantitative indicators of goodness (Chang et al., 2009); polysemy, topic to document inspection, and human coder agreement represents qualitative alternatives or complements. Then, I coded as external those articles providing comparisons between topic probabilities and exogenous events (Quinn et al., 2010). Lastly, those researches that evaluate topic modeling looking at task's effectiveness were coded as assessing the statistical model. This evaluation practice is populated by several metrics described in Appendix A.2.

1.4 Semantic Interest

Authors can approach topic modeling with or without 'substantial semantic interest'. To have a 'substantial semantic interest' means to possess some theoretical concern regarding the semantic space. This goal can be pursued by analyzing single topics' meanings or studying the topology, thus explaining relations among topics in time or space. Authors not possessing any theoretical goal are interested in classifications or in crafting qualitative variables. In so doing, authors perceive the semantic space as instrumental and are not concerned with appreciating meanings, relations or variations.

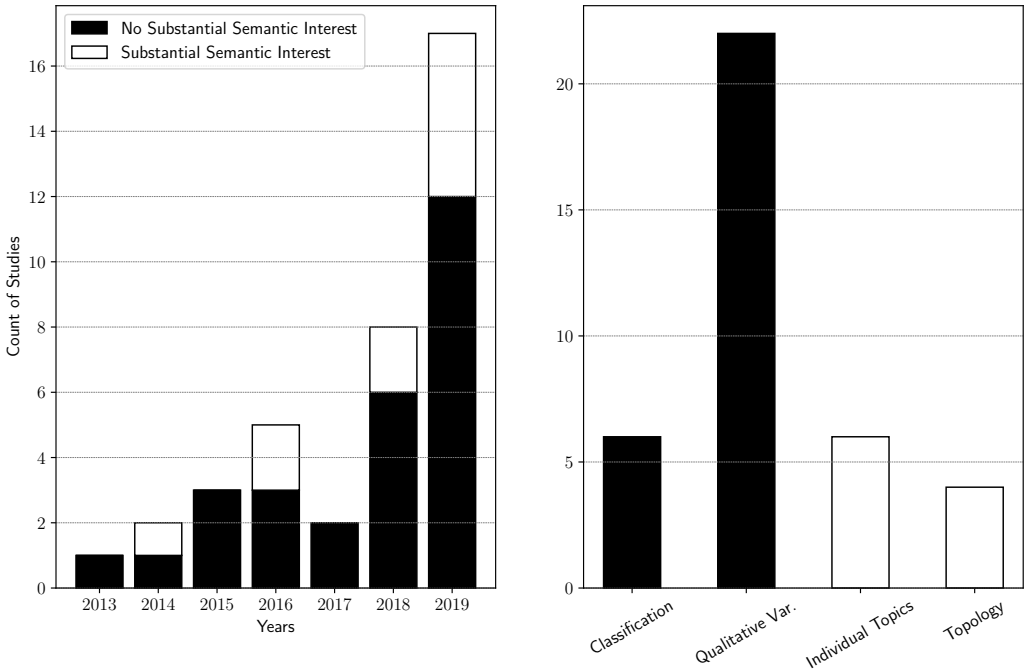
Figure 1.2 describes the change of semantic interest by years. In general, management research is largely populated by studies without substantial semantic interest. The wide majority of these articles approach topic modeling to craft qualitative variables. Nonetheless, growing attention seems to raise around meanings, thus, for a substantial semantic interest. In particular, authors show concern for both individual topics and topology.

1.4.1 Without Substantial Semantic Interest

Six studies employed topic modeling to perform Classification tasks within the sampled articles, while twenty-two to craft Qualitative Variables. The community of information systems scholars mainly populates both groups. Nonetheless, many strategy and organization authors show interest in the second unit of analysis.

Articles concerned with Classification employ topic modeling to organize documents in homogeneous groups. In particular, management scholars use topic modeling to classify patents (Ruckman and McCarthy, 2017), apps (Wang, Li, and Singh, 2018), conversations (Abbasi, Zhou, Deng, and Zhang, 2018), and

Figure 1.2: Semantic interest, number of articles per year



Notes. The year 2019 includes two articles In-press by January 2020. At this date, the total population of retrieved articles amounts to 38. The count of articles by unit of analysis show: 6 Classification, 22 Qualitative Variables, 6 Individual Topics, and 4 Topology.

detect constructs fallacy (Larsen and Bong, 2016). For example, Ruckman and McCarthy (2017) aimed to compare licensed patents to not licensed alternatives. To this end, the authors applied topic modeling matching these two groups based on their similarity. The core idea is to create a set of alternative patents that could have been licensed, but — in fact — have not. In so doing, authors were able to better control for patents' characteristics, and uncover the effect of licensors. Another example is provided by Larsen and Bong (2016), who employed topic modeling (and other natural language processing techniques) to investigate behavioral constructs overlaps. In particular, their goal was to find constructs that reference similar phenomena, hence grouping them based on 'construct identity'. Similarly, Wang, Li, and Singh (2018) employed topic modeling as a baseline comparison algorithm in the task of app classification. Here, the aim was to match the original mobile application with their copycats.

Articles concerned with crafting Qualitative variables employ posterior probabilities to create infra-topic or infra-sample dissimilarity measures or to be included as regressors into the econometric analysis. In so doing, scholars studied novelty (Kaplan and Vakili, 2015), distinctiveness (Haans, 2019), rhetorical signals (Antons, Joshi, and Salge, 2019), and culture (Corritore, Goldberg, and Srivastava, 2020). For example, Choudhury, Wang, Carlson, and Khanna (2019) investigate the tendency of leaders to vary thematic focus. To operationalize this tendency, the authors applied topic modeling on a corpus of leaders' interviews to obtain the hidden semantic structure. Then, computed an entropy measure for each interview to grasp its heterogeneity regarding the topic discussed. Corritore, Goldberg, and Srivastava (2020) offer another great example. In particular, the authors aimed at capturing cultural heterogeneity, crafting two variables, intrapersonal and interpersonal cultural heterogeneity. In so doing, they leveraged on posterior probabilities obtained training the topic modeling algorithm on employees' reviews, to then calculate the Jensen–Shannon divergence and Herfindahl index. Besides, Bapna, Benner, and Qiu (2019) provide an example for the inclusion of posterior probabilities within the econometric model. This research aims to understand how companies nurture online communities, looking at what firms publish. In so doing, authors employed topic modeling to control latent topics associated with engagement. Hence, they trained the LDA algorithm on posts and included the topics obtained as regressors.

1.4.2 With Substantial Semantic Interest

Ten studies show a substantial semantic interest. Of these, six studied meanings conveyed by single topics, while four embarked on the study of typology. Among this set of articles, information systems scholars' presence drastically decreases (a single study); thus, diversified management interests emerge (e.g. leader-

ship, organization, innovation).

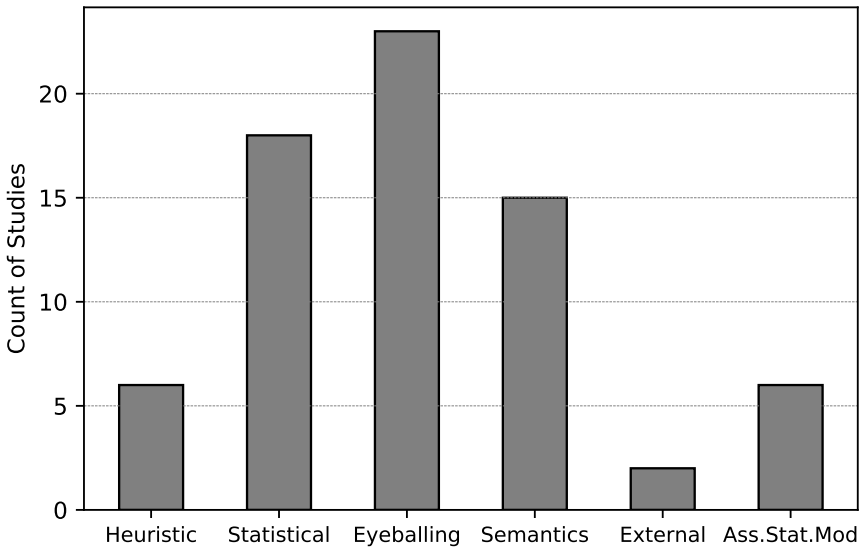
Authors approaching Individual Topics are interested in uncovering gender differences (Doldor, Wyatt, and Silvester, 2019; Wu, 2013), analyzing research fields (Sieweke and Santoni, 2020), discovering financial risks (Bao and Datta, 2014), capturing firms' recruiting signals (Banks et al., 2019), and examining online responses (Lappas, Sabnis, and Valkanas, 2016). For example, Doldor, Wyatt, and Silvester (2019) aimed to reveal gender differences in leadership development. Hence, analyzing the text of leaders' developmental comments, they employed topic modeling to derive 12 first-order topic summaries. Then, scrutinizing each topic in-depth, authors identified second-order messages by gender, revealing a sensible feedbacks' heterogeneity. Analogously Wullum Nielsen and Börjeson (2019) studied research outcomes in relation to the authors' gender. In so doing, authors leveraged on topic modeling to identify topics of interest for 25,000 articles, to later compare with gender concentration. Another interesting application comes from Banks et al. (2019). The authors' goal is to understand domestic and international recruiting signals of multinational enterprises. To do so, they analyzed companies' website data through topic modeling, uncovering hidden topics; hence the sought recruiting signals.

Articles with interest in topology attempt to appreciate meanings' relationships and evolution. Here, the authors' interest span from literature stream evolution (Antons, Kleer, and Salge, 2016; Hopp, Antons, Kaminski, and Salge, 2018), to legitimation mechanisms (Croidieu and Kim, 2018), and changes in the cultural context (Giorgi, Maoret, and J. Zajac, 2019). As an example, Croidieu and Kim (2018) studied mechanisms of legitimation in the radio movement of US amateurs. In this case, topic modeling informed researchers' effort towards theorizing. Indeed, raw topics were analyzed and classified as the first-order construct and further grouped in second-order ones. Then, exploring the historical path of these second-order themes, authors could reveal four mechanisms of legitimation. Similarly, Hopp, Antons, Kaminski, and Salge (2018) analyzed journal articles uncovering a set of themes able to give an insightful topical map of disruption research. In so doing, they individuated two disconnected components in the topic network and offered an in-depth analysis of the evolutionary trajectories of each topic both individually and in relation to others. Giorgi, Maoret, and J. Zajac (2019) provide another compelling example. Here, the authors use topic modeling to study the emergence and diffusion of meanings in the automobile field.

1.5 Evaluation Practices by Unit of Analysis

In this section, evaluation practices are discussed by the unit of analysis (see Table A.2 in the Appendix for further details). Overall, Figure 1.3 offers details on each evaluation approach popularity within management research. In particular, management scholars largely rely on Eyeballing (60,5%), Statistical (47,4%) and Semantic (39,5%) practices. A smaller portion of our sample evaluated topic modeling using the other three approaches (Heuristic 15,8%, Assessment of the Statistical Model 15,8%, and External 5,3%). For the following paragraphs, Figures 1.4 and 1.5 summarize several information such as practices popularity and their combination by the unit of analysis (see figures description).

Figure 1.3: Evaluation practices within management research



1.5.1 No Substantial Semantic Interest: Classification

As shown in Figure 1.4, articles concerned with Classification follow similar evaluation paths. Indeed, all articles in this category chose practices of the 'Assessment of the Statistical Model' group. For example, Ruckman and McCarthy (2017) experiment three models with different numbers of topics and choose the best one based on performance: "The 500 topic model performed well on random checks of the patents" (p. 22). More rigorously, Larsen and Bong (2016) created a 'gold standard' dataset annotated by experts that "is used as the "solution" against which the design performance is assessed" (p. 563). Then, they compared the obtained results focusing on the false positive and false negative. They further compute some typical

measures, namely precision, recall, and F-measure. A similar evaluation path is followed by Wang, Li, and Singh (2018), Abbasi, Zhou, Deng, and Zhang (2018), Guo et al. (2017), and He, Fang, Liu, and Li (2019). Noteworthy, this group is the sole to chose 'Assessment of the Statistical Model' techniques to evaluate topic modeling.

1.5.2 No Substantial Semantic Interest: Qualitative Variables

As Figure 1.4 highlights, authors that aim to craft 'Qualitative Variables', embrace different evaluation strategies [4]. Eyeballing, and in particular keywords inspection, is the most common one. For example, Huang, Lehavy, Zang, and Zheng (2018) provided a table showing the top 20 words for the top 10 topics retrieved, and labeled each topic to "validate that LDA is able to discern the underlying economic content of the topics" (p. 2836). Similarly, Corritore, Goldberg, and Srivastava (2020) validated their topics showing the highest-weighted words for a topics' subsample. A less employed eyeballing technique is the visual inspection. A good example is offered by Gong, Abhishek, and Li (2018) that evaluate the obtained topics through two visualizations. The first (p. 814) is a network of the top 10 words for each topic that shows topic membership and links among topics based on shared words. The second (p. A3) depicts a sample of keywords and topics, where bubbles represent the posterior probability of a word to be part of a topic; the higher the bubble size higher the probability.

The second strategy for popularity is statistical evaluation. Five indicators are employed and here presented in decreasing order of usage: Griffiths and Steyvers (2004), Arun, Suresh, Madhavan, and Narasimha Murty (2010), Cao et al. (2009), Deveaud, SanJuan, and Bellot (2014), and Perplexity. The latter measure is used by both Huang, Lehavy, Zang, and Zheng (2018) and Yue, Wang, and Hui (2019) to identify and validate the optimal number of topics. Often, authors prefer to combine the results of these indicators to provide a more robust evaluation. For example, Choudhury, Wang, Carlson, and Khanna (2019) identified the optimal number of topics "triangulating across several different measures of model fit" (p.19), using all the first four indicators cited above. Analogously Geva, Oestreicher-Singer, and Saar-Tsechansky (2019) collapsed the same four metrics into a single sensitivity score to find the optimal number of topics.

The third most employed evaluation strategy is the semantic one, in particular topic to document inspection. For example, Haans (2019) validated the obtained topics and their ability to capture homogeneity and diversity, by returning "to the website of several highly average firms, finding them to indeed be very

[4]For completeness, Huang, Lehavy, Zang, and Zheng (2018, p. 2836) is the sole study leveraging on external evaluation; therefore, it is not discussed here.

similar" (p. 14). The author provides a series of examples corroborating the ability of topic modeling "to model the (in)distinctiveness" (p. 15). For what concerns other semantic evaluation practices, Huang, Lehavy, Zang, and Zheng (2018) offer a good example of human coder agreement and polysemy inspection. For what concerns the former, authors "compare the LDA's topic assignment to that of a human coder" (p. 2836), thus analyzing differences among automated topic assignment and manual assignment. In the case of polysemy, the author shows that the word price "is related to both "valuation" and "raw materials and input price" in the capital good industry" (p. 2836); thus, providing evidence for the ability of topic modeling to disentangle multiple meanings.

Following the example of Kaplan and Vakili (2015), five articles embrace a heuristic approach. Indeed, the former authors pointed out some criticisms towards best-fit models (statistical evaluation), therefore choosing to "constrain the number of topics" (p. 1442) to 100. In so doing, they followed the guidance of some computer science studies such as Blei and Lafferty (2007) and Hall, Jurafsky, and Manning (2008). In partial compliance with this approach, Haans (2019) "set the number of topics to 100 – balancing between having too many topics to be interpretable and too few to allow meaningful variation" (p. 13). On the other hand, Corritore, Goldberg, and Srivastava (2020) employed heuristics but from a different angle. In this case, the authors' goal is "not to maximize the coherence or distinctiveness of the topics" as they are interested only "in the distribution of content between and within reviewers" (pp. 14-15). Thus, they run topic modeling with a larger number of topics, 500.

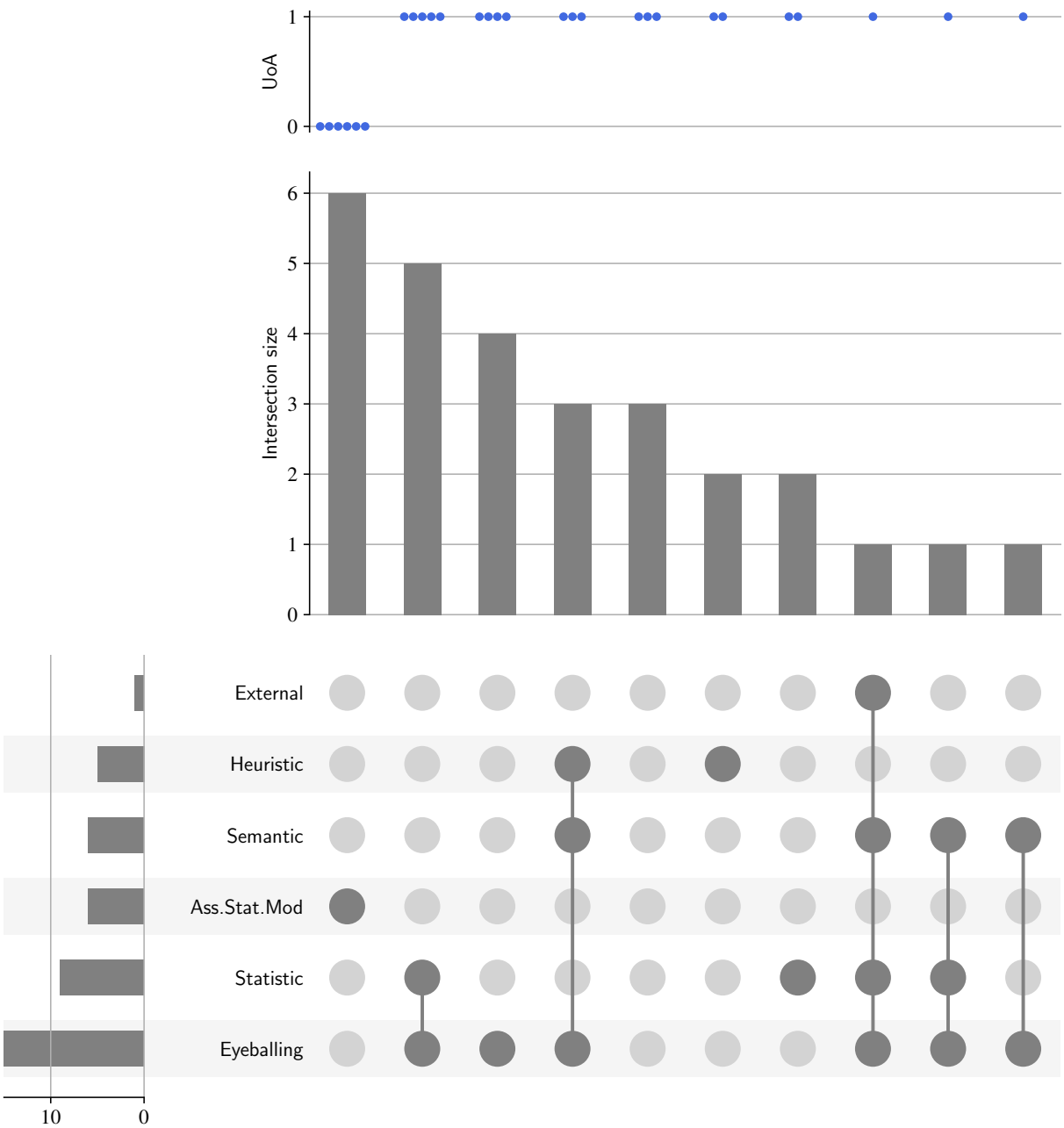
These techniques are not mutually exclusive and, as shown in Figure 1.4, eleven studies combined two or more practices. In particular, eyeballing and statistic evaluation is the most common dyad with seven occurrences. In five cases, eyeballing is also related to semantic evaluation.

1.5.3 Substantial Semantic Interest: Individual Topics

Figure 1.5 highlights that articles concerned with studying individual topics have a preference for semantic, statistical, and eyeballing practices. Noteworthy, the heuristic approach does not receive any attention among these articles, while Bao and Datta (2014) is the sole study leveraging on external evaluation.

The most employed semantic validation practice is topic to document inspection. Banks et al. (2019) offer a great example. Indeed, to "confirm that the topics emerging from the text were robust and represent the underlying data", they "compared the text data to the emerging topics as well as back to the existing literature on the topic" (p. 486). Authors completed this evaluation focusing on topics similarity, dissimilar-

Figure 1.4: Evaluation practices of articles without semantic interest



Notes: the upset chart (Lex et al., 2014) collects the following information. Full dots connected by lines characterize the set of evaluation practices used. The horizontal bar chart shows the frequency for each evaluation practice, the vertical bar chart ('Intersection size') describes the count of studies interested in a particular set of practices, and the scatter plot ('UoA') provides information on the unit of analysis of each study interested in that set ('0': Classification; '1': Qualitative Variables)

ity, and reasons for dissimilarity. Another example is provided by Doldor, Wyatt, and Silvester (2019) that validated topic modeling results analyzing the "65 most representative feedback comments for each topic (top 5%)" (p. 6). Another semantic practice employed is to ask a human coder to evaluate the results of topic modeling. For example, Bao and Datta (2014) asked two human annotators to go through topic modeling results in order to assess the goodness of retrieved topics. Similarly, Banks et al. (2019) use two experts to revise their results. Lastly, both Bao and Datta (2014) and Doldor, Wyatt, and Silvester (2019) employ word intrusion, thus providing quantitative metrics to show the human ability to select the intruder word.

Statistical and eyeballing practices show a similar occurrence in this cohort of articles. For what concerns the former, six indicators are employed and here presented in decreasing order of usage: perplexity, semantic coherence, dispersion of residuals, frequency and exclusivity, document-completion held-out likelihood, and silhouette coefficient. The majority of articles employ a single measure of fit, in particular perplexity (Lappas, Sabnis, and Valkanas, 2016; Wullum Nielsen and Börjeson, 2019) and semantic coherence (Sieweke and Santoni, 2020). On the contrary, Doldor, Wyatt, and Silvester (2019) compare different indicators (semantic coherence, dispersion of residuals, frequency and exclusivity, and document-completion held-out likelihood) to select the number of topics, while Bao and Datta (2014) choose the number of topics through perplexity and further validated cluster quality on the basis of silhouette coefficient metric comparing different topic modeling alternatives. For what concerns eyeballing, Sieweke and Santoni (2020) provide an example of both keyword and visual inspection. Indeed, they report a term-topic matrix, showing the top ten words per topic, and a dynamic visualization, displaying inter-topic distances and top-30 most frequent and salient terms per topic.

As shown in Figure 1.5, all articles in this group combine multiple evaluation techniques to assess the goodness of topic modeling results. In particular, semantic and statistic evaluations are the most common dyad, occurring in five cases out of six. Often, eyeballing is added to the former evaluation practices.

1.5.4 Substantial Semantic Interest: Topology

Similarly to articles with interest in individual topics, Figure 1.5 highlights that topology studies still embrace semantic, statistical, and eyeballing practices. In so doing, all four articles go through statistical evaluation. In particular, Antons, Kleer, and Salge (2016), Croidieu and Kim (2018), and Hopp, Antons, Kaminski, and Salge (2018) rely on the Griffiths and Steyvers (2004) metric. Alternatively, Giorgi, Maoret, and J. Zajac (2019) compared results of Arun, Suresh, Madhavan, and Narasimha Murty (2010) and Deveaud, SanJuan, and Bellot (2014) metrics "to check for the correctness of the assumed range of topics

chosen" (p.825). Additionally, eyeballing and semantic evaluation practices are largely diffused. For example, Antons, Kleer, and Salge (2016) combine both. Hence, they firstly went through "the top 15 terms per each topic together with the title, abstract, keywords, and author information of the five highest loading articles per topic" (p.732). Then, the same information is given to a pool of 14 academic researchers and disagreements are resolved through discussion. Also in this cohort, Figure 1.5 shows that all articles combine different practices. Indeed, statistic is often used together with semantic and eyeballing.

1.6 Stratification in Evaluation practices

This section aims to deepen the relationship between units of analysis and evaluation practices. In particular, the goal is to investigate whether the evaluation choice is lead by popularity or by the emergence of crystallized practices by unit of analysis [5]. In so doing, the empirical distribution is compared with randomly simulated data with two aims. First, I want to assess whether evaluation in each unit of analysis is concentrated among a smaller set of practices than what expected under random assignment. Second, I investigate whether any evaluation practice has higher chances to occur within a unit of analysis than what expected with random choice.

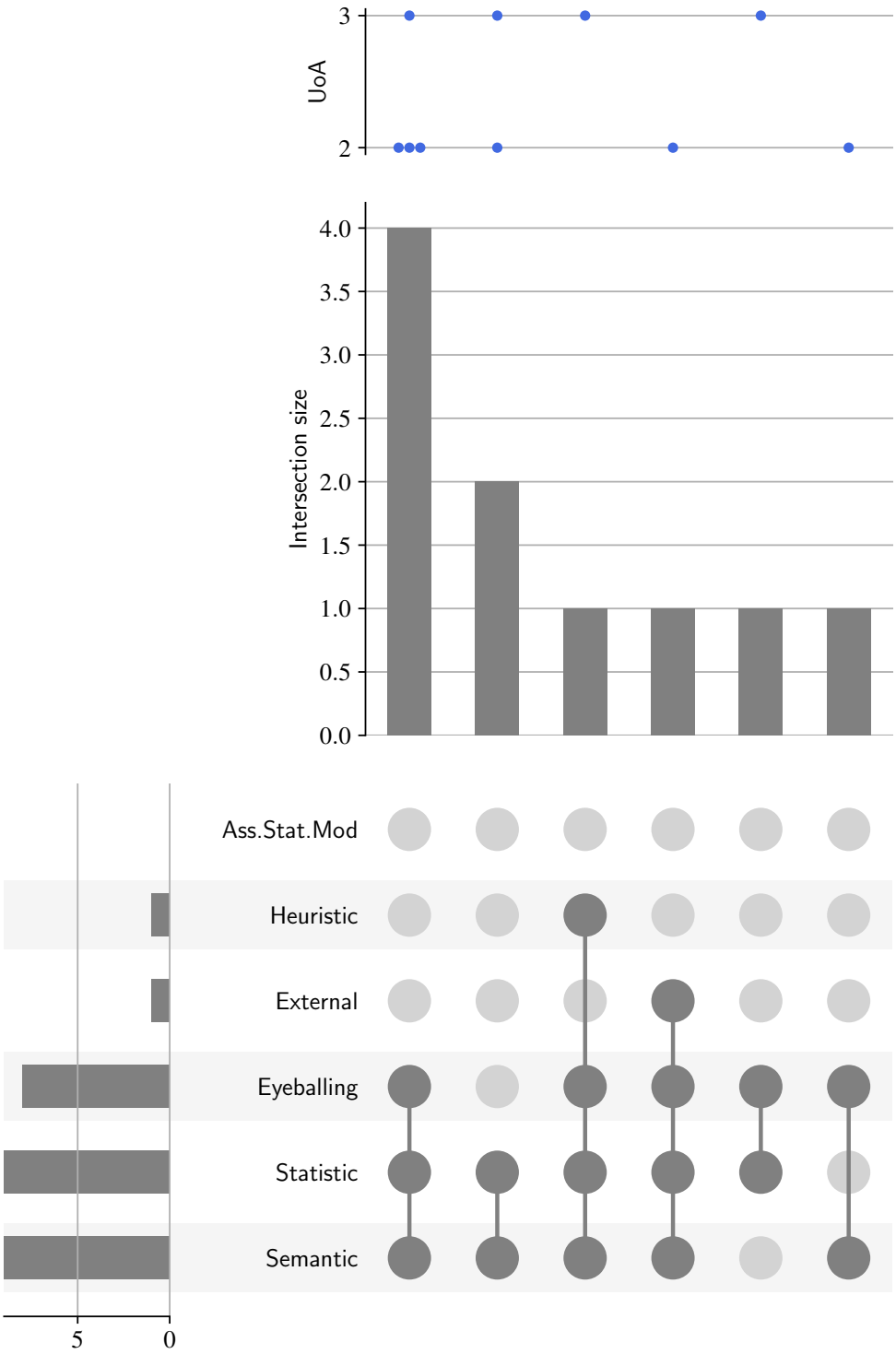
To perform these analyses, the author went through the following steps. First, the independent probabilities of occurrence for both units of analysis and evaluation practices were calculated over the empirical sample. Then, 38 articles and the relative choices were simulated, taking into account the empirical probabilities. In this way, the real popularity of each scope and practice is considered. At this point, the mean of evaluation practices employment by the unit of analysis are generated (obtaining a matrix 4×6). Then, a Herfindahl score is calculated per each unit of analysis:

$$H = \sum_{e} \left(\frac{\sum_{A}^{a} D_{ea}}{E}\right)^2 \tag{1.1}$$

D is a dichotomic variable that takes value 1 if a certain evaluation e is employed by an article a, grouped within a specific unit of analysis. Hence, $\sum_{A}^{a} D_{ea}$ identifies the times a practice is employed by articles of the same scope. This is divided by E that identifies the total amount of evaluation decisions by that scope category and then squared. Finally, the Herfindahl score is the sum of all individual scores for each evaluation category within the unit of analysis under investigation. This process is then repeated 10,000 times. Lastly, I calculated the mean and standard deviation for both H-indexes and means.

[5] The analysis provided by Zuckerman, Kim, Ukanwa, and Von Rittmann (2003) offers crucial guidance to this end

Figure 1.5: Evaluation practices of articles with substantial semantic interest



Notes: the upset chart (Lex et al., 2014) collects the following information. Full dots connected by lines characterize the set of evaluation practices used. The horizontal bar chart shows the frequency for each evaluation practice, the vertical bar chart ('Intersection size') describes the count of studies interested in a particular set of practices, and the scatter plot ('UoA') provides information on the unit of analysis of each study interested in that set ('2': Individual Topics, '3': Topology).

To answer the first goal, the Herfindahl measure (H) for the empirical distribution has been compared with mean (μ_{HS}) and standard deviation (σ_{HS}) of the simulated Herfindahl index through the z-score:

$$Z_H = \frac{H - \mu_{HS}}{\sigma_{HS}} \tag{1.2}$$

Results are displayed in Table 1.2. Comparing z-scores obtained with standard normal distribution, it emerges a significant concentration within Classification. This means that articles falling in this category rely on a smaller set of evaluation possibilities than what expected. On the contrary, the other scopes do not highlight any significant concentration if compared to a random assignment.

Table 1.2: Concentration of evaluation practices by unit of analysis

		Simulated Data		
UoA	Observed Score	Mean	Std	Z-Score
Classification	1.000	0.393	0.180	3.373
Qualitative Variable	0.284	0.267	0.043	0.399
Individual Topics	0.301	0.393	0.183	-0.501
Topology	0.289	0.477	0.240	-0.782

To answer the second goal, the mean (M) for the empirical distribution has been compared with mean (μ_{MS}) and standard deviation (σ_{MS}) of the simulated means through the z-score:

$$Z = \frac{M - \mu_{MS}}{\sigma_{MS}} \tag{1.3}$$

Results are displayed in Table 1.3. In this case, several interesting and significant [6] z-scores emerge. For what concerns Classification, z-score for the Herfindahl index find a clear explanation. Indeed, this cohort of articles widely prefer Assessment of the Statistical Model, obtaining a z-score of 7.387. In the case of Qualitative Variables subgroup, the choice for Heuristic and Eyeballing is largely higher than under random assignment, with z-scores values of 2.371 and 3.426, respectively. Looking at Individual Topics category, four evaluation practices are significant: Statistic, Eyeballing, Semantic, and External. Here, Semantic is the one receiving the greatest z-score equal to 4.242. This is also the second-largest score in Table 1.3. Lastly, Topology articles have a large preference for both Statistical and Semantic practices than expected, with 2.965 and 2.244, respectively.

What emerges from Tables 1.2 and 1.3 depicts the emergence of practice stratification within each unit of analysis. These results are robust to single evaluation practice popularity. Framing Table 1.3 in terms

[6] With a P at least 0.95.

Table 1.3: Preference for evaluation practices by unit of analysis

	Z-Scores					
UoA	Heuristic	Statistical	Eyeballing	Semantic	External	Assessment
Classification	-0.684	-1.317	-1.556	-1.152	-0.377	7.387
Qualitative Variable	2.371	1.632	3.426	0.656	0.483	-1.407
Individual Topics	-0.677	2.997	2.363	4.242	1.847	-0.677
Topology	1.057	2.965	1.614	2.244	-0.293	-0.541

of 'substantial semantic interest', it appears that articles concerned with meanings chose both Statistical and Semantic practices more than what expected under random assignment. When there is a lack of interest in meanings, diversified paths based on the unit of analysis emerge.

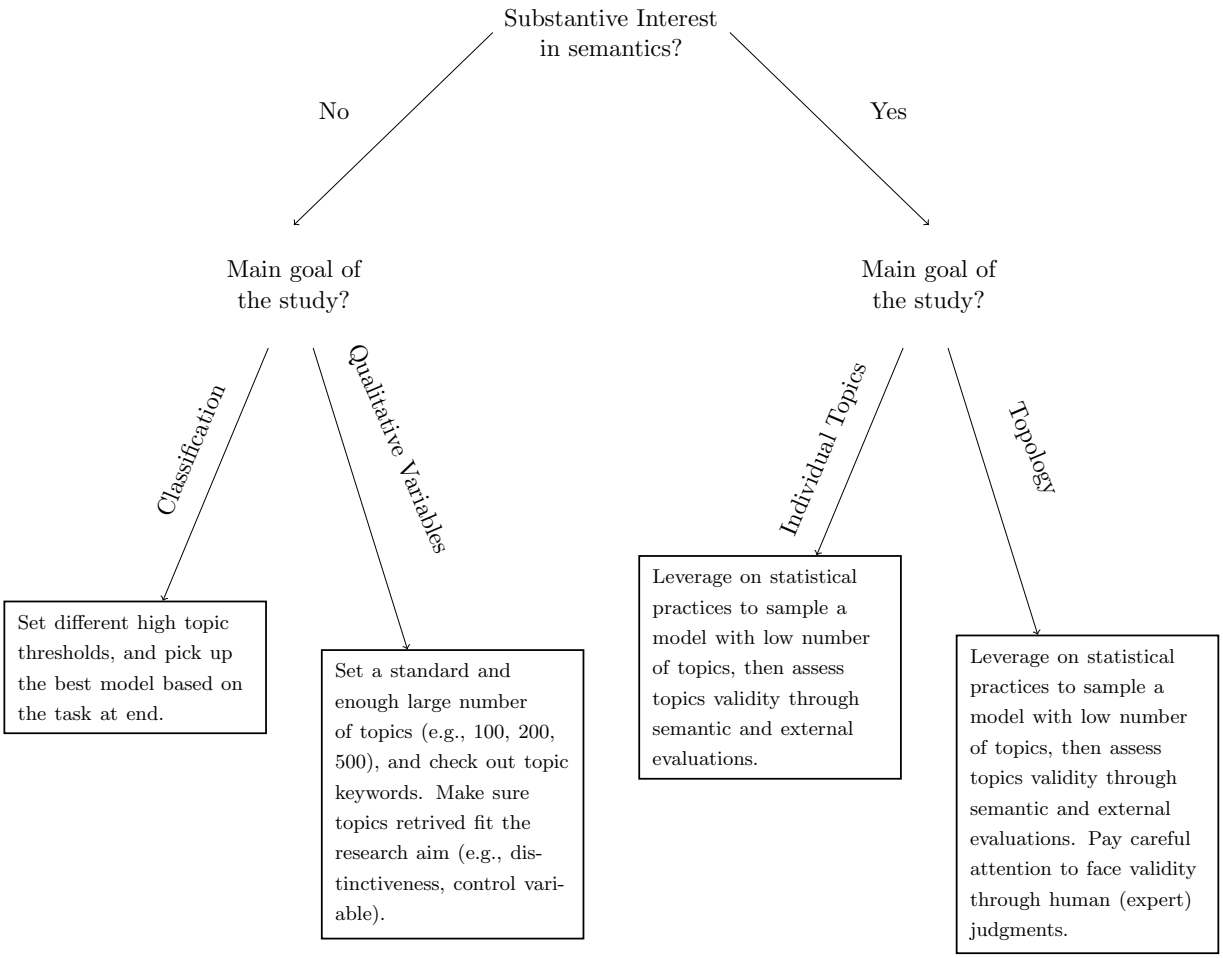
1.7 Discussion

The results described in the previous paragraphs offer interesting insights for those approaching topic modeling. In particular, the distinction between having or not having a substantial semantic interest clarifies this methodology's potential immediately. Subsequently, the separation between the different units of analysis allows researchers to have useful reference points. Once the studies are placed in their respective fields of interest, the analysis goes through evaluation practices, aiming to highlight the relationships between the purpose of the research and the evaluation of the results. Figures 1.4 and 1.5, together with the analysis of the z-scores, made it possible to grasp a nascent stratification of practices by scope. In particular, each unit of analysis highlights the adoption of some specific valuation practices more than a random choice.

The aggregation of these results can be summarized in the flow chart represented in Figure 1.6. In particular, the first question that any researcher interested in this methodology must ask is whether or not there is a substantive semantic interest. In our sample, twenty-eight studies do not show this interest, while ten do. Going down the left branch, the second question to ask is what the primary purpose of the search is. Six articles opted to use topic modeling as a Classification method at this crossroads, while twenty-two opted to use qualitative variables. On the contrary, going down the right branch, scholars interested in meanings can choose between an individual study of the topics or a topological study. To this choice, six articles preferred the first unit of analysis, while four preferred the second.

At this point, we can enter into the specificity of the evaluation practices for each unit of analysis. Regarding the Classification studies, the use of procedures to evaluate the model in carrying out the specific task emerges unequivocally. In this case, the ideal would be to sample multiple models with many topics

Figure 1.6: From scope to evaluation, emerging guidelines



and then compare them based on the traditional indices that consider the relationship between true positives, true negatives, false positives, and false negatives. Thus, obtaining an evaluation of the task for which topic modeling is used. An excellent example is offered by Larsen and Bong (2016). They use an annotated database to compare the algorithm's results and calculate the classic measurements of Precision, Recall, and F-1. With regards to studies interested in crafting Qualitative Variables, the indication that emerges from the z-scores is to set a high level of topics (e.g., 100, 200, 500), which must be evaluated with standard eyeballing practices. A suggestion that can be grasped between the lines is to make sure that the topics emerged are useful for the authors' purpose. In particular, measuring uniqueness presupposes comparing the actors on issues suited to the sense of distinctiveness sought. Haans (2019) and Kaplan and Vakili (2015) offer an example of combining a heuristic approach to eyeballing, with a particular focus on the significance of the qualitative measures created.

The authors' attention shifts to different practices in studies with a substantive interest in meanings. Regarding the studies interested in analyzing the individual topics, we tend to prefer a relatively low topic sampling and guided by statistical methods. Once the statistical optimum has been identified, the resulting topics are semantically analyzed. Adding an external analysis to these two steps can give a more robust justification of the meanings given to the individual topics. On the contrary, being reduced to just the use of eyeballing practices seems insufficient and superfluous if combined with more rigorous analyzes of a semantic or external nature. Doldor, Wyatt, and Silvester (2019) offer an example of combining different statistical methods to find the best model, sampling 12 topics. Subsequently, by analyzing the meanings, the authors can offer sufficient evidence of the identified topics' goodness. Concerning topology studies, statistical practices are once again required to sample a relatively limited number of topics to be semantically validated. However, additional effort is required for topology studies involving experts who can guarantee face validity. For example, Antons, Kleer, and Salge (2016) statistically sample 57 topics then semantically evaluated by the authors. Later, these same topics are first assessed by a group of 14 academics. Then, for each topic where the authors and scholars' interpretations differ, an in-depth discussion is conducted with a panel of experts.

The flow chart in Figure 1.6 summarizes these results by offering a clear outline useful for structuring a coherent research design. It also allows us to understand how the research aims are related to the necessary evaluation methodologies. Finally, it offers a further opportunity in the search for reproducibility.

1.8 Conclusion

From the analysis of 38 management articles that use topic modeling, this research has tried to highlight the link between purposes and methods of evaluation. In this way, the study stands as a complement to Hannigan et al. (2019), which has the great merit of having summarized technicalities, challenges, and opportunities related to topic modeling. Concerning the Annals piece, this study explores the relationship between what the authors identify as 'rendering of topics' and 'rendering of theoretical artifacts'. Therefore, by placing the selected articles in homogeneous classes and comparing their evaluation practices, the author drew up a flow chart that helps to clarify the relationship between researchers' scope and evaluation practices. In so doing, this work sheds light on the expectations that authors and reviewers must have by offering the following contributions. First, it provides an overview of the various evaluation practices present, promoting convergence in their use. Second, it provides a simple and generalizable classification of topic modeling uses, necessary to connect evaluation purposes and tools. Finally, it offers useful insights to improve the reproducibility of topic modeling in managerial sciences.

The study has limitations too. Having analyzed topic modeling in the early stages of introduction into management research, the categories identified may be subject to change. In this sense, the author's effort has been to favor generalizability rather than specificity. Besides, the list of assessment practices identified cannot be considered exhaustive and is inevitably subject to becoming obsolete. For example, in the first steps of topic modeling, perplexity was the primary statistical measure. Today there are at least nine more. Finally, having focused attention on the management literature was necessary not to disperse the research's objective. Nevertheless, this choice has apparent limitations.

Before concluding, some final reflections on threats and opportunities may be shared. As researchers, we must critically reflect on what a technique is made for, avoiding to burdening it with expectations that it cannot and must not meet. In particular, topic modeling cannot substitute the researcher's effort to obtain valid findings or interpretations. Topic modeling is no *"Deus ex Machina"*. At the same time, social scientists must be aware of the challenges involved in this methodology by consciously learning and taking inspiration from computer scientists' work. In so doing, we must avoid methodologically blind applications. Convergence and dialogue among social and computer scientists represent an enormous opportunity to advance science (e.g., DiMaggio, Nag, and Blei, 2013), exploring synergies with different epistemological paradigms toward novel ways of theorizing and novel theories.

This page is intentionally left blank.

Chapter 2

Proximity and innovation: Tracking down management theory with topic modeling

Abstract

This literature review focuses on the relationship between proximity and innovation. In particular, it proposes a flavored managerial analysis of the research advances up to 2019. In so doing, this chapter analyses 98 articles from top journals through topic modeling, applying a folding-in strategy and a structured coding scheme. Combining computational text analysis with qualitative analysis offers greater chances to represent the knowledge produced fairly. Firstly, the topic modeling algorithm is trained over a sample of more than 12,000 abstracts from top-tier management journals. This step allowed to 'project' the 98 sampled research over the trained topic modeling, uncovering the managerial conversations in which each article took part. Secondly, the review deepens proximity and innovation dimensions definitions and operationalizations. Then, it focuses on theoretical claims unveiling mechanisms of influence of each proximity dimension on innovation. Lastly, it classifies articles by innovation outcome interest and produces tables collecting empirical findings. The evidence provided by the topic modeling and coding scheme offered the chance to articulate ambiguities and inconsistencies and what contribution this research stream supplied to the greater management one.

Keywords: Proximity, innovation, management, topic modeling.

2.1 Introduction

The count of studies with interest in external factors leading to innovation has risen in management research. In particular, the last twenty years have seen an increasing amount of contributions dealing with actors' proximity from many management literature fields. This fervent debate on the relationship between proximity and innovation has seen developing the French school's initial work on proximity (Torre and Gilly, 2000) to more sophisticated and complex frameworks of analysis (Balland, Boschma, and Frenken, 2015; Boschma, 2005; Knoben and Oerlemans, 2006; Moodysson and Jonsson, 2007; Zeller, 2004). Also, this transfer from economic geography to management research, together with the access to more sophisticated methodologies, has broadened the scope of this research stream. However, the beauty of such a variety of contributions is not without drawbacks, leaving the reader dazed and confused.

This work aims to provide a review on this topic and refocus the debate on proximity and innovation. In particular, it deals with the scarce theoretical and analytical clarity arising and the consequent inconsistency of empirical findings (Capaldo and Petruzzelli, 2014; Lazzeretti and Capone, 2016). In so doing, it sets the research boundaries to contributions with interest to management. Hence, it posits three research questions: *What are the managerial conversations laying under this debate? What are the theoretical mechanisms investigated? What is the stratified contribution?* Reviewing 98 articles from top management journals, which show a relevant interest in the relationship between proximity and innovation, this chapter provides evidence of research areas that need further discussion. In particular, it emerges scarce attention towards proximity as a multi-dimensional concept, a low agreement on variable measurements, a scarce theoretical and empirical debate on proximity influence mechanisms over innovation, and a consequent lack of theoretical falsification that could benefit the development of this research stream.

To decrease the ambiguities arising and avoid widening too much the conceptual boundaries of proximity, which will cause an inevitable loss of significance, the proximity debate must be refocused. Hence, this chapter aims to provide the following contributions. Firstly, it offers momentum on the theoretical claims at the ground of proximity and innovation relation. Secondly, it places each contribution into its respective management conversation, defining the proximity debate's boundaries into the management literature field. Thirdly, it uncovers the stratified contribution offered by proximity scholars to the innovation

debate. Lastly, it provides a discussion of what topics are needed to advance this stream of research.

The chapter is structured in eight sections. In the next one, a brief introduction to the proximity debate is provided, recalling the key contributions which lay the foundation of this research stream. In section three, methodological strategies are covered, discussing articles retrieval, detection of managerial topics, and the coding schema employed. Then, the sampled articles' characteristics are discussed, focusing on managerial topics, proximity, and innovation dimensions. The fifth section describes theoretical claims discussed in the sampled articles, while the sixth deals with theoretical and empirical results. The last two sections are devoted to discussions and conclusions.

2.2 Proximity, a long lasting debate

Actors' closeness and its effect on economic activities has attracted scholars since the first half of the nineteenth century, with the famous contribution of Marshall (1920). The idea of an 'industrial atmosphere' pushed scholars to further investigate relations among actors in a spatial framework. Indeed, space and co-location were at the heart of industrial district (Becattini, 1990) and cluster (Porter, 1990) literature, which provided evidence of both pros and cons of such economic structures. Then, the discussion has expanded including new dimensions of closeness (Jaffe, 1989; Kirat and Lung, 1999; Torre and Gilly, 2000). This academic debate has been formalized by two great contributions — Zeller (2004) and Boschma (2005) — which provided analytical frameworks with their roots in both sociology, economics, and management.

Hereinafter, the multi-dimensional essence of proximity has been consecrated to the broader academic community, with ambiguities too. This point was made clear by both Knoben and Oerlemans (2006) and Moodysson and Jonsson (2007), which offered their conceptualization of the proximity framework. Even though these contributions aimed at decreasing complexity and ambiguity, the path to always more granular and varied dimensions of proximities has been unaltered. This path has been recognized by several authors, with a call for community-based debate to solve conceptual inconsistencies (Capaldo and Petruzzelli, 2014; Lazzeretti and Capone, 2016).

In the last years, the evolution of this framework has also faced the advent of globalization, which legitimized many famous writers to claim the inevitable loss of significance of space (Friedman, 2006). To these claims many theoretical and empirical answers were offered, such as Morgan (2004)[1] or Morescalchi

[1] 'Because information diffuses rapidly across organizational and territorial borders, it wrongly assumes that understanding does too' Morgan (2004, p. 3)

et al. (2015), and today geographical proximity still remains the pillar of this literature stream (as this study will show). Apart from the role of space, the framework has been further enriched by a temporal component, which takes into account the evolution of relations and the consequent role of proximity through time (Balland, Boschma, and Frenken, 2015). In so doing, the framework has been applied to study collaboration networks (Crescenzi, Nathan, and Rodríguez-Pose, 2016; Ter Wal, 2014), power dynamics (Hansen and Mattes, 2018), and learning processes (Davids and Frenken, 2018). In this progress, proximity essence is always more laying at the intersection of sociology, economics, and management literature.

2.3 Methods

This research focuses on theoretical and empirical contributions analyzing the influence of proximity on innovation. In particular, the focus is on both spatial and non-spatial proximity dimensions, their influence on innovation, and on the consequent contribution provided to management literature. In line with this aim, 98 research articles have been selected and analyzed as follows.

2.3.1 Proximity and Innovation articles retrieval process

Articles retrieval has been performed through Scopus, querying for the terms 'proxim*' and 'innovat*' among title, abstract, and keywords of papers published in journals from social sciences, business, and economics between 1990 and 2019[2]. This search provided 1,227 articles. Hence, I selected only contributions from journals ranked 3, 4, and 4* by the Academic Journal Guide 2018, which guarantees high publication standards from a great variety of management sub-fields. So, 329 studies were selected.

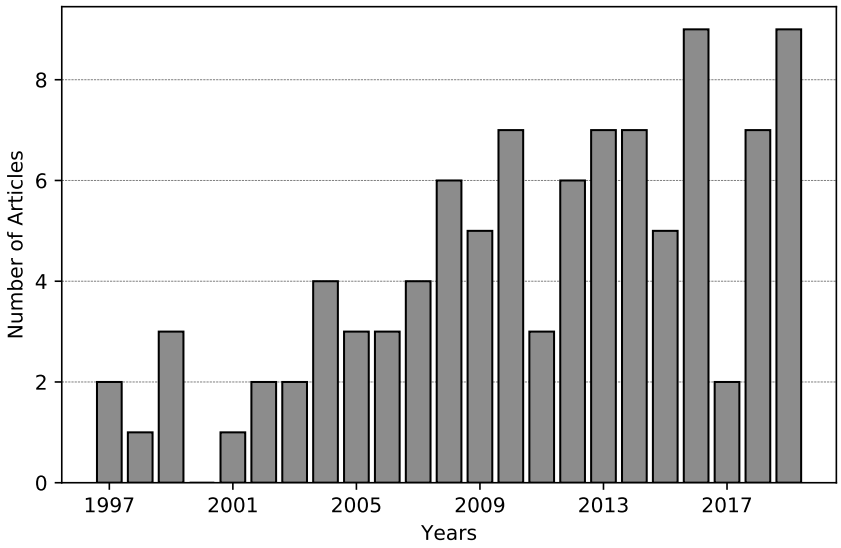
Then, the author went through the abstract or full paper (when necessary) of each article, in order to select those contributions with a substantive interest for the relationship of proximity and innovation. In particular, contributions with an exclusive interest on knowledge, research papers interested on relations among proximity dimensions and not innovation, articles employing any proximity dimension only as control variable, and literature reviews were discarded.

The final sample is composed by 98 articles from 32 journals (see Table B.1 in the Appendix for number of retrieved articles by journal). The proximity and innovation articles included were published between 1997 and 2019, see Figure 2.1. A part from year 2000, for which there are no observation, the

[2]The following query has been performed the 9^{th} of Jannuary 2020: TITLE-ABS-KEY ("proxim*" AND "innovat*") AND (EXCLUDE (PUBYEAR, 2020) OR EXCLUDE (PUBYEAR, 1989) OR EXCLUDE (PUBYEAR, 1988) OR EXCLUDE (PUBYEAR, 1987) OR EXCLUDE (PUBYEAR, 1986) OR EXCLUDE (PUBYEAR, 1985) OR EXCLUDE (PUBYEAR, 1984) OR EXCLUDE (PUBYEAR, 1982) OR EXCLUDE (PUBYEAR, 1981) OR EXCLUDE (PUBYEAR, 1979) OR EXCLUDE (PUBYEAR, 1978) OR EXCLUDE (PUBYEAR, 1977) OR EXCLUDE (PUBYEAR, 1975) OR EXCLUDE (PUBYEAR, 1974) OR EXCLUDE (PUBYEAR ,1973) OR EXCLUDE (PUBYEAR, 1972) OR EXCLUDE (PUBYEAR, 1970) OR EXCLUDE (PUBYEAR, 1969)) AND (LIMIT-TO (SUBJAREA, "SOCI") OR LIMIT-TO (SUBJAREA, "BUSI") OR LIMIT-TO (SUBJAREA, "ECON")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "re"))

academic attention towards this topic seems constant and — in general — increasing.

Figure 2.1: Number of articles per year



2.3.2 Mangement conversations

This chapter leverages on *topic modeling* (Blei, 2012; Blei, Ng, and Jordan, 2003) to identify the managerial conversations to which proximity and innovation articles refer. The potentials of topic modeling for social sciences are extremely broad (DiMaggio, Nag, and Blei, 2013; Hannigan et al., 2019), as some researches from the management field prove (Antons, Kleer, and Salge, 2016; Sieweke and Santoni, 2020). In particular, this methodology enables the analysis of enormous amount of text, for which manual coding will be not feasible. In the analysis, I employ the *Latent Dirichlet Algorithm* (LDA), the most common and versatile topic modeling algorithm (Blei, 2012). The outcomes of LDA are a posterior probability linking a word to a topic, and a posterior probability linking a topic to a document. Hence, a topic is described by a collection of words, and a document is described by a collection of topics. Notably, each topic is assigned with different probabilities to each document (abstracts, in my case). This enables the retrieval of multiple topics of interest in a single document. For example, a paper dealing with technology adoption and competitive dynamics in international markets, will be characterized by two topics referring to 'technology' and 'competition'.

Since the proximity and innovation articles come from different research fields (e.g. economic geography, innovation, entrepreneurship), I followed the methodological strategy of Sieweke and Santoni (2020). In particular, I build a database composed by 12,762 abstracts from eight top organization and management

theory journals[3] (see Table B.2 for journals and number of articles). This database (training set) has been used to train the topic modeling algorithm in order to learn management topics. Then, the abstracts of the 98 proximity and innovation articles were analyzed through the trained algorithm, thus retrieving the same topics learned in the training set.

The execution of this methodological design required five steps. Firstly, the training set was pre-processed to remove unnecessary noise from the dataset. Secondly, I selected an appropriate number of topics for the LDA. In particular, since this algorithm requires to specify an a-priori number of topics, I run a coherence score metrics (Mimno, Wallach, Talley, Leenders, and McCallum, 2011) on a range of ± 20 topics from 64, which is the number of keywords assigned by the Academy of Management to the Organization and Management Theory division. Indeed, these keywords should provide a fair representation of theoretical and empirical topics discussed in the training set (composed by organization and management theory journals). Forty-seven is the number of topics for which the coherence score achieves its maximum value, thus indicating the most suitable solution. Thirdly, LDA-Mallet[4] was run on the training set, obtaining the distribution of the 47 topics in the training set. During this step I applied some eyeballing techniques (textual and visual inspection of most frequent words per topic) in order to evaluate the goodness of the retrieved topics. Then, I applied similar pre-processing techniques on the 98 proximity and innovation abstracts as in the first step. Lastly, I retrieved the topic learned during the third step in the proximity and innovation abstracts. The topic assignment to each document has been extensively analyzed through semantic evaluation and some topics dealing with methodological or technical issues have been excluded from this analysis. Appendix B.2 provides a detailed discussion of the steps here summarized.

2.3.3 Setting-up the coding schema

In order to uncover underlying structures and paths, I took a qualitative approach to complement the topic modeling analysis (see Doldor, Wyatt, and Silvester, 2019 or Croidieu and Kim, 2018 for great examples). In particular, I set up a coding schema to extract rich textual information on both proximity and innovation relations.

The coding schema has been outlined through an iterative process. Firstly, the author analyzed a set of ten papers. At this stage the author collected data on proximity variables, innovation dimensions, the underlying theoretical mechanisms, and findings. After a second screening on a similar set of articles, the

[3] The collection has been performed the 11^{th} of January 2020

[4] Mallet is a Java-based topic modeling toolkit that uses an optimized version of collapsed Gibbs sampling

author came up with the final version of the coding schema. This schema was employed to analyze the full sample of 98 papers, its structure is reported in Table 2.1.

The schema is divided in five distinctive domains: Scope, Proximity, Innovation, Theoretical Mechanisms, and Findings. The former gathers information on both the Research Aims and Literature Gaps that the paper aims to address. Given the richness and diversity of scopes moving researchers, both data were stored as text. For what concerns Proximity domain, on top of the five dimensions identified by Boschma (2005) (geographic, cognitive, institutional, organizational, and social), I gathered information also on 'technological' proximity (Knoben and Oerlemans, 2006) and on 'others' (e.g. functional or relational, see Moodysson and Jonsson, 2007). Each proximity dimension was coded as covered [Yes = 1] or not [No = 0]. Furthermore, I also stored the definition given and the operationalization (for quantitative studies) of each proximity dimension. For what concerns the innovation domain, I distinguished three different interests: (i) outcome, containing studies analyzing product or process innovation, radical or incremental innovation, or innovation performance; (ii) process, including studies that analyze different phases of innovation process, and those interested in innovation diffusion and adoption; (iii) network, grouping studies with a substantive interest on tie formation in innovation collaborations. The presence [Yes = 1] or absence [No = 0] of each innovation focal interest has been retrieved, with the variable operationalization (stored as text). Furthermore, the Theoretical Mechanism domain aims to collect the reasons put forward for the sought relationship. Hence, for each proximity dimension, I stored quote from text disentangling the underlying mechanism. In line with the recommendations of Martin (2015), I also summarized the intellectual skeleton of theoretically complex papers. Lastly, the Findings domain stores information on research findings and explanation for the effect found. Also, for each proximity dimension I collected the uncovered effect as present [Significant (Important) = 2], somewhat present [Mixed = 1] or not [Not Significant (Not Important) = 0].

2.4 Sample Description

This section introduces the three sets of variables analyzed in this review, namely management conversations, proximity dimensions, and innovation dimensions, and their characteristics within the sampled articles.

2.4.1 Management conversations

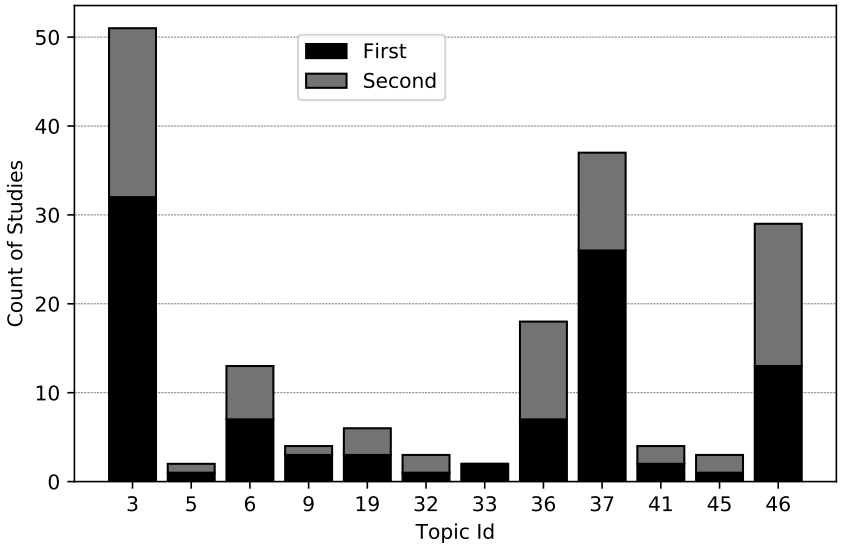
Among the 47 topics obtained training the LDA on the sample of 12,762 organization and management theory abstracts, 12 topics result as most discussed in the 98 proximity and innovation articles selected. In

Table 2.1: Coding Schema

Domain	Variables [Synopsis]
Scope	Literature Gap [quote from text]
.....	Research Aim [quote from text]
Proximity	Geographic proximity [No=0; Yes=1]
.....	Cognitive proximity [No=0; Yes=1]
.....	Institutional proximity [No=0; Yes=1]
.....	Organizational proximity [No=0; Yes=1]
.....	Social proximity [No=0; Yes=1]
.....	Technological proximity [No=0; Yes=1]
.....	Other proximity [No=0; Yes=1]
.....	Proximity definitions [quote from text]
.....	Proximity measurements [quote from text]
Innovation	Innovation outcome [No=0; Yes=1]
.....	Innovation process [No=0; Yes=1]
.....	Innovation network [No=0; Yes=1]
.....	Innovation measurements [quote from text]
Theoretical Mechanisms	Geographic proximity and innovation [quote from text]
.....	Cognitive proximity and innovation [quote from text]
.....	Institutional proximity and innovation [quote from text]
.....	Organizational proximity and innovation [quote from text]
.....	Social proximity and innovation [quote from text]
.....	Technological proximity and innovation [quote from text]
.....	Other proximity and innovation [quote from text]
.....	Intellectual skeleton [summary]
Findings	Geographic proximity effect [Not significant (important) = 0; Mixed = 1; Significant (important) = 2]
.....	Cognitive proximity effect [Not significant (important) = 0; Mixed = 1; Significant (important) = 2]
.....	Institutional proximity effect [Not significant (important) = 0; Mixed = 1; Significant (important) = 2]
.....	Organizational proximity effect [Not significant (important) = 0; Mixed = 1; Significant (important) = 2]
.....	Social proximity effect [Not significant (important) = 0; Mixed = 1; Significant (important) = 2]
.....	Technological proximity effect [Not significant (important) = 0; Mixed = 1; Significant (important) = 2]
.....	Other proximity effect [Not significant (important) = 0; Mixed = 1; Significant (important) = 2]
.....	Research findings and explanation [quote from text]

particular, I looked at the most 'relevant' topics (in terms of posterior probability) per each document, and retained the first (topic with the highest posterior probability for that document). Figure 2.2 shows the count of studies in which the 12 topics appeared as first or second best. Here, Topic 3 emerges as the most relevant, followed by Topic 37 and 46. Topic 6 and 36 received somewhat attention, while the others characterized a small set of the sample. Table 2.2 displays the ten most important lemmas for each of the 12 topics (see Table B.4 for the full set of 47 topics). To interpret and label each topic I leveraged on these keywords and the proximity and innovation articles in which they appear as dominant.

Figure 2.2: First and Second most relevant topic occurrences



As Figure 2.2 shows, Topic 3, 6, 36, 37, and 46 represents the most prominent conversations. Topic 3 collects articles with an interest on *Technology & Production*. In particular, authors attention includes technological spaces of interactions (Ardito, Messeni Petruzzelli, Pascucci, and Peruffo, 2019; Phene, Fladmoe-Lindquist, and Marsh, 2006), technological intensive companies (Li, Qiu, and Wang, 2019) and their expansion (Zeller, 2004), or technological production (mainly in the form of patents, e.g. Schwartz, Peglow, Fritsch, and Günther, 2012; Sonn and Storper, 2008). Topic 6 is characterized by a focus on *Portfolio of Alliances*. Hence, researchers deal with topics such as portfolio diversity (Van de Vrande, 2013), knowledge integration within alliances (Capaldo and Petruzzelli, 2014), and global pipelines connecting international partners (Fitjar and Rodríguez-Pose, 2011). Topic 36 shifts the attention towards *Knowledge & Learning*. Here, papers explores determinants and effects of knowledge spillovers (Ghio, Guerini, and Rossi-Lamastra, 2016; Ponds, Oort, and Frenken, 2010), knowledge creation (Moodysson, 2008; Moodysson, Coenen, and Asheim, 2008), transfer (Weidenfeld, Williams, and Butler, 2010), and acquisition (Parra-Requena, Ruiz-

Ortega, García-Villaverde, and Rodrigo-Alarcon, 2015). Topic 37 contains researches with an appeal for *Networks & Embeddedness*. In particular, authors deal with network structures and characteristics, such as direct/indirect ties (Liang and Liu, 2018), network centrality (Whittington, Owen-Smith, and Powell, 2009), and triadic closure (Ter Wal, 2014), or collaboration in general (Lazzeretti and Capone, 2016; Steinmo and Rasmussen, 2016) and embeddedness (Knoben and Oerlemans, 2012; Owen-Smith and Powell, 2004). Topic 46 includes articles with an interest in *Local & Distant* relations. In so doing, authors focus on local innovation capacity (Cabrer-Borras and Serrano-Domingo, 2007), institutional and cross border flows (Morescalchi et al., 2015; Weidenfeld, 2013), and the coexistence of local and non-local dynamics (Thomas III, 2004; Wang and Wu, 2016).

Topic 5, 9, 19, 32, 33, 41 and 45 refer to managerial conversations toward which less attention has been devoted. In particular, Topic 5 is focused on innovation as an evolutionary process (Kirat and Lung, 1999), Topic 9 on corporate governance (such as the role of 'outside' directors in innovation activities, see Balsmeier, Buchwald, and Stiebale, 2014), Topic 19 on trust and collaboration (Bunduchi, 2013), Topic 32 on experiences (such as pre-entry experiences, see Weterings and Koster, 2007), Topic 33 on business and entrepreneurship (Letaifa and Goglio-Primard, 2016), Topic 41 on firms entry, growth, and survival (Weterings and Koster, 2007) and Topic 45 on the strategic role of markets for product development (such as market intelligence, see Cornish, 1997).

2.4.2 Proximity dimensions

Figure 2.3: Number of articles per year

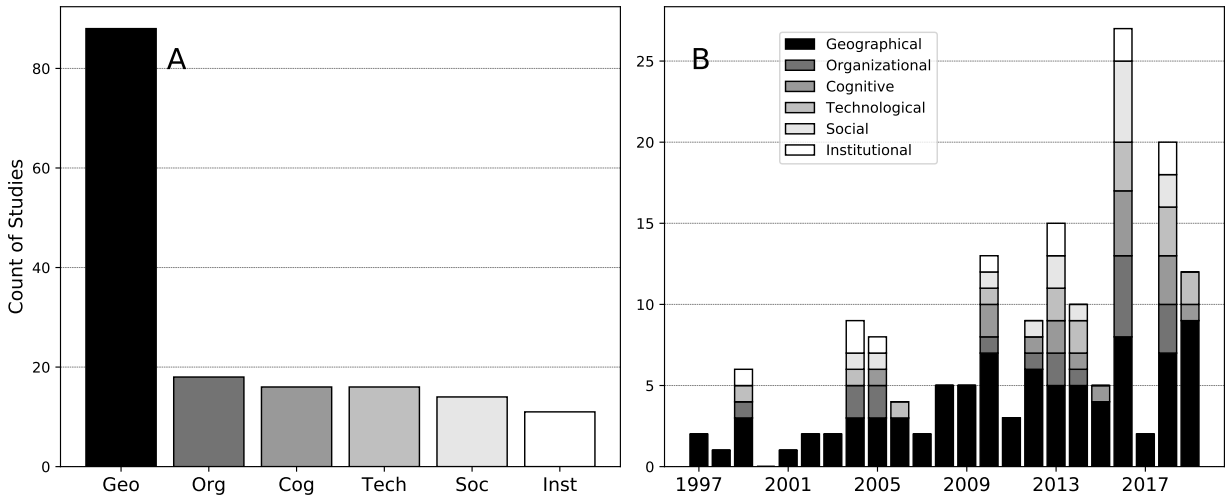


Table 2.2: First ten lemmas and posterior probabilities per topic

Id	Label	Lemmas									
3	Technology &	technology	technological	production	company	firm	manufacturing	patent	technical	datum	productivity
	Production	0.082	0.042	0.026	0.023	0.019	0.018	0.017	0.016	0.016	0.014
5	Innovation as	process	innovation	dynamic	mechanism	develop	innovative	case	explain	evolution	configuration
	Process	0.117	0.092	0.052	0.025	0.023	0.019	0.015	0.014	0.013	0.013
6	Portfolio of	alliance	activity	partner	benefit	increase	firm	find	selection	portfolio	formation
	Alliances	0.097	0.079	0.063	0.053	0.024	0.024	0.022	0.022	0.019	0.017
9	Corporate	corporate	board	stakeholder	governance	director	company	reputation	financial	firm	corporation
	Governance	0.097	0.056	0.042	0.038	0.036	0.031	0.022	0.02	0.019	0.018
19	Trust &	relationship	trust	exchange	relational	mechanism	contract	governance	supplier	transaction_cost	firm
	Collaboration	0.059	0.05	0.043	0.033	0.026	0.024	0.023	0.02	0.019	0.016
32	Experiences	experience	venture	capital	prior	event	find	examine	firm	signal	initial
		0.104	0.047	0.046	0.035	0.034	0.026	0.021	0.02	0.018	0.017
33	Business &	business	opportunity	success	entrepreneurial	entrepreneurship	entrepreneur	failure	develop	successful	firm
	Entrepreneurship	0.123	0.055	0.036	0.035	0.03	0.03	0.028	0.023	0.02	0.018
36	Knowledge &	knowledge	learning	external	learn	internal	transfer	source	firm	creation	suggest
	Learning	0.190	0.066	0.047	0.044	0.041	0.032	0.027	0.017	0.014	0.012
37	Network &	network	tie	social	position	collaboration	structure	relationship	firm	embeddedness	interorganizational
	Embeddedness	0.132	0.049	0.033	0.032	0.023	0.019	0.017	0.016	0.014	0.013
41	Entry, Growth	industry	entry	rate	growth	incumbent	early	firm	find	survival	exit
	& Survival	0.119	0.034	0.031	0.027	0.021	0.02	0.019	0.019	0.016	0.014
45	Markets	market	product	competitive	diversification	competition	cost	industry	copyright	firm	customer
		0.153	0.075	0.047	0.029	0.027	0.026	0.026	0.025	0.02	0.02
46	Local &	acquisition	country	international	foreign	local	subsidiary	global	enterprise	national	japanese
	Distant	0.050	0.034	0.033	0.03	0.027	0.026	0.023	0.019	0.019	0.018

Fig 2.3 sub-plot A shows the frequency of occurrence for each proximity dimension (studies concerned with 'other proximity' dimensions are discussed in Appendix B.3). The sample shows a clear unbalance of research attention toward geographical proximity, with almost all articles showing their interest to this category. On the other hand, non-spatial proximities are by far less investigated. This is in clear contrast with the examples and calls for the analysis of non spatial proximities received from several authors such as Torre and Gilly (2000), Zeller (2004), Boschma (2005). The reason for this lack may be sought in the words of Moodysson and Jonsson (2007, p. 118):

> *'However, all these conceptualizations face problems when applied in empirical studies like the one carried out in this article. The categories employed by Torre and Gilly are too vague and loosely defined to be operational, while those of Boschma and Zeller include too many points of overlap to provide a consistent framework.'*

Unfortunately, neither the Moodysson and Jonsson (2007) categorization of functional and relational proximities, nor the great analysis and conceptualization offered by Knoben and Oerlemans (2006) lead to a path brake in the analysis of non spatial proximities in innovation research.

Even though geographical proximity has kept the scholars attention towards the whole period considered, Fig 2.3 sub-plot B shows a slight increase of interest towards some non-spatial dimensions at least in the last few years. After 2009, a rising attention towards organizational, cognitive, technological, and social proximity emerges. On the contrary, the institutional dimension falls behind. To further dig into the multi-dimensional characterization of the proximity concept, Fig 2.4 traces out the different combination of proximity investigated in my sample of articles. In so doing, it focuses on 28 studies (theoretical or empirical) with an interest in at list two dimensions [5]. Even in this sub-sample, the spatial dimension is over-represented, with all studies combining this proximity with others. Noteworthy, geographic and organizational dimensions represent the most explored dyad (18 studies). The Boschma, 2005 framework has been approached as a whole by four more articles (Davids and Frenken, 2018; Hall and Jacobs, 2010; Lazzeretti and Capone, 2016; Letaifa and Rabeau, 2013), while other five researches leaved the institutional dimension out (Broekel and Boschma, 2012; Crescenzi, Nathan, and Rodríguez-Pose, 2016; Dolfsma and Van der Eijk, 2016; Jespersen, Rigamonti, Jensen, and Bysted, 2018; Steinmo and Rasmussen, 2016). Furthermore, two qualitative studies (Brink, 2018; Cantù, 2010) combined the technological and cognitive dimensions, two

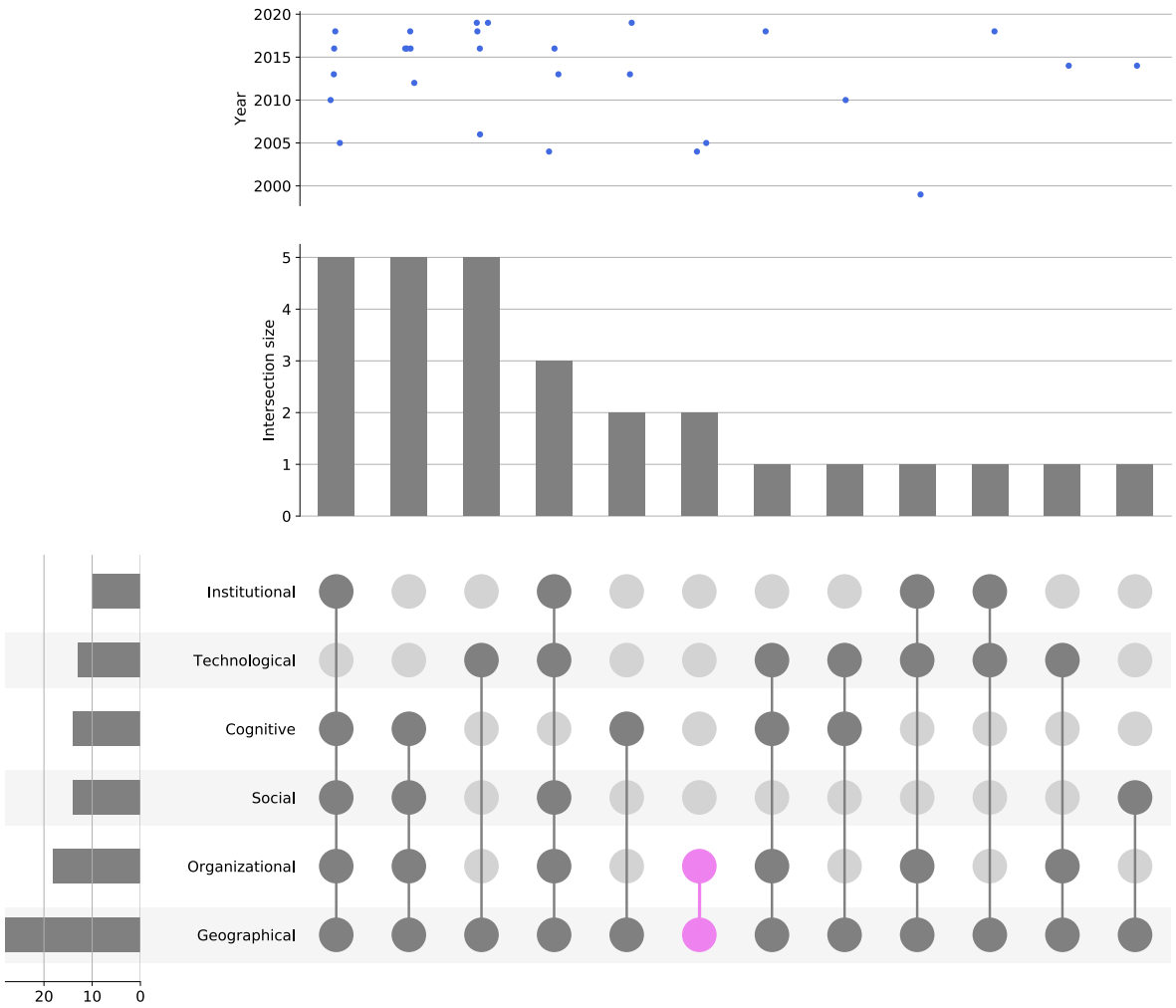
[5] Hence, the figure excludes 57 articles interested in the sole geographical proximity, 2 in cognitive, 1 in institutional, 3 in technological, and 6 articles considering 'other' dimensions alone or together with geographical proximity.

categories not easy to distinguish (see Knoben and Oerlemans, 2006). Lastly, Fig 2.4 provides the temporal distribution of articles for each set. In particular, only 6 articles concerned with multiple proximity dimensions were published before 2010, while 13 came out in the last five years. Multi-dimensional proximity is still an emergent topic.

For what concerns definition and measurement of proximity dimensions, the sampled articles show — in general — a great level of agreement on definitions, but a scarce convergence on measurement strategies. The following discussion tries to summarize and display the most evident trends inside my sample (a collection of all measurements is available in the Appendix B.3). Even though consensus has diffused on defining geographical proximity as *'physical or spatial distance'* among actors (Capaldo and Petruzzelli, 2014; Crescenzi, Nathan, and Rodríguez-Pose, 2016; Letaifa and Rabeau, 2013), great variety has risen in its measurement. Indeed, while some authors measure this proximity as a crude distance among two actors (Chu, Tian, and Wang, 2019; Shearmur, 2011; Ter Wal, 2014), others add different degrees of 'human intrusion'. In so doing, many authors prefer (i) 'time travel' distances which are tied to human infrastructures (Drejer and Ostergaard, 2017; MacPherson, 1998; Ponds, Oort, and Frenken, 2010), (ii) co-location in the same cluster, region, or nation (Kapetaniou and Lee, 2019; Lazzeretti and Capone, 2016; Presutti, Boari, Majocchi, and Molina-Morales, 2019), or (iii) individual perception of distance (e.g. Fernandes and Ferreira, 2013 and Stephan, 2014). In the case of organizational proximity, two main definitions coexist in my sample. The first is inspired by the *logic of belonging and similarity* posed by Torre and Gilly (2000) and Torre and Rallet (2005), which involves also a cognitive valuation. The other comes from the Boschma (2005) framework and distinguishes among the *rate of autonomy and degree of control* exerted by partners. The variable measurement reflects these positions, with some authors looking at membership to the same research or business group (Capaldo and Petruzzelli, 2014; Crescenzi, Nathan, and Rodríguez-Pose, 2016), or at partnership with actors of the same type (Broekel and Boschma, 2012) or sharing similar organizational mechanisms (Jespersen, Rigamonti, Jensen, and Bysted, 2018), while others at hierarchies and at the distance generated by those hierarchies (Dolfsma and Van der Eijk, 2016). For what concerns cognitive proximity, a great part of authors agreed on defining this dimension as the similarity in the way actors *'perceive, interpret, understand and evaluate the world'* (Cantù, 2010; Lazzeretti and Capone, 2016; Letaifa and Rabeau, 2013; Steinmo and Rasmussen, 2016; Taura and Radicic, 2019) [6]. A similar cohesion does not recur in the variable measurement, where some employ patents or technological similarity (Broekel

[6]Furthermore, some authors link cognitive proximity to the overlap on 'knowledge bases' (Broekel and Boschma, 2012; Jespersen, Rigamonti, Jensen, and Bysted, 2018), other to a 'similar knowledge background' more than knowledge bases (Davids and Frenken, 2018), or to the 'degree of similarity in routines, culture, habits, common values and norms between organizations' (Parra-Requena, Ruiz-Ortega, García-Villaverde, and Rodrigo-Alarcon, 2015, p. 151).

Figure 2.4: Combination of proximity dimensions per year



Notes: the upset chart (Lex et al., 2014) collects information from 28 studies investigating (theoretically or empirically) more than one dimension of proximity in their research. The set of proximities explored are characterized by full dots connected by lines. The horizontal bar chart shows the frequency for each proximity within the 28 articles, the vertical bar chart ('Intersection size') describes the count of studies interested in a particular set, and the scatter plot ('Year') provides information on the publication year of each study interested in that set. The most recurrent dyad is highlighted in violet.

and Boschma, 2012; Crescenzi, Nathan, and Rodríguez-Pose, 2016), while others ask respondent to assess closeness on 'knowledge bases' or relatedness of technological disciplines (Dolfsma and Van der Eijk, 2016; Jespersen, Rigamonti, Jensen, and Bysted, 2018; Taura and Radicic, 2019). On the other hand, authors with an interest towards technological proximity show higher levels of agreement on both definition and measurement. Indeed, this dimension seeks to capture shared technological knowledge base and experience of economic actors (Cantù, 2010; Guan and Yan, 2016; Zeller, 2004), and the great majority of sampled articles leveraged on differences or similarities among patents' classes [7] to measure actors' proximity (Ardito, Messeni Petruzzelli, Pascucci, and Peruffo, 2019; Balsmeier, Buchwald, and Stiebale, 2014; Capaldo and Petruzzelli, 2014). For what concerns social proximity, authors are mostly aligned with Boschma (2005) definition, which is deeply rooted in the works of Granovetter (1985) and Uzzi (1997). Hence, this proximity dimension aims to represent the social embeddedness of actors, which involves trust based on friendship and experience (Broekel and Boschma, 2012; Letaifa and Rabeau, 2013; Ter Wal, 2014). Most researches tend to measure social proximity on the base of experience in previous collaborations (Broekel and Boschma, 2012; Crescenzi, Nathan, and Rodríguez-Pose, 2016; Lazzeretti and Capone, 2016; Marrocu, Paci, and Usai, 2013), while other asked respondents to state some relations' qualities (Dolfsma and Van der Eijk, 2016; Jespersen, Rigamonti, Jensen, and Bysted, 2018). Lastly, the researches dealing with Institutional proximity refer to this dimension as the cultural codes and economic institutions (Davids and Frenken, 2018), which provide the context (Boschma, 2005; Thomas III, 2004) or framework (Marrocu, Paci, and Usai, 2013; Zeller, 2004) that regulate business and non-business transactions (Letaifa and Rabeau, 2013). The few measures adopted in the sample take as unit of analysis economic actors (Lazzeretti and Capone, 2016; Liang and Liu, 2018), or international discrepancies (Marrocu, Paci, and Usai, 2013; Thomas III, 2004).

2.4.3 Innovation dimensions

The innovation dimensions investigated in the sample have been grouped in three sets, namely: *Outcome*, *Network*, and *Process*. In particular, the first set is composed by articles with an interest in the role of proximity in fostering or hindering the outcome of innovation, such as product or process, and radical or incremental innovation. On the contrary, studies aiming to investigate collaborative ties formation, compose the *Network* set. Lastly, articles dealing with the influence of proximity in the different phases of innovation are grouped into the *Process* set. Both theoretical and empirical articles have been grouped on the basis of this classification. Also, articles with multiple aims that fall in different categories, have been included in

[7] Most of authors employed or adapted the technological proximity measure of Jaffe (1986, p. 986) and Jaffe (1989, p. 88)

both sets, but discussing only the relevant research component to the relative innovation category.

Figure 2.5: Number of articles per year

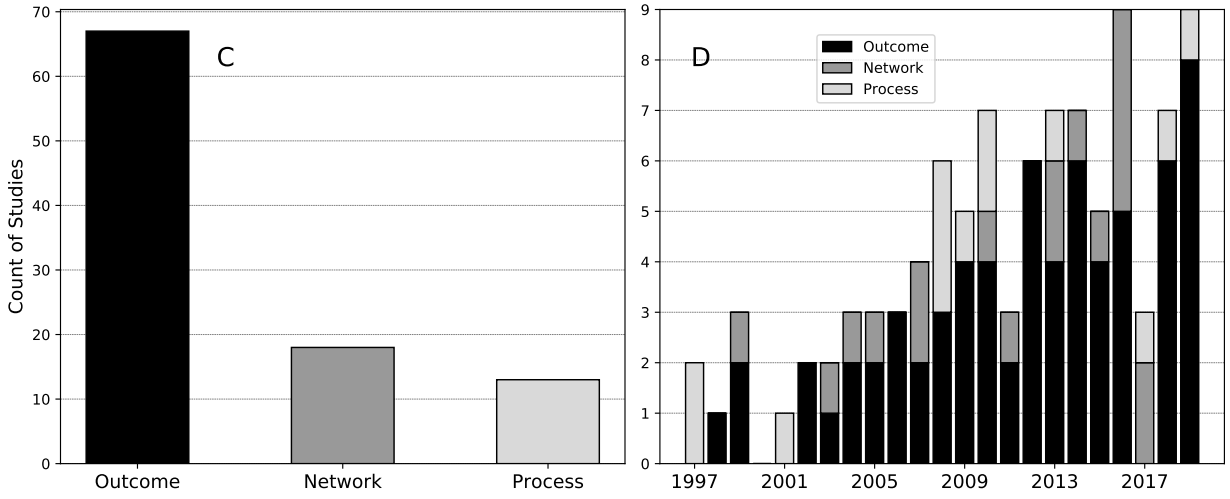


Figure 2.5 subplot C shows the vast majority of sampled articles (a count of 67) being clustered into the *Outcome* set. *Network* and *Process* studies account for a third of the total, with 18 and 13 studies respectively. Figure 2.5 subplot D displays the temporal evolution of authors interest. What emerges is a increase of articles dealing with innovation *Outcome* after 2005, while *Network* studies maintain a somewhat constant but lower attention. Articles dealing with *Process* do not show any clear temporal pattern, resulting in some spotted contributions. Similarly to proximity dimensions, the complexity of data collection and variable measurement could be ascribed as the cause of this great unbalance.

In terms of variable measurement, the *Outcome* set has been distinguished in four sub-groups on the basis of the specific innovation measures employed, namely: technological innovation, product & process innovation, innovation novelty, and economic value of innovation. The articles in the first set build their measures on patents applications or citations. Those dealing with product & process innovation mainly employ a count of introduced innovations or a dichotomous indicator for that introduction. The third sub-set mainly measures novelty distinguishing between new-to-the-market or new-to-the-firm, while the last sub-group employs turnover as a weighting strategy. For what concerns *Network* studies, information on collaboration ties have been collected, with some articles adding a temporal component to ties. Lastly, the *Process* sub-group has fewer quantitative contributions, mainly dealing with innovation adoption measured as a dichotomous variable. The full collection of innovation measures is summarized from Table B.12 to Table B.17 in Appendix B.3.

2.5 Theoretical claims and mechanisms

In this sections, theoretical claims on the influence of proximity dimensions over innovation are discussed. In particular, a report is provided of the most prominent explanations of how and why each dimension could contribute or not to innovation outcome, network, or process. Even though theoretical positions varies, the central components are the voluntary or involuntary relations among economic actors. Hence, proximity dimensions are a mean to explain the essence of dyadic and network-level structures of innovation-related actions.

2.5.1 Spatial proximity

To explain how and why spatial propinquity is related to innovation, authors took different stances. In particular, the most discussed argument concerns knowledge, thus its degree of tacitness and appropriability. Indeed, the key point raised is that 'tacit knowledge resides in the heads and practices' of its producers and requires 'direct, face-to-face interactions' to be transferred (Geerts, Leten, Belderbos, and Van Looy, 2018, p. 155). Hence, actors should benefit more of closeness when the knowledge involved in a project is tacit (Ter Wal, 2014). Also, other authors highlight that knowledge is hard to be fully appropriated by its producer, this in turn makes that knowledge spills over firms' boundaries, but with a pace that decays with distance (Wang and Wu, 2016). Indeed, Oerlemans and Meeus (2005, p. 94) describe the 'non-rival nature of the locally accumulated knowledge' which 'spills unintentionally and in particular over to firms located in the region'. Torre (2008) devotes a part of his work on the analytical mismatch of these two claims. In particular, he points out the 'contradiction between the appropriability of the knowledge involved in spillovers and the non-appropriability of tacit knowledge' (p. 873). This concern seems to be left apart by later contributions in the sample.

Together with knowledge externalities, being proximate is seen to offer other economic advantages to take into account for partner selection and coordination effectiveness in innovation-related projects. Indeed, 'considerations of knowledge spill-over and sunk cost advantages are integral parts of this management problem' (Broström, 2010, p. 1312). In particular, the access to closer knowledge could be motivated by cost economizing choices (Divella, 2017), or by the decrease of risk perceived in the context of asymmetry (Maietta, 2015), or by lower transaction or re-orientation costs (Bindroo, Mariadoss, and Pillai, 2012; Taura and Radicic, 2019). Also, actors bounded rationality impedes them from complete searches, thus pushing them to choose 'the spatially and socially most proximate and cognitively satisfying solutions' (Dornbusch and Neuhäusler, 2015, p. 1362).

A part from knowledge externalities and transaction costs, geographical proximity has other mechanisms with which could influence innovation paths. For example, propinquity may enhance serendipitous encounters, reciprocity, and trust (Funk, 2014; Geerts, Leten, Belderbos, and Van Looy, 2018; Morgan, 2004). Also, it may enhance the likelihood of relations such those of the 'club cliche' cited by Davis and Greve (1997), or local 'buzzes' which are not only related to knowledge but also to inspiration (Funk, 2014). Furthermore, being co-located may enhance the creation of common 'codes of communication' (Weterings and Boschma, 2009, p. 747), or the 'emergence of idiosyncratic languages' (Capaldo and Petruzzelli, 2014, p. 68). Also, knowledge infrastructures may explain the influence of proximity on innovation in the light of the Christallerian central place theory, as discussed by Shearmur and Doloreux (2015). Lastly, other authors cited the accessibility to a skilled labor force and to high job turnovers as possible means of innovation (Leten, Landoni, and Van Looy, 2014; Sonn and Storper, 2008). In this case, the labor force is seen as a vector of knowledge, and propinquity as a facilitator of 'searching and matching' (Sonn and Storper, 2008, p. 1022).

2.5.2 Non-spatial proximity

In the case of organizational proximity, authors take diversified initial theoretical positions such as the embeddedness of Granovetter (1985), bounded rationality and routines as per Nelson and Winter (1982), the faultline theory as cited by Dolfsma and Van der Eijk (2016), or the Boschma (2005) framework. However, the majority agrees with the idea that this proximity enables actors lowering uncertainty, opportunistic behavior, and - thus - transaction costs. The diffused concern is monitoring partners (Crescenzi, Nathan, and Rodríguez-Pose, 2016; Marrocu, Paci, and Usai, 2013). Also, authors characterizing organizational proximity with a cognitive dimension, underline this as a way to foster interactive learning between actors (Oerlemans and Meeus, 2005; Zeller, 2004).

For what concerns cognitive proximity, authors mainly cite the contributions of Cohen and Levinthal (1990) and Nooteboom (1999) as theoretical roots. Indeed, this dimension enables communication and understanding on the base of actors' 'prior related knowledge' (Enkel and Heil, 2014, p. 244). The affiliation to Cohen and Levinthal (1990) idea of absorptive capacity lays at the bases of technological proximity too. However, the difference relies on the type of knowledge considered, at least theoretically (Cantù, 2010; Knoben and Oerlemans, 2006).

Articles dealing with social dimension lay their theoretical foundations on strong and weak ties (Granovetter, 1973), embeddedness (Granovetter, 1985), and social capital (Coleman, 1988). Even in this case

the reason of the effect on innovation are due to costs that actors bare during transactions. Indeed, 'in situations of high risk and high cost to opportunistic behavior, actors have a clear preference to form ties embedded in dense structures' (Ter Wal, 2014, p. 597). Hence, the trust rising in this relations should foster knowledge flow and consequently innovation (Crescenzi, Nathan, and Rodríguez-Pose, 2016; Steinmo and Rasmussen, 2016). Similarly, institutional proximity is acknowledged as decreasing actors transaction costs. Indeed, 'institutions, such as laws and norms, can provide a set of standard procedures and mechanisms which are shared by agents and, therefore, taken for granted' (Marrocu, Paci, and Usai, 2013, p. 1485). This in turn decrease actors uncertainty and bared costs (Liang and Liu, 2018).

Much of non-spatial proximity discussion deals also with negative influences of an excessive degree of proximity, and many of them are aligned with the Boschma (2005) framework. In particular, the paradoxes of embeddedness (Uzzi, 1997) and competency traps (Levitt and March, 1988) are highlighted as potentially detrimental.

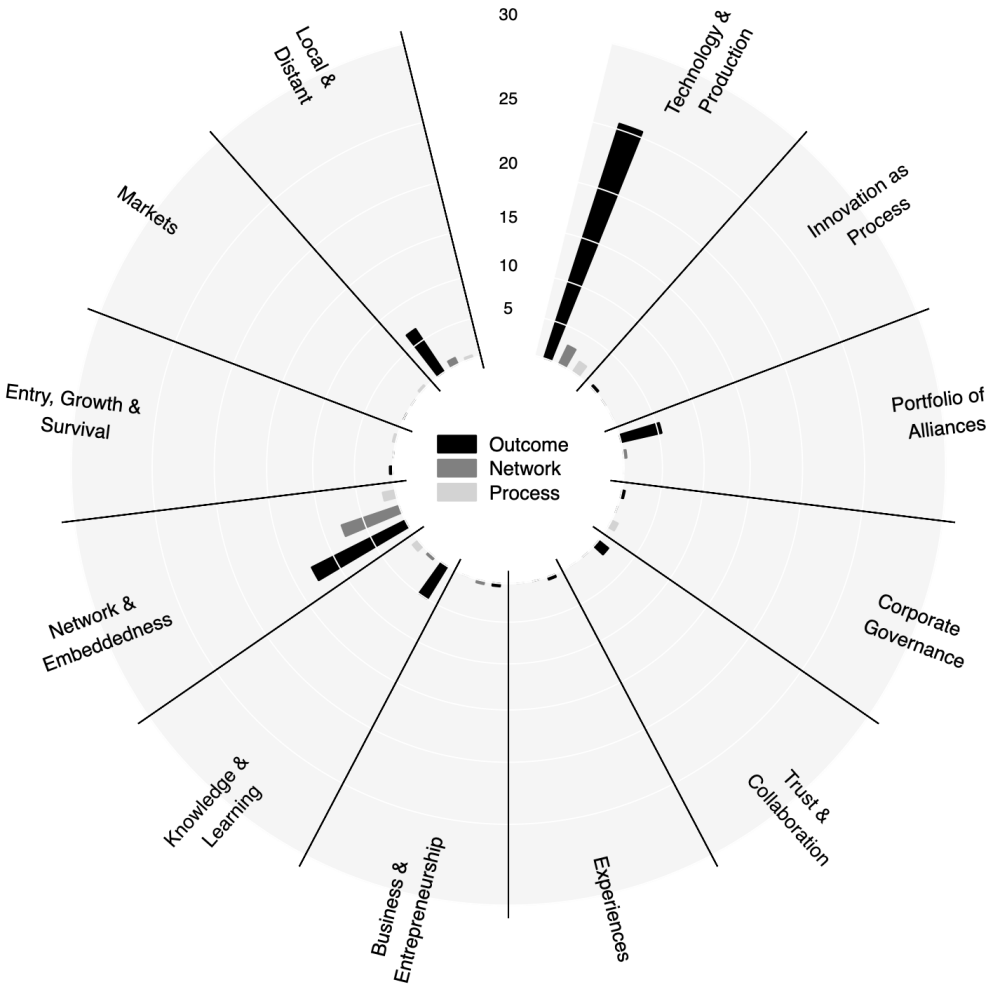
2.6 Theoretical and Empirical results

In this section, a detailed discussion of research findings is provided. In particular, per each innovation dimension investigated, managerial conversations and research findings are reported. As Fig 2.6 depicts, the vast majority of authors closer to *Technology & Production* are interested on innovation outcomes, and the same is true also for *Portfolio of Alliances*, *Knowledge & Learning*, and *Local & Distant* conversations. Studies with an interest toward the *Network & Embeddedness* topic frequently deal with innovation outcome and networks.

2.6.1 Outcome

In this section studies with an interest toward innovation outcome are collected. With a count of 67 articles, this group is the largest and it has been divided in four sub-groups of interest, namely: technological innovation, product & process innovation, innovation novelty, and economic value of innovation. Some articles fall outside this second layer of classification, showing interest toward less discussed topics, such as innovation frequency (Taura and Radicic, 2019), speed (Carbonell and Rodriguez, 2006), capability (Romijn and Albaladejo, 2002), or other innovation peculiarities (Dornbusch and Neuhäusler, 2015).

Figure 2.6: Managerial conversations and innovation dimensions



Note: this figure has been realized on the basis of a python script shared by the Bokeh library

Technological Innovation

Twenty articles show an interest toward the relationship between proximity and technological performance, which has been explored through quantitative means. In so doing, six researches touch upon *Technology & Production* (Topic 3), five *Knowledge & Embeddedness* (Topic 37), three *Local & Distant* (Topic 46), and two *Portfolio of Alliances* (Topic 6). Also, *Corporate Governance* (topic 9), *Trust & Collaboration* (Topic 19), *Business & Entrepreneurship* (Topic 33), and *Knowledge & Learning* (Topic 36) are represented conversations.

Researchers closer to *Technology & Production* span from technological conglomerates and access to new technology (Li, Qiu, and Wang, 2019), technological exploration and exploitation (Geerts, Leten, Belderbos, and Van Looy, 2018), the influence on technological performance of buyers-suppliers or universities-firms relations (Isaksson, Simeth, and Seifert, 2016; Leten, Landoni, and Van Looy, 2014), to the study of project characteristics and the consequent influence over innovation (Schwartz, Peglow, Fritsch, and Günther, 2012). Articles dealing with *Knowledge & Embeddedness* deepen the relation between network structural characteristics and innovation (Funk, 2014; Liang and Liu, 2018; Owen-Smith and Powell, 2004; Whittington, Owen-Smith, and Powell, 2009). Those researches dealing with *Local & Distant* theme focus on inter- and intra-regional relations (Cabrer-Borras and Serrano-Domingo, 2007; Marrocu, Paci, and Usai, 2013; Parent and Riou, 2005), while those close to *Portfolio of Alliances* topic dig the structure and effects of knowledge-related alliances (Capaldo and Petruzzelli, 2014; Van de Vrande, 2013). Lastly, the article dealing with Topic 9 discusses the role of executives in the innovation of firms they advice (Balsmeier, Buchwald, and Stiebale, 2014), the one close to Topic 19 touch upon the effect of customer-supplier relation (Chu, Tian, and Wang, 2019), that closer to Topic 33 deals with the role of science park (Link and Scott, 2003), while the one approaching Topic 36 discusses academic knowledge spillovers and their effect on regional innovation (Ponds, Oort, and Frenken, 2010).

For what concerns results, this sub-sample shows a great level of agreement on the effect of geographical proximity on technological innovation. This is not without some caution, as suggested by Whittington, Owen-Smith, and Powell (2009):

> 'These results demonstrate that location in one of three key clusters is a source of advantage but show that these effects depend on the degree to which an organization is centrally connected, both within and beyond a home region. (...) Propinquity and centrality are both wellsprings of innovation, however, they offer at least partially exclusionary advantages. Thus understanding

> the effects of networks on innovation requires concern with physical location, and vice versa' (pp. 114-115)

Hence, we should avoid taking the effect of space as autonomous. Other authors refers to knowledge spillovers to explain spatial effects detected (Marrocu, Paci, and Usai, 2013; Ponds, Oort, and Frenken, 2010). For what concerns technological dimension, this proximity seems to play a manly positive role in innovation (Balsmeier, Buchwald, and Stiebale, 2014; Li, Qiu, and Wang, 2019), even though it is not always significant (Isaksson, Simeth, and Seifert, 2016; Liang and Liu, 2018). Lastly, the other dimensions are far less investigated. Noteworthy is the disagreement over organizational proximity influence, which shows manly negative results (Capaldo and Petruzzelli, 2014; Dolfsma and Van der Eijk, 2016). Empirical evidences are summarized in Table 2.3.

Product & Process Innovation

This group of studies is composed by seventeen articles, of which twelve investigate both product and process innovation, four only product innovation, and one only process innovation. For what concerns management conversations, the most represented topic is *Technology & Production* (Topic 3) with a count of seven studies. Also, four studies deal with *Network & Embeddedness* (Topic 37), three with *Local & Distant* (Topic 46), two with *Portfolio of Alliances* (Topic 6) , and one with *Knowledge & Learning* (Topic 36). Of these studies, fifteen employed a quantitative design, a single study a qualitative one, and one offers a theoretical contribution.

Articles dealing with *Technology & Production* are generally interested on the effects of spatial determinants on the innovative performance. In so doing, authors explore knowledge intensive business services (Brunow, Hammer, and McCann, 2020; Shearmur and Doloreux, 2015), industrial clusters dynamics (Callois, 2008; Molina-Morales and Martínez-Fernández, 2010), the space of university-industry collaborations (Maietta, 2015), the relation between space and firms characteristics (Presutti, Boari, Majocchi, and Molina-Morales, 2019), and more in general the influence of face-to-face contact on firms outcome (Weterings and Boschma, 2009). On the other hand, articles with an interest in *Network & Embeddedness* discuss the characteristics of actors' relationships as influenced by proximity dimensions, and the consequent effect on innovation (Hall and Jacobs, 2010; Huggins and Johnston, 2010; Jespersen, Rigamonti, Jensen, and Bysted, 2018; Oerlemans and Meeus, 2005). Also, authors closer to the *Local & Distant* conversation deals with different layers of actors' location (Doran, Jordan, and O'Leary, 2012; Rammer, Kinne, and Blind, 2020), and influence of foreign firms on locals (Wang and Wu, 2016). Lastly, researches dealing with *Portfolio of*

Table 2.3: Effect detected on technological innovation

Studies	Geo	Org	Cog	Tech	Soc	Inst
Ardito, Messeni Petruzzelli, Pascucci, and Peruffo, 2019	*NS*			*NS*		
Balsmeier, Buchwald, and Stiebale, 2014				+		
Capaldo and Petruzzelli, 2014	+	−				
Chu, Tian, and Wang, 2019	+					
Dolfsma and Van der Eijk, 2016	*NS*	Bus.: *NS*	+		*NS*	
		Hier.: −				
Funk, 2014	+					
Geerts, Leten, Belderbos, and Van Looy, 2018	+					
Isaksson, Simeth, and Seifert, 2016				*NS*		
Leten, Landoni, and Van Looy, 2014	+					
Li, Qiu, and Wang, 2019	+			+		
Liang and Liu, 2018	−			*NS*		*NS*
Link and Scott, 2003	*NS*					
Marrocu, Paci, and Usai, 2013	+	+		+	+	
Owen-Smith and Powell, 2004	+					
Parent and Riou, 2005	±					
Ponds, Oort, and Frenken, 2010	+ | *NS*					
Schwartz, Peglow, Fritsch, and Günther, 2012	*NS*					
Van de Vrande, 2013				∩		
Whittington, Owen-Smith, and Powell, 2009	*NS* | +					

Note: for consistency, the effect of those articles investigating 'distance' and not 'proximity' has been reversed. For what concerns symbols, + stands for significant and positive, − significant and negative, ± significant but conflicting, ∩ for inverted-U shaped relation, and *NS* not significant. Proximity dimensions treated as control variables are not included.

Alliances discuss collaboration with geographical heterogeneous private or public partners (Beise and Stahl, 1999; Fitjar and Rodríguez-Pose, 2011), while the article facing *Knowledge & Learning* topic talks about knowledge spillovers and transfer (Fernandes and Ferreira, 2013).

For what concerns results, it is not clear weather geographical proximity has an influence or not. For example, Fitjar and Rodríguez-Pose (2011, p. 1264) find a non significant effect of closeness in product/process innovation, thus concluding:

> 'Firms that develop international partnerships are likely to innovate, firms that rely on national and local interaction are not, meaning that the transfer mechanisms of knowledge and innovation within close geographical proximity are either broken or less prominent than previously thought.'

An even more negative conclusion on the effect of spatial closeness is pointed out by Presutti, Boari, Majocchi, and Molina-Morales (2019), who claim a lock-in effect able to isolate firms. On the other hand, authors such as Fernandes and Ferreira (2013) or Rammer, Kinne, and Blind (2020) provide a somewhat different conclusion. In so doing, it emerges a weak equilibrium result which recalls the Boschma (2005) idea of a spatial proximity that is neither sufficient nor necessary (Weterings and Boschma, 2009). Table 2.4 summarizes the results of this sub-group.

Innovation Novelty

This sub-sample is composed by twelve articles dealing with innovation novelty mainly through quantitative means (only two articles take a qualitative angle). For what concerns managerial conversations, seven articles discuss the *Technology & Production* theme (Topic 3), while only two deal with *Network & Embeddedness*. Also, *Portfolio of Alliances* (Topic 6), *Trust & Collaboration* (Topic 19), and *Local & Distant* (Topic 46) are touched by one article each.

Authors interested in *Technology & Production* are generally involved in the influence of external knowledge sources and spatial characteristics in innovation (Phene, Fladmoe-Lindquist, and Marsh, 2006; Shearmur, 2011). In particular, articles discuss the effects of closeness to supplier or customer clusters (Bindroo, Mariadoss, and Pillai, 2012; Pillai and Bindroo, 2019), open innovation (Kapetaniou and Lee, 2019), recombinative innovation (Guan and Yan, 2016), and knowledge intensive business services (Doloreux and Shearmur, 2012). For what concerns *Network & Embeddedness* conversation, researchers dis-

Table 2.4: Effect detected on product & process innovation

Studies	Geo	Org	Cog	Tech	Soc	Inst
Beise and Stahl, 1999	*NS* | +					
Brunow, Hammer, and McCann, 2020	Product: + | *NS*					
	Process: +					
Doran, Jordan, and O'Leary, 2012	Product: ±					
	Process: ±					
Fernandes and Ferreira, 2013	Product: +					
	Process: +					
Fitjar and Rodríguez-Pose, 2011	*NS*					
Hall and Jacobs, 2010	*M*	*M*	*M*		*M*	*M*
Huggins and Johnston, 2010	+|*NS*					
Jespersen, Rigamonti, Jensen, and Bysted, 2018	*NS* | −	+ | *NS*	+ | −*		*NS* | +	
Maietta, 2015	Product: +					
	Process: *NS*					
Molina-Morales and Martínez-Fernández, 2010	+					
Oerlemans and Meeus, 2005	+|*NS*					
Presutti, Boari, Majocchi, and Molina-Morales, 2019	−					
Rammer, Kinne, and Blind, 2020	+					
Shearmur and Doloreux, 2015	*NS*					
Wang and Wu, 2016	+					
Weterings and Boschma, 2009	*NS*					

Note: for consistency, the effect of those articles investigating 'distance' and not 'proximity' has been reversed. For qualitative studies *M* stands for mixed proximity required. For quantitative studies, + stands for significant and positive, − significant and negative, ± significant but conflicting, and *NS* not significant. Theoretical studies and articles concerned with'other proximity' dimensions are not displayed in this table. * : here the positive symbol is for 'cognitive-technology proximity', while the negative for 'cognitive-market proximity'

cuss the influence of ego-network configuration (Knoben and Oerlemans, 2012) and boundary stretching collaborations (Enkel and Heil, 2014) on novelty. Lastly, the article dealing with Topic 6 takes care of local knowledge and global pipelines (Fitjar and Rodríguez-Pose, 2011), the one closer to Topic 19 concerns the role of trust in partner selection (Bunduchi, 2013), while the last article deals with micro-geography and knowledge spillovers from the angle of Topic 46 (Rammer, Kinne, and Blind, 2020).

For what concerns results, this sub-sample show a great heterogeneity. Indeed, it is not clear whether spatial proximity is positively associated with novelty (e.g. Bindroo, Mariadoss, and Pillai, 2012), or if it has a not significant or negative effect (e.g. Shearmur, 2011). Also, not clear results are shown by those studies dealing with other dimensions. Results are summarized in Table 2.5.

Table 2.5: Effect detected on innovation novelty

Studies	Geo	Org	Cog	Tech	Soc	Inst
Bindroo, Mariadoss, and Pillai, 2012	$+$					
Bunduchi, 2013	*M*					
Doloreux and Shearmur, 2012	Rad.Proc: $-$					
	Rad.Prod.: *NS*					
Enkel and Heil, 2014			*M*			
Fitjar and Rodríguez-Pose, 2011	*NS*					
Guan and Yan, 2016				$\cap$		
Kapetaniou and Lee, 2019	Firm: $+$					
	Mark: *NS*					
Knoben and Oerlemans, 2012	*NS*					
Phene, Fladmoe-Lindquist, and Marsh, 2006				$\pm$*		
Pillai and Bindroo, 2019	*NS*					
Rammer, Kinne, and Blind, 2020	$+$					
Shearmur, 2011	Rad.Proc.: $-$					
	Rad.Prod.: *NS* | $+$					

Note: for consistency, the effect of those articles investigating 'distance' and not 'proximity' has been reversed. For qualitative studies *M* stands for mixed proximity required. For quantitative studies, $+$ stands for significant and positive, $-$ significant and negative, $\pm$ significant but conflicting, $\cap$ for inverted-U shaped, and *NS* not significant. Theoretical studies and articles concerned with'other proximity' dimensions are not displayed in this table. *: the effect of technological proximity is by construction linked to the geographical dimension.

Economic value of innovation

Seven articles are members of this last sub-group, investigating innovation performance through quantitative means. Of these, *Portfolio of Alliances* (Topic 6) and *Network & Embeddedness* (Topic 37) attracted two articles each, while *Technology & Production* (Topic 3), *Entry, Growth, & Survival* (Topic 41), and *Local & Distant* (Topic 46) are discussed by a single contribution each. In particular, articles closer to Topic 6 are concerned with stakeholders and university-industry collaboration (Li, Xia, and Zajac, 2018; MacPherson, 1998), those dealing with Topic 37 take care of knowledge networks and firms' ego-networks (Broekel and Boschma, 2012; Knoben and Oerlemans, 2012), the article with an interest in Topic 3 discusses face-to-face interactions (Weterings and Boschma, 2009), the one closer to Topic 41 pre-entry working experiences of the founder (Weterings and Koster, 2007), and the last article deals with Topic 46 showing interest toward regional characteristics and their influence on innovation.

For what concerns results, geographical proximity is either not significant or it exerts a positive influence over the economic value of innovation (MacPherson, 1998; Weterings and Koster, 2007). A single study deals with multiple dimensions of proximity, showing a positive influence of social proximity, a non significant effect of organizational proximity, and a negative effect of cognitive dimension (Broekel and Boschma, 2012). Table 2.6 summarizes the results found.

Table 2.6: Economic value of innovation novelty

Studies	Geo	Org	Cog	Tech	Soc	Inst
Broekel and Boschma, 2012	+	*NS*	−		+	
Czarnitzki and Hottenrott, 2009	+ | *NS*					
Knoben and Oerlemans, 2012	*NS*					
Li, Xia, and Zajac, 2018	+					
MacPherson, 1998	+					
Weterings and Boschma, 2009	*NS*					
Weterings and Koster, 2007	*NS*					

Note: for consistency, the effect of those articles investigating 'distance' and not 'proximity' has been reversed.+ stands for significant and positive, − significant and negative, ± significant but conflicting, and *NS* not significant. Results concerning 'other proximity' dimensions are not displayed in this table.

2.6.2 Network

In the sample, eighteen studies show an interest toward proximity influence over tie formation in innovation related projects. In so doing, nine studies jump into the conversation of *Network & Embeddedness* (Topic

37), four of *Technology & Production* (Topic 3), and two of *Local & Distant* (Topic 46). The last three articles discuss *Portfolio of Alliances* (Topic 6), *Business & Entrepreneurship* (Topic 33), and *Knowledge & Learning* (Topic 36) respectively. Of these studies, eleven employed a quantitative design, five a qualitative one, and one offers a purely theoretical momentum.

Researchers dealing with *Network & Embeddedness* focus on individual networks (Crescenzi, Nathan, and Rodríguez-Pose, 2016; Ter Wal, 2014), firms knowledge collaborations (Moodysson and Jonsson, 2007), ties between public research organizations and firms (Drejer and Ostergaard, 2017; Steinmo and Rasmussen, 2016), and collaboration at cluster level (Giuliani, 2007; Lazzeretti and Capone, 2016; Letaifa and Rabeau, 2013). Also, some of them describe proximity influence over tie formation through time (Crescenzi, Nathan, and Rodríguez-Pose, 2016; Lazzeretti and Capone, 2016; Ter Wal, 2014). In this conversation, Boschma (2005) contributes uncovering theoretical mechanisms of proximity influence over coordination. On the other hand, authors interested in *Technology & Production* topic, discuss factors influencing collaboration relationships on technological knowledge (Cantù, 2010), firms' international expansion of R&D and technology (Zeller, 2004), cooperation on technological capability development (Divella, 2017), or the relationship between innovation-related external linkages and internal resources (Freel, 2003). Furthermore, both articles aligned with *Local & Distant* debate are concerned with the influence of physical distance in innovative linkages (Morescalchi et al., 2015; Sternberg, 1999). Lastly, the article closer to the *Portfolio of Alliances* topic discusses the influence of space and university quality in firms' collaboration choices (Laursen, Reichstein, and Salter, 2011), the one closer to *Business & Entrepreneurship* discusses collaborative or individual entrepreneurship, and the one interested in *Knowledge & Learning* talks about knowledge spillovers and collaboration with higher education institutions (Freel, 2003).

For what concerns research findings, authors tend to agree on the relevance of geographical proximity for tie formation, but disagree on the effect. On one side, Morescalchi et al. (2015, p. 661) state:

> 'Contrary to the widespread notion that the importance of distance has been decreasing over time due to globalization and technological advancement, our results show that the constraint imposed by geographical distance on R&D inter-regional links seem to have actually increased in three of the networks analyzed'

Prior to this, Zeller (2004) highlighted the geographical selectivity of knowledge flow, and the tendency of rivals to 'fight over privileged access to the spatially concentrated technological bases' (p. 105). Addi-

tionally, from an evolutionary perspective, a group of studies shows the greater influence played by spatial closeness in early stages, when knowledge is mainly tacit, consequently diminishing on later ones (Cantù, 2010; Lazzeretti and Capone, 2016; Ter Wal, 2014). However, taking a critical stance from the Marshallian idea of 'industrial atmosphere', Giuliani (2007) describes the 'uneven and selective' (p. 162) distribution of knowledge inside clusters[8]. Hence, she highlights the scarce relevance of closeness in the absence of a strong knowledge base. Similarly, Letaifa and Rabeau (2013) shows the inability of geographical proximity to foster collaborative ties. Also, Freel (2003) points out that 'novel innovator' are characterized by a spatially reach network of collaborations. Apart from geographical proximity, authors seem to agree on a positive association of social, organizational, and cognitive dimensions with collaboration ties (Crescenzi, Nathan, and Rodríguez-Pose, 2016; Lazzeretti and Capone, 2016; Ter Wal, 2014). Also, some evidences of a positive association of technological (Cantù, 2010) and institutional (Lazzeretti and Capone, 2016) proximity emerge. Table 2.7 summarizes the empirical results of this sub-group.

2.6.3 Process

Thirteen studies show an interest toward the relationship of proximity dimension and processes of innovation. In particular, eight disentangle effects of closeness on diffusion and adoption, while five on several innovation phases. For what concerns management conversations, *Technology & Production* (Topic 3) and *Network & Embeddedness* (Topic 37) are discussed by three articles each, while *Corporate Governance* (Topic 9) and *Knowledge & Learning* (Topic 36) by two articles. Also, *Entry, Growth & Survival* (Topic 41), *Markets* (Topic 45), and *Local & Distant* (Topic 46) find their space in this sub-group. Of these studies, six employed a quantitative design, three a qualitative one, two mixed qualitative and quantitative evidences, and two articles offer a theoretical framework of analysis.

Authors closer to *Technology & Production* conversation discuss diffusion of technological process (Baptista, 2001), development of technological capabilities (Divella, 2017), and university-industry cooperation in relation with different phases of the R&D cycle (Broström, 2010). Those with a foot on *Network & Embeddedness* cover the temporal dimension of co-location (Torre, 2008), bench-marking and emulation (Still and Strang, 2009), and the relationship between proximities and type of knowledge produced (Davids and Frenken, 2018). Authors interested on *Corporate Governance* depict mechanisms of diffusion of corporate practices (Davis and Greve, 1997; Mohliver, 2019), while those closer to *Knowledge & Learning* dig into the spatial dimension of learning and knowledge transfer (Moodysson, 2008; Weidenfeld, Williams,

[8]Giuliani (2007) does not directly test the influence of proximity on knowledge network, but it is an indirect result given by her research design as discussed at page 162

Table 2.7: Effect detected on collaboration and tie formation

Studies	Geo	Org	Cog	Tech	Soc	Inst
Cantù (2010)	I^+		I^+	I^+		
Crescenzi, Nathan, and Rodríguez-Pose (2016)	$\pm$	$+$	$\pm$		$+$	
Divella (2017)	$\pm$					
Drejer and Ostergaard (2017)	$+$					
Fernandes and Ferreira, 2013	$+$					
Freel (2003)	$\pm$					
Giuliani (2007)	*NI*					
Laursen, Reichstein, and Salter (2011)	$\pm$					
Lazzeretti and Capone (2016)	$+$	$+$	$+$		$+$	$\pm$
Letaifa and Rabeau (2013)	I^-				I^+	
Letaifa and Goglio-Primard (2016)		I^+			I^+	I^+
Morescalchi et al. (2015)	$+$					
Steinmo and Rasmussen (2016)	*M*	*M*	*M*		*M*	
Sternberg (1999)	$\pm$					
Ter Wal (2014)	$+$				$+$	
Zeller (2004)	I^+	*M*		*I*	*M*	

Note: for consistency, the effect of those articles investigating 'distance' and not 'proximity' has been reversed. For qualitative studies I^+ stands for positive influence, I^- stands for negative influence, *NI* no influence, and *M* for mixed influence. For quantitative studies, $+$ stands for significant and positive, $-$ significant and negative, $\pm$ significant but conflicting, and *NS* not significant. Theoretical studies and articles concerned with'other proximity' dimensions are not displayed in this table.

and Butler, 2010). Lastly, articles dealing with topic 41, 45, and 46 discuss respectively the adoption of corporate venture capital programs (Gaba and Meyer, 2008), the process of market intelligence in innovation (Cornish, 1997), and the influence of international mobility on innovation diffusion (Weidenfeld, 2013).

For what concerns results, among articles dealing with adoption and diffusion, a positive influence of geographical proximity prevails. For example, Baptista (2001, p. 43) stated that 'the diffusion of process innovations tend to grow stronger as the geographical unit of reference becomes smaller'. A similar conclusion is depicted by Gaba and Meyer (2008, p. 992), who highlight that 'adoption by peers in close geographic proximity to a focal firm increases its probability of adoption'. Also some corporate strategies seem to follow some spatial logic in their diffusion (Davis and Greve, 1997; Mohliver, 2019). On the other hand, Still and Strang (2009) show that closer firms are not more likely than others to be taken as benchmark, and Divella (2017) finds a not significant relationship (or sometimes positively related to distance) between partner location and adoption. For what concerns studies that investigate differential influence of proximity in innovation phases, results seems to scantly converge. This divergence could be associated with the differences in innovation stages investigated. In general, it emerges a complex influence which varies phase by phase. Lastly, only Davids and Frenken (2018) deal with multiple proximity dimensions. Table 2.8 provides an overview of the studies discussed.

2.7 Discussion

The following section aims to join the different components discussed in previous sections, offer answers to my research questions, and provide some guidance on future actions to take. In so doing, Fig 2.7 summarizes the relationships investigated in the sampled articles, and it is a useful reference for the following discussion.

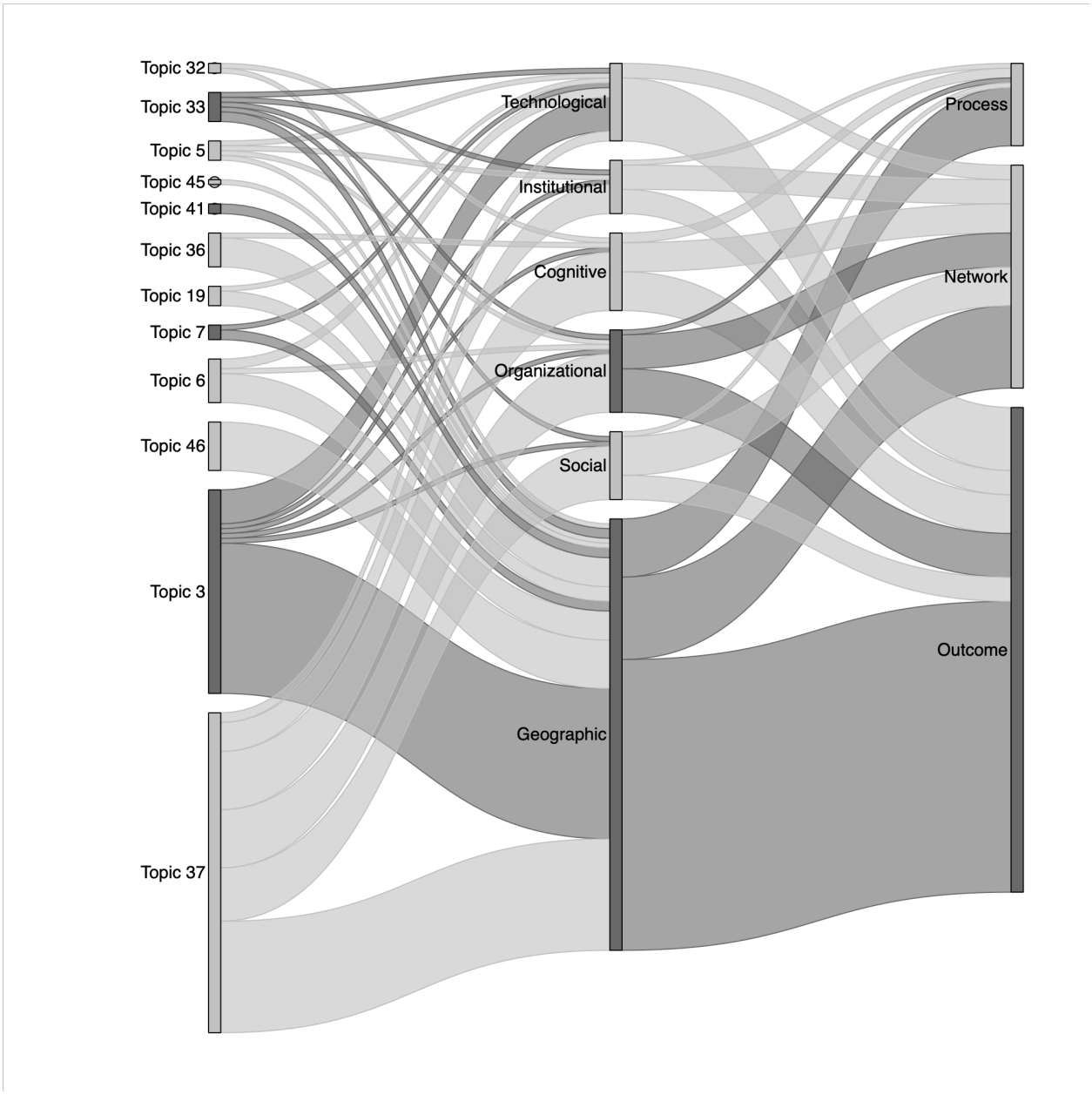
Some introductory and general considerations are necessary. First, even though the multi-dimensional essence of proximity seems to be conceptually well-established, it is not in practice. Hence, less than a third of the articles deals with more than one dimension, and the vast majority focuses only on the geographical one. Here, two main explanations are possible: (i) either there is no real agreement among management scholars on the necessity to have multiple dimensions, or (ii) ambiguities are so wide to prevent researchers from actually using them. Also, as a third explanation, this could be ascribed to difficulties in retrieving data, which — in the case is true — should be in part smoothed by the enormous amount of data offered by the Internet. Second, disagreement is not on definitions, but more on measures, except technological proximity. This is true also for geographical proximity, in which translation into numbers is sometimes

Table 2.8: Effect detected on processes of innovation

Studies	Geo	Org	Cog	Tech	Soc	Inst
Baptista (2001)	$+$					
Broström (2010)	Early: $+$					
	Late: $+\|NS$					
Cornish (1997)	Early: I^+					
	Late: M					
Davids and Frenken (2018)	Research: L	L	H		M	H
	Develop.: H	H	L		M	H
	Market: H	L	L		M	H
Davis and Greve (1997)	$+\|NS$					
Divella (2017)	$-\|NS$					
Gaba and Meyer (2008)	$+$					
Mohliver (2019)	$+$					
Moodysson (2008)	Brainst.: H					
	Prob. Solv.: L					
	Des. & Redes.: M					
	Diss.: L					
	Prot.: H					
	Comm.: L					
Still and Strang (2009)	NS					
Weidenfeld, Williams, and Butler (2010)	I^+					

Note: for consistency, the effect of those articles investigating 'distance' and not 'proximity' has been reversed. For qualitative studies I^+ stands for positive influence, H for high proximity, L low proximity, and M for mixed proximity required. For quantitative studies, $+$ stands for significant and positive, $-$ significant and negative, $\pm$ significant but conflicting, and NS not significant. Theoretical studies and articles concerned with'other proximity' dimensions are not displayed in this table.

Figure 2.7: Beaten paths and open routes



Notes: the size of lines does not refer to the count of studies in each set, but the count of relations each set has with the others. For example, an article of topic 37 discussing both geographic and technological proximity will generate two distinct relations to be represented.

less straightforward than its definition. Lastly, even though great methodological advancements have been achieved and applied to proximity and innovation research, conceptual inconsistencies remain that hinder these new methods' potential explanatory power.

Keeping in mind these considerations, I can now answer the three research questions moving this article. First, proximity and innovation scholars have mainly contributed to two management topics, namely *Technology & Production* (Topic 3) and *Network & Embeddedness* (Topic 37). Other topics arousing some interest are *Local & Distant* (Topic 46), *Knowledge & Learning* (Topic 36), and *Portfolio of Alliances* (Topic 6). This first representation of authors' interests suggests too wide attention placed on outcomes more than on reasons of that influence. Second, scarce attention has been paid to mechanisms, leaving theoretical claims as for 'granted'. This is a great gap hindering the scholars' discussion over proximity influence, with few articles analyzing in detail 'how' and 'why' each dimension matters. The risk is to focus only on regression signs and p-values, speculating on results without sound empirical evidence. Third, given scarce theoretical clarity, the consequent empirical contribution remains limited. This also emerges from the heterogeneous results discussed in the previous section. Moreover, given the scarce multi-dimensional attention, we can discuss the stratification of results only for the geographic dimension. This has a role for sure, but it can not be considered as an 'independent' effect (Boschma, 2005; Giuliani, 2007; Whittington, Owen-Smith, and Powell, 2009).

In light of these findings, two main recommendations for further research arise. First, we should re-focus the discussion on each proximity dimension's theoretical roots, leveraging that multidisciplinary background to further enhance the debate. In so doing, it is needed to explore further the relations with other research streams, such as the homophily literature (McPherson, Smith-Lovin, and Cook, 2001). The proximity scholars' engagement with economic sociology has always been beneficial, as some later contributions prove (Balland, Boschma, and Frenken, 2015). Also, as Fig 2.7 shows, scarce attention has been paid to some crucial management conversations, such as *Trust & Collaboration* (Topic 19), or not enough to others, such as *Knowledge & Learning* (Topic 36). A greater discussion over these topics could further enhance the proximity framework. Second, a greater emphasis on empirical studies (both qualitative and quantitative) to uncover and prove how each dimension affects innovation would be essential to foster a productive discussion.

2.8 Conclusion

The multidisciplinary nature of proximity literature is both a strength and a weakness, requiring writers to synthesize a great variety of knowledge in vectors representing specific characteristics of dyads. In the last years, always more interest has risen over the proximity topic, with contributions growing in numbers. This large amount of theoretical and empirical works contributed to generate a diffused confusion. In such a scenario, this article tries to offer the following contributions. Firstly, it provides a summary of fundamental dimensions and measurements to uncover not only operational but also theoretical mismatches. Secondly, it defines the thematic boundaries as nested in the broader managerial literature, thus identifying the audience to which each contribution is directed. Thirdly, it reports both established and less prominent claims over the relationship between proximity and innovation, with a particular focus on crucial theoretical inputs. Also, it tries to highlight the contribution offered to literature with a great effort of synthesis of empirical results. Furthermore, it reports some suggestions for issues that require the whole academic attention to enhance a community discussion. Lastly, the appendix of this chapter offers a collection of all measures employed by the sample. Nonetheless, this chapter comes with some limitations. First, the synthesis required by this research has inevitably imposed the loss of more rich information contained in the retrieved articles. Second, such a research design's subjectivity has been smoothed only partially by applying the topic modeling technique. Lastly, the narrow attention to top journals' managerial contributions is mainly due to the writer's limited capacity.

A final reflection on the steps I believe necessary to advance the proximity framework is here sketched. A great absence from the articles reviewed is a sound discussion on both ontology and epistemology. The dearth of debate on the former generated literature full of variance contributions and almost nonexistent process ones. Whereas, in the case of epistemology, a clear hegemony of neopositivists stances seems to impede other scientific paradigms to take part in the debate on proximity. In so doing, proximity literature is stuck in a deterministic approach to science, that conceives change as separate and almost independent states. If we want to create momentum, we need a more inclusive field not just on multidisciplinarity. At the same time, we need to reflect on the assumptions our field bears. Questioning assumptions, critically reflecting on what resonates with society & science advances and what does not, it is a must. Then, we need to ask more "how" and "why". Understanding mechanisms underlying the observed effects to really test our theories and, perhaps, discard them.

Unfortunately, the next chapter will disappoint the expectations expressed therein about how to ad-

vance this field of research. Nonetheless, the chapter introduces a mediator who can explain, albeit partially, how proximity can count in the innovation process. In particular, it leverages on a strategic argument, that is the business model, which links themes such as cognition, learning, and embeddedness. These topics can act as bridges and advance our understanding of the proximity effect within companies strategizing.

This page is intentionally left blank.

Chapter 3

The missing link between Proximity and Innovation: An exploratory study on the role of Business Model

Abstract

This exploratory study focuses on the link between proximity and innovation, investigating the business model's role in mediating this relation. In particular, the chapter looks at whether geographical, cognitive, and social proximity influence business model, and if the latter can mediate proximities' effects over radical and incremental innovation. These relationships are explored in a sample of 198 Tuscan firms. Cognitive proximity shows a positive influence on both novelty- and efficiency-centered business model, social proximity positively influences the choice of efficiency-centered models, and geographical proximity harms the novelty-centered model. Furthermore, the econometric analysis highlighted evidence of mediation effects, with both model designs showing positive influence over radical innovation. Results are explained in the light of transaction cost theory and absorptive capacity.

Keywords: Proximity, Business Model, Radical Innovation, Incremental Innovation.

3.1 Introduction

Identifying the effect of proximity on innovation has been a goal that has put together scholars from different disciplines, such as economics (Broekel and Boschma, 2012; Ter Wal, 2014), sociology (Owen-Smith and Powell, 2004; Whittington, Owen-Smith, and Powell, 2009), and management (Funk, 2014; Lazzeretti and Capone, 2016). In so doing, an astonishing amount of knowledge has been offered to academia and practitioners, as shown in Chapter two. This research tries to contribute to this debate by introducing a strategy topic to explore the influence of proximity on innovation, thus focusing on the business model's role.

Proximity literature was born in the local systems community, to which several management scholars have directly or indirectly participated (Dahl and Sorenson, 2012; Porter et al., 1998). This debate is still catching a lot of attention from top management journals (Chu, Tian, and Wang, 2019; Crescenzi, Nathan, and Rodríguez-Pose, 2016; Mohliver, 2019). Here, the central insight is that different dimensions of closeness between partners do influence innovation. This influence is not necessarily positive as the effects of lock-in (Boschma, 2005) and proximity paradox (Broekel and Boschma, 2012) have shown. On the other side, business model research has its roots in strategy literature, taking into account less discussed topics by mainstream strategy scholars (Massa, Tucci, and Afuah, 2017). As demonstrated by an exponential increase of contributions over the years, this is still a hot topic (Klang, Wallnöfer, and Hacklin, 2014; Massa, Tucci, and Afuah, 2017; Wirtz, Pistoia, Ullrich, and Göttel, 2016). Noteworthy, the business model can be both a determinant (Hu and Chen, 2016) and a dimension (Foss and Saebi, 2017) of the firm innovation effort. However, for the sake of clarity, this chapter treats the business model as a determinant of innovation and as an 'attribute of a real firm' (Massa, Tucci, and Afuah, 2017).

Proximity and Business Models are not independent. Indeed, a business model comprises those activities performed 'either within the firm, or beyond it through cooperation with partners, suppliers or customers' (Zott and Amit, 2010, p. 217); thus, it 'must link the workings inside the firm to outside elements' (Baden-Fuller and Mangematin, 2013, p. 413). Simultaneously, proximity literature is all about cooperation with partners, focusing on geographical, cognitive, organizational, institutional, and social closeness (Boschma, 2005). Hence, studying the firm's 'model' paying substantial attention to partners' characteristics — or 'outside elements' — it is likely to offer novel insights on firms' ability to innovate. In so doing, this research aims to explore the following question: *Can the business model mediate the relation between proximity and innovation?*

Among the different ways in which proximity may play a role in the innovation process, this chapter decided to take as the units of analysis firms and their business model. The choice to focus on firms reflects the literature acknowledgment for firms' heterogeneity and the consequent uneven knowledge distribution within spatial agglomerations (e.g., Giuliani, 2007; Giuliani and Bell, 2005). Firms' heterogeneity depends on both internal and external resources and capabilities, that determine their short- and long-term competitive advantage (Barney, 1991; Teece, Pisano, and Shuen, 1997). Within firms, the business model translates those internal and external elements into value and sets the foundations for value capturing (Zott and Amit, 2010). In this chapter, I am especially interested in value creation and, in particular, to explore whether closeness enables firms to achieve different degrees of novelty, that is radical or incremental innovation. As discussed in the previous chapter, the relationship between proximity and innovation novelty displays a great heterogeneity of results, leaving wide margins of contribution.

To investigate my research question, a structured survey has been elaborated and submitted via email to a sample of Tuscan firms. Hence, a data set containing information on 198 respondents has been put together. At first, three proximity dimensions (geographical, cognitive, and social) are analyzed together with Novelty and Efficiency centered business models (Zott and Amit, 2007, 2008). Here, the regression results show a significant association between these variables. Then, radical and incremental innovation are introduced in the model, uncovering both direct and indirect influences. In particular, findings provide initial evidence for a mediation effect of business model configurations over proximity dimensions on the generation of radical innovation. Overall, the results highlight that different degrees of novelty and efficiency combinations can convey proximity effects on innovation search.

This exploratory study offers several hints to the literature. Firstly, it highlights the direct influence of local and distant ties on firms' approach to business. Secondly, it shows that not only spatial closeness, but also non-spatial dimensions can modify that approach. Thirdly, it offers a re-thinking of influences and causal links (not tested here), opening future research space. Also, the indirect effects emerging suggest that strong cooperation between firms and policymakers is required to foster actors and regional competitiveness. Indeed, business model configurations should take advantage of the critical resources offered by the local socio-economic network. Lastly, it provides an empirical analysis of two deficient research areas. Nonetheless, this exploratory study suffers from several limitations; thus, it calls for further research to confirm or reject results that emerged.

The chapter is structured as follows: section two provides a review of the relevant literature and the

conceptual model; section three discusses methodology and sample characteristics; section four provides the research findings; section five presents discussion; section six conclusions and limitations of this research.

3.2 Literature Review

The following section briefly introduces theoretical claims and empirical findings concerning proximity and innovation literature (for an extensive review on the topic, see Chapter Two). Then, some critical contributions to business model design are discussed in light of a recent literature review offered by Massa, Tucci, and Afuah (2017). The final section is devoted to introducing the conceptual model proposed.

3.2.1 Proximity and Innovation

The interest toward actors' closeness and its effect on the economy and society involved scholars since the 20^{th} century (Marshall, 1920). The debate has evolved from a simple spatial issue to a more complex and detailed classification of distance through the years. This evolution has also seen the birth of different literature streams, such as the re-conceptualization of embeddedness offered by Zukin and DiMaggio (1990) or Hess (2004). In this article, I follow the Proximity conversation as per Boschma (2005) and Knoben and Oerlemans (2006). The community of authors raising around this literature stream offered several theoretical and empirical research evidence to uncover the influence of geographical, institutional, organizational, social, and cognitive proximity with partners on economic actors' activity. In particular, a substantial contribution comes with a great effort to detail the interrelation between proximity and innovation.

The vast amount of articles from economics, management, geography, and sociology prove the excitement around this topic, but this multivocality is not without drawbacks. Dissonance and ambiguities have made conversation and debate on proximity difficult. Nonetheless, the Boschma (2005) framework represents a connective link between antecedent and posterior contributions; thus, this chapter leverages Boschma's original classification. In particular, geographical, cognitive, and social dimensions are critical variables for this research [1]. *Geographical proximity* refers to the physical distance among agents, *Cognitive proximity* represents the affinity in individual perceptions and cognition of phenomena, and *Social proximity* identifies characteristics of social bonds between actors.

[1] Two dimensions of the 2005 framework and one from Knoben and Oerlemans (2006) are not discussed. First, Institutional proximity — "associated with the institutional framework at the macro-level" (Boschma, 2005, p. 67) — is not investigated here due to the regional sample that impedes heterogeneity. Second, Organizational proximity — "the rate of autonomy and the degree of control that can be exerted in organizational arrangements" (Boschma, 2005, p. 65) — raised several theoretical and empirical ambiguities which hinder its ability to contribute. Lastly, given the overlap between technological and cognitive proximity and the great variety of actors (thus of technologies) involved, this chapter considers only the cognitive dimension.

Different mechanisms have been detailed to justify the influence of each dimension on innovation and performance from a theoretical standpoint. In particular, spatial closeness may foster knowledge transfer among actors, especially when a great degree of tacit information is involved (Ter Wal, 2014). Indeed, tacit knowledge can hardly be conveyed to other agents without face-to-face contact. Also, proximity to other actors may generate unintentional knowledge spill-overs, thus fostering diffusion (Oerlemans and Meeus, 2005). Apart from knowledge-related explanations, being close also generates benefits related to transaction costs. Closeness enables agents to cope with asymmetric information and bounded rationality (Dornbusch and Neuhäusler, 2015; Maietta, 2015). Similarly, social closeness — that has its theoretical roots within the works of Coleman (1988) and Granovetter (1985, 1973) — enables actors to decrease the risk and cost of opportunistic behaviors (Ter Wal, 2014). Indeed, being socially close, actors should develop mutual trust with a positive influence on knowledge transfer and innovation. Lastly, cognitive proximity is associated with actors' knowledge bases, thus with the theory of absorptive capacity (Cohen and Levinthal, 1990). Hence, actors with similar backgrounds should communicate and integrate their knowledge and processes more efficiently and effectively.

Empirical researches describe the effects of these dimensions on economic activities. In particular, studies on technological innovation tend to agree on actor closeness's positive influence in spatial terms (Funk, 2014; Geerts, Leten, Belderbos, and Van Looy, 2018). However, several authors highlighted (Boschma, 2005; Whittington, Owen-Smith, and Powell, 2009) that the geographical proximity effect is not autonomous; thus, it should be investigated with other variables of actors' coordination. Unfortunately, the same agreement cannot be found in studies concerning innovation novelty (Knoben and Oerlemans, 2012; Pillai and Bindroo, 2019), or economic performance of innovation (Knoben and Oerlemans, 2012; Weterings and Boschma, 2009). Hence, even though geographical closeness may foster patent development, this does not automatically translate into gains in market-novelty and profitability. In the case of social and cognitive proximity, the influence found is mainly positive (Lazzeretti and Capone, 2016; Ter Wal, 2014). However, qualitative and quantitative contributions to these dimensions are still scant.

3.2.2 Business model

Research on Business Model is continuously evolving, with new theoretical and empirical challenges to solve. This fact is shown by the last decade's several review contributions, that focus on scaling down the massive amount of information we have (Wirtz, Pistoia, Ullrich, and Göttel, 2016; Zott, Amit, and Massa, 2011), or solving ambiguities (Klang, Wallnöfer, and Hacklin, 2014; Massa, Tucci, and Afuah, 2017).

Notwithstanding some significant criticisms (Porter and Gibbs, 2001), the business model has proven its usefulness for firms' competitiveness (Massa, Tucci, and Afuah, 2017), and it has become itself an object of innovation (Foss and Saebi, 2017). As will become apparent in this research, the business model is considered an *attribute of real firms* (Massa, Tucci, and Afuah, 2017) and a driver of innovation.

Even though a great debate exists on the business model's definition (Klang, Wallnöfer, and Hacklin, 2014; Pucci, 2016; Zott, Amit, and Massa, 2011), Baden-Fuller and Mangematin (2013) highlight what scholars agree on. In particular, the business model should explain how the firm creates and monetize value, leveraging on inside and outside elements. Hence, the business model consists of a 'system of specific activities' (McDonald and Eisenhardt, 2020), which determines value creation and capture. As a system, the business model should link different components, defining what, when, how, and who performs those activities (Massa, Tucci, and Afuah, 2017). Also, viewing the business model as an attribute implies the effort to classify firms on activities performed. Among the various classification emerged (Aversa, Furnari, and Haefliger, 2015; Cucculelli and Bettinelli, 2015; Pucci, 2016), this chapter leverages on *novelty-* and *efficiency-centered* categories as per Zott and Amit (2007). These have provided useful insights into several empirical analysis (Balboni, Bortoluzzi, Pugliese, and Tracogna, 2019; Pati et al., 2018; Zott and Amit, 2008).

Novelty- and efficiency-centered business models are based on two distinct theoretical standpoints (Amit and Zott, 2001; Zott and Amit, 2007, 2008). In particular, Schumpeterian economics informs the notion of a novelty-centered business model. For Schumpeter, innovation generates change and disequilibrium, which in turn, fosters value creation (Schumpeter, 1934). Hence, novel ways of conducting economic exchanges, novel activities, or new bonds between unconnected participants define a novelty oriented business model. On the other hand, the Williamson (1975) transaction costs theory lays the ground for efficiency-centered business model. Here, reducing uncertainty, complexity, and asymmetric information is a primary goal; thus, all the activities performed by a firm to decrease those 'transaction costs' define an efficiency oriented business model.

The literature provides several pieces of evidence on how this translates into performance. In particular, Zott and Amit (2007) showed novelty-centered models to have a positive influence on firm performance (measured as stock market value) regardless of environmental conditions (e.g., resource scarcity). In the case of an efficiency-centered model, the authors highlighted a positive effect on adverse environmental conditions, but the empirical evidence is relatively weak. From a theoretical standpoint, authors justified

these effects through novelty effect and cost-reduction [2]. Similar results were obtained by Zott and Amit (2008) studying the interaction between business model designs and product market strategies. More recently, Pati et al. (2018) employed the same design classification to study the effect on firm performance (a self-reported measure of deterioration/improvement) as moderated by firm age and external environment. Also, in this case, novelty design showed a positive influence, while Efficiency's effect is not significant or sometimes negative. Here, the authors explain the latter result as a consequence of a lack of mutual trust. From an evolutionary perspective, Balboni, Bortoluzzi, Pugliese, and Tracogna (2019) provided evidence of how business model design changes affect growth performance (in terms of full-time equivalent workers). Noteworthy, the search for efficiency becomes crucial for later stages of the firm life-cycle. Lastly, Hu and Chen (2016) studied the effect of novelty- and efficiency-centered designs on technological innovation performance (self-reported measure). They found a positive and independent impact of both dimensions. In particular, they noted that *"business models focusing on efficiency can encourage firms to make full use of established and reliable existing activity systems, partnership networks, and relationship capital, while making technological innovation high efficiency and low cost"* (Hu and Chen, 2016, p. 595).

More in general, scholars agree on the relationship between business model, technological innovation, and success. Indeed, even the most incredible invention does not generate success without a good business model (Chesbrough, 2010; Teece, 2010). A famous example of this is the many commercial failures of Thomas Edison, one of the most prolific inventors of all time (Teece, 2010). In line with this, great technological leap requires firms to design and experiment novel business models, thus creating a tension between the established and the emerging model (Chesbrough, 2010). However, the model can also affect innovation. For Baden-Fuller and Haefliger (2013), the choices concerning openness and users' role influence development. Both are firms' decisions that fall within their business model design. For example, the authors discuss the case of software development in online communities or the T-shirt business of some textile firms. In both cases, "permeability of the company boundaries" (p. 423) and customers' engagement are key factors leading innovation. However, more effort "to unpick the interdependencies between business model choice, technology development, and success" is needed (Baden-Fuller and Haefliger, 2013, p. 423). This research chooses to explore one way of this interrelation, thus the influencing of business model design on technological development.

[2] "[...] efficiency enhancements that reduced their transaction costs, simplified transactions, and sped up processes." (Zott and Amit, 2007, p. 191)

3.2.3 Proximity, Business Model, and Innovation

This research aims to explore the role of the business model in mediating the influence of proximity and innovation (see Figure 3.1). As noted in the previous paragraphs, a general agreement exists among scholars on the relationship between proximity and innovation, and business model and innovation. However, apart from some spare contributions (Burt, Johansson, and Dawson, 2016; Mason and Chakrabarti, 2017), the interrelation between closeness and firms' model has been overlooked. Nonetheless, the relationship between proximity and business model is likely to emerge in at least two phases of the model design: learning and coordination.

Learning — a process not necessarily formalized but often informal — as a means to superior competitive advantage is a crucial topic of proximity research. Indeed, "what unites the different dimensions of proximity is that they reduce uncertainty and solve the problem of coordination, and, thus, facilitate interactive learning and innovation" (Boschma, 2005, p. 62). The emulation and adaptation process to peers' networks in which firms are embedded can explain the spread of strategic and organizational practices (Davis and Greve, 1997). This diffusion can often be realized through informal meetings (e.g., "club cliche"), where propinquity plays a significant role (Funk, 2014; Geerts, Leten, Belderbos, and Van Looy, 2018; Morgan, 2004). In a recent contribution, Mohliver (2019) described the spread of legally questionable reporting practices among American corporations. A remarkable finding concerns the spread mechanism of these risky practices. In particular, geographically constrained networks and trust among decision-makers were two significant diffusion factors. More in general, as Funk (2014) highlighted, proximity may enable actors to take advantage of knowledge spillovers and support from close economic agents.

In the business model literature, phases of experimentation and learning are considered crucial to design good business models. After the Xerox case's excellent analysis (Chesbrough and Rosenbloom, 2002), Chesbrough (2010) stressed the need to explore novel ways to create and capture value. In particular, the author showed that it "is not a matter of superior foresight ex-ante", but it involves both "trial and error" and "adaptation ex post" (Chesbrough, 2010, p. 356). Similarly, discussing the rivalry between Blockbuster and Netflix, Teece (2010) introduced "business model learning and adjustment". Also, Amit and Zott (2015, p. 393) added that "a business model designer can draw inspiration (i.e., borrow) by observing existing firms, or ways of organizing activities and exchanges, and by talking to investors, mentors, or colleagues who might be able to offer advice". More recently, McDonald and Eisenhardt (2020) provided empirical evidence of how startups come to design effective business models in nascent markets. Key evidence of their

research is that ventures with higher chances to build effective models are those that "treated peers as sources of ideas and resources that they could borrow to design their business models quickly and cheaply" (p.495). This process of borrowing from peers comes out to be useful for two main reasons. First, it accelerates designing an initial business model prototype, thus diminishing the time and money needed. Second, it saves resources that can be directed to improve further the model, therefore exploring viable alternatives or enhancements.

The "problem of coordination" (p. 62) is another critical topic of interest for proximity scholars (Boschma, 2005). In particular, researchers interested in actors' networks provide several insights on collaboration dynamics, thus investigating partnerships between individual inventors (Crescenzi, Nathan, and Rodríguez-Pose, 2016), firms (Lazzeretti and Capone, 2016), and firms and public research organizations (Steinmo and Rasmussen, 2016). For example, Ter Wal (2014) offered an interesting analysis of the evolution of biotech innovation networks in Germany. He showed that geographical proximity was a primary driver of collaboration at the initial stages of the network. In contrast, in later stages, partnerships were mainly formed via triadic closure [3] (social proximity). For the author, spatial closeness is essential in initial stages where knowledge is mainly tacit, while in later ones — focused on value appropriation — the development of mutual trust is fundamental to decrease the uncertainty caused by possible involuntary spillovers.

Designing a business model also requires to "link the workings inside the firm to outside elements" (Baden-Fuller and Mangematin, 2013, p. 413). In particular, Baden-Fuller and Haefliger (2013) included in their business model typology the "value delivery and link dimension". Here, the "architecture of information flows and system governance" (p. 345) describes linkages that can also transcend the traditional value chain. The importance of external elements for value creation was already noted by Amit and Zott (2001), who saw possible gains in integrating "resources, capabilities, roles and relationships among suppliers, partners and customers" (p. 496). More recently, Amit and Zott (2015) propose "stakeholders' activities" — thus collaboration — as an antecedent of business model design. Indeed, once the firm acquires knowledge of the activities that need to be performed and their complementarities, it should decide who is appointed to perform those activities. These choices are the result of a system-level analysis of activities required by the model chosen.

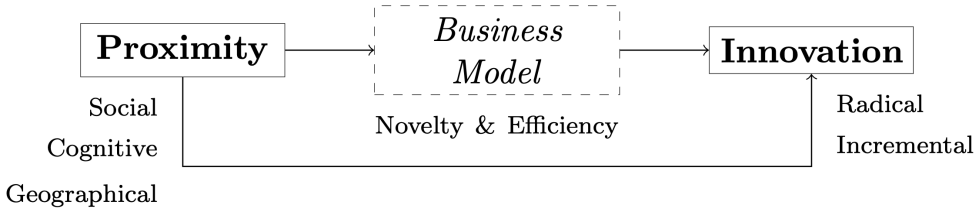
At this stage, the relationship between proximity dimensions and business model design is evident.

[3] "that partners of partners become directly connected, closing a triad (a set of three nodes) in the network" (Ter Wal, 2014, p. 590).

In particular, actors should benefit from proximity to stakeholders for both learning and coordination. In the former case, geographical and social proximity may provide greater chances to borrow from peers and to receive insightful information. However, understanding insights received from peers also requires actors to leverage on their cognitive abilities. For this reason, sharing a similar knowledge base (cognitive proximity) should enable effective communication. Hence, this learning process from proximate actors should translate into lower *trial & error costs*, thus raising the chances to design a suitable business model. In the "coordination" case, geographical proximity may reduce costs faced by the focal firm, thus offering better monitoring chances, timely feedback, and decreased productions costs (such as transportation and storage). Second, social proximity should build mutual trust among partners, thus offering chances of informal advice or market insights and decreasing uncertainty. Lastly, the focal firm should benefit from cognitive proximity, enhancing the integration of knowledge. How does this affect the choice of novelty- and efficiency-centered designs? As discussed by Amit and Zott (2015), "mindfully" borrowing from others "may increase efficiency-centered and/or novelty-centered design of the resulting business model“ (p. 340). In terms of effects, it is likely to find a correlation between geographical and social proximity with the efficiency-centered business model. Indeed, these proximity dimensions deal with the decrease of transaction costs that are pillars of an efficient model. On the other hand, being cognitively close should benefit both designs. Indeed, both exploration (novelty) and exploitation (efficiency) requires a certain degree of cognitive proximity among partners involved in value creation.

As shown in Figure 3.1, the relationship explored by this chapter is a mediation. Indeed, the business model is what makes a firm profitable (Chesbrough, 2010; Teece, 2010), not proximity per se. A firm can gain from proximity, but it needs a model to do that. However, the model design and its application are also a matter of proximity (e.g., learning from peers). Hence, this chapter aims to understand how this turns into innovative performance, thus investigating radical and incremental innovation in percentage terms over total turnover.

Figure 3.1: Conceptual Model



3.3 Methods

The research question moving this chapter has been empirically explored on a sample of Tuscan firms. Tuscany hosts 7.3% of Italian firms, employing 6.6% of the national labor force (ISTAT, 2020). Overall, the size of the Tuscan firms ranges from 1.3 to 9 employees. Noteworthy, the COVID-19 pandemic crisis and the consequent lock-down had a significant influence on the regional economy. ISTAT (2020) reports that only 46.9% of firms continued their activities during the lock-down, with a loss in terms of turnover that is above the national average. This last datum has clear relevance in terms of the following research results.

3.3.1 Data retrieval and sample characteristics

Data has been collected through a structured survey submitted to a sample of Tuscan firms at the beginning of July 2020. The survey is composed of three sections: (i) the first concerns firms' business model configurations; (ii) the second collects insights on proximity/distance of the focal firm from partners; (iii) the last section focuses on some firms' characteristics to be used as control variables. The questionnaire with a presentation letter was sent by email to a list of more than 96.405 contacts, of which only 4.353 opened the email, and 202 answered the questionnaire (redemption of 4.6% [4]). Excluding four observations coming from non-regional actors, I obtained 198 usable answers [5].

Figure 3.2 shows the count of firms by province. The large majority of respondents are located in the area of Florence (30.3%), Arezzo (14.14%), Pisa (11.6%), and Lucca (11.1%). Figure 3.3 shows the size and the year of inception of respondents. In particular, the Q_2 for the year of inception is 1997, with a mean of 1990 and a standard deviation of 22.13; while the Q_2 for the number of employees is 7, with a mean of 23.29 and a standard deviation of 61.89. The count of respondents from province area and the quartile distribution for the number of employees are in line with recent reports from national statistical bodies on Tuscan firms (e.g., InfoCamere, 2019, ISTAT, 2020). However, given the small sample (compared to the whole Tuscan population of firms) and the heterogeneity of businesses involved (based on the ATECO codes), this study can not claim any representativeness.

[4] Such redemption rate is somewhat smaller than what is achieved by other studies, but it should be considered in lights of the COVID-19 pandemic crisis.

[5] To further develop this study, the research team is still collecting responses, and it will do so until full representativeness is reached. For this chapter, I limit the sample to answers gathered during July 2020

Figure 3.2: Count of firms by province

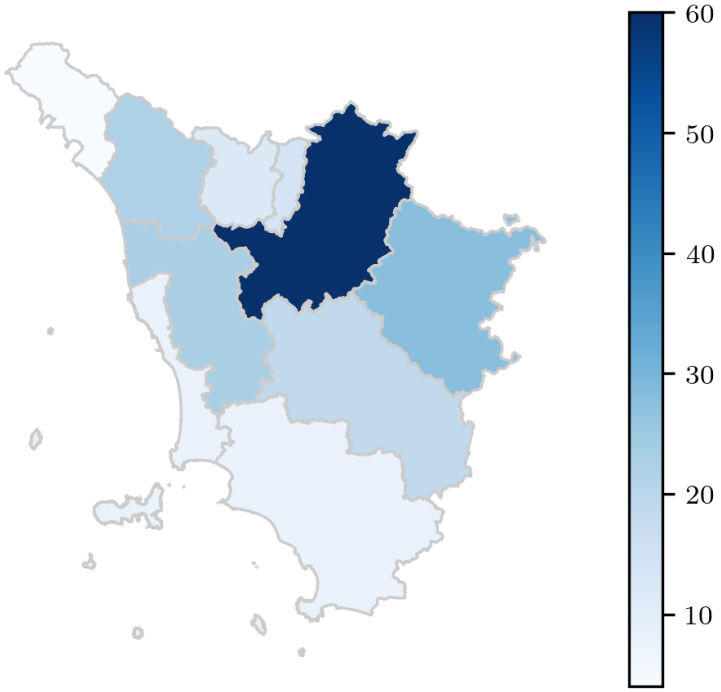
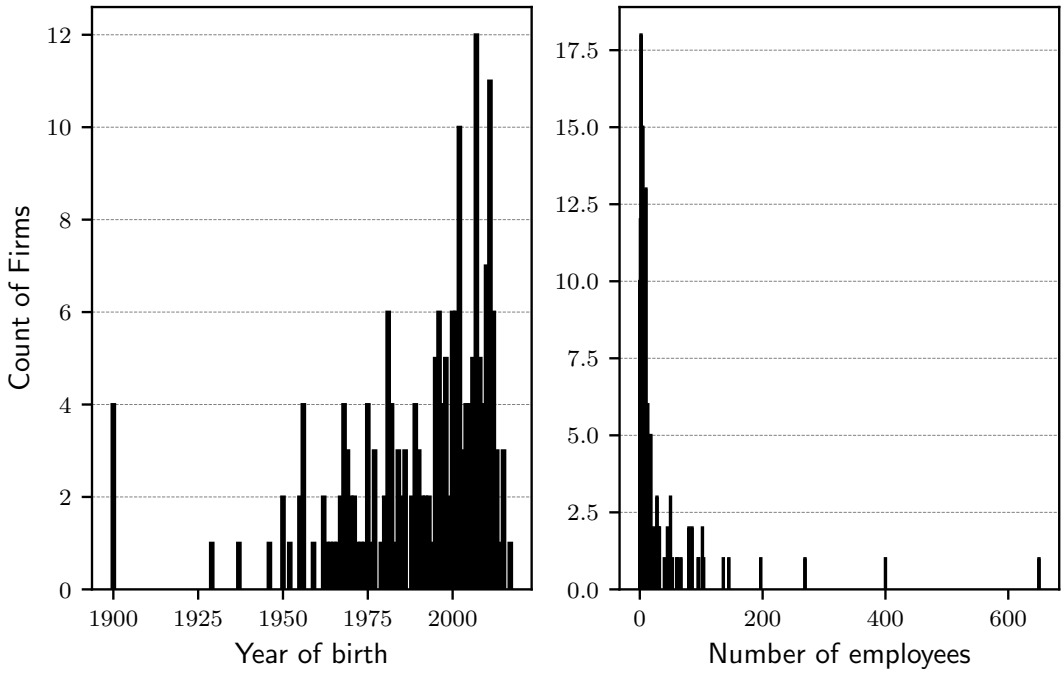


Figure 3.3: Sample Year and Size



3.3.2 Variables description

In line with previous studies, two *Dependent variables* are analyzed: radical and incremental innovation (Broekel and Boschma, 2012; Knoben and Oerlemans, 2012). In particular, firms were asked to state what percentage of their turnover has been generated either by radical or incremental innovation. This operational choice ended-up with two continuous variables bounded between 0 and 100.

For what concerns the *Mediating variables*, I relied on the well-established categorization of Efficiency and Novelty Business Model (Pati et al., 2018; Zott and Amit, 2007, 2008). In so doing, I adapted the Zott and Amit (2007) original construct, thus obtaining a 7-items scale to measure Efficiency configuration and a 5-items scale for Novelty one. In both cases, respondents are required to assess their agreement on items on a 5-point Likert scale. Table 3.1 offers a detailed description of each item, while Table C.1 in the Appendix shows factor loading and alpha values. Both constructs show values for factor loading (> 0.5), Cronbach's alpha (> 0.8), and CR (> 0.8) that meet literature requirements (Hair, Black, Babin, and Anderson, 2009).

The six *Independent variables* for this study originated in the proximity literature (Boschma, 2005; Torre and Gilly, 2000). To account for geographically proximate partners, respondents were required to state the percentage of their partners located within the region (see Bindroo, Mariadoss, and Pillai (2012) and Pillai and Bindroo (2019) for similar operational designs). In so doing, I obtained a continuous variable bounded between 0 and 100. For what concerns Cognitive and Social proximity, several items have been independently designed and tested. For each item, respondents are required to assess their agreement on a 5-point Likert scale. Hence, Cognitive Proximity items aim to assess both closeness in the knowledge-base and the ability to integrate that knowledge. While the first declination of cognitive proximity follows the Boschma (2005) definition, the second tries to capture the cognitive dimension underlying the organizational closeness claimed by Torre and Gilly (2000) and Torre and Rallet (2005). On the other hand, Social Proximity items capture three different peculiarities of relationships, namely: lasting (time duration), strength (magnitude), and personal bond (more than a business partnership). These variations are informed by the study on strong and week ties of Granovetter (1973), the study on the embeddedness of Uzzi (1997), the Boschma (2005) framework, and some recent empirical works (Dolfsma and Van der Eijk, 2016; Jespersen, Rigamonti, Jensen, and Bysted, 2018). In conclusion, the proximity literature has shown several 'interpretative shades', which translate into a great richness of definitions and measurements [6]. Hence, this

[6]The Appendix of chapter two provides clear evidence.

analysis wants to take advantage of that 'multivocality'. In so doing, no data reduction has been performed on proximity items.

Lastly, three *Control variables* are included in the model. Size is the natural logarithm of the number of employees, Age is the natural logarithm of the years since inception (at 2020), and Graduated is the percentage of employees holding a university degree over the total number of firms labor force.

3.3.3 Econometric strategy

After multicollinearity analysis, the conceptual model (see Figure 3.1) has been tested in three steps. First, the Efficiency and Novelty Business Model have been regressed over control and independent variables. Then, Incremental and Radical Innovation has been regressed over the mediators, independent variables, and controls. As a last, a mediation analysis has been performed. In the first and second steps, OLS regression has been employed. While to test mediation I followed both Baron and Kenny (1986), Preacher and Hayes (2008), and Valeri and VanderWeele (2013).

The classic approach of Baron and Kenny (1986) requires the following steps:

$$Y = \alpha_0 + \beta_0 X + \varepsilon_0 C \tag{3.1}$$

$$M = \alpha_1 + \beta_1 X + \varepsilon_1 C \tag{3.2}$$

$$Y = \alpha_2 + \beta_2 X + \gamma M + \varepsilon_2 C \tag{3.3}$$

To have mediation, the following conditions should be satisfied: (i) β_0 is significant, so the predictor (X) has an effect on the outcome (Y); (ii) β_1 is significant, so the predictor (X) has an effect on the mediator (M); (iii) γ is significant, so the mediator (M) has an effect on the outcome (Y); (iv) β_2 is not significant, so when controlling for the mediator (M), there is no effect on the outcome (Y) by the predictor (X). If all these four conditions are satisfied, there is a chance that M *fully* mediates the predictor-outcome relation. On the other hand, if step (i), (ii), and (iii) are met, but not step (iv), M is a candidate for *partial* mediation. More recently, scholar consensus has risen over criticism for step (i) (Valeri and VanderWeele, 2013). Indeed, it is still possible to have *inconsistent* mediation if the direct and indirect effects of the predictor over outcome have opposite signs. In conclusion, to have a candidate for mediation, it is at least necessary to have a significant β_1 and γ.

Once the candidate relationships are individuated, the amount of mediation should be calculated and

Table 3.1: Variables description

Variable	Item	Synopsis	Scale	Source (Adapted from)
		Innovation		
Radical		What is the percentage of turnover realized with radically innovative products?	0-100	Broekel and Boschma, 2012, Knoben and Oerlemans, 2012
Incremental		What is the percentage of turnover realized with incrementally innovative products?	0-100	
		Business Model		
Efficiency		Row mean of ebm_0 - ebm_7		Zott and Amit, 2007, Zott and Amit, 2008
α : 0.8231	ebm_0	Clients can easily interact with the focal firm.	1-5	
CR : 0.87	ebm_1	During these interactions, our business model enables a low number of errors.	1-5	
	ebm_2	Our business model does not generate additional costs for our partners.	1-5	
	ebm_3	Our business model allows us to handle small and large size transactions.	1-5	
	ebm_4	Our business model enables us to make informed decisions.	1-5	
	ebm_5	Within our firm, transactions are transparent: flows of knowledge, services, and goods can be verified.	1-5	
	ebm_6	Our business model allows fast transactions.	1-5	
	ebm_7	Overall, our business offers high transaction efficiency.	1-5	
Novelty		Row mean of ibm_0 - ibm_4		Zott and Amit, 2007, Zott and Amit, 2008
α : 0.8664 CR : 0.91	ibm_0	Our business model offers new combinations of products, services, and information.	1-5	
	ibm_1	Our business model brings together different actors or participants to shared projects.	1-5	
	ibm_2	Our business model enhances the development of innovative relationships in terms of quality and relation depth.	1-5	
	ibm_3	The firm has continuously introduced innovation in its business model.	1-5	
	ibm_4	Overall, the firm business model is novel.	1-5	
		Proximity		
Geographic		What is the percentage of partners located within the region?	0-100	Bindroo, Mariadoss, and Pillai, 2012, Pillai and Bindroo, 2019
Cognitive				Parra-Requena, Ruiz-Ortega, García-Villaverde, and Rodrigo-Alarcon, 2015, Dolfsma and Van der Eijk, 2016
	cpx_1	Our partners have similar knowledge to the focal firm.	1-5	
	cpx_2	The focal firm can easily integrate its knowledge with those of partners.	1-5	
Social				Dolfsma and Van der Eijk, 2016, Jespersen, Rigamonti, Jensen, and Bysted, 2018
	spx_1	We prefer to collaborate with partners with which a long term relationship exists.	1-5	
	spx_2	We establish strong relationships with our partners.	1-5	
	spx_3	With our partners we establish personal relationships that go beyond business relations.	1-5	

tested for significance. Different strategies have been proposed to capture the magnitude and significance of the indirect effect. Often, this is simply calculated as the product $\beta_1\gamma$, and the standard error for this product is then used to test for significance (Baron and Kenny, 1986; Valeri and VanderWeele, 2013). Another popular way to test mediation is bootstrapping, which is based on a re-sampling with replacement strategy performed several times (Preacher and Hayes, 2008). In this article, four different tests are applied to estimate each candidate relationship: coefficient product, parametric regression models[7], bootstrapping[8], and structural equation modeling (SEM). 5000 replications are applied to both bootstrapping and SEM. All computations are performed through Stata 16.0.

3.4 Results

Descriptive statistics and collinearity tests for all the variables and models included in this study are in the appendix. Hence, Table C.2 provides mean, standard deviation, min, and max, Table C.3 reports correlation coefficients, and Table C.4 displays VIF-scores and Tolerance levels for the different models tested. Low correlation coefficients (< 0.4), low VIF-scores (< 1.5), and high Tolerance levels (> 0.65) exclude any problem of multicollinearity.

Table 3.2 provides coefficients, standard errors, and p-values for Efficiency and Novelty business models regressed over controls and independent variables. R^2 and Adjusted R^2 show good levels in both cases, but Novelty appears to be better predicted by the involved regressors. Here, geographical proximity has a negative coefficient, which is significant at 5%, while cpx_2 has a positive coefficient significant at 1%. Also, both Size and Graduated variables are strongly significant. In the case of the Efficiency business model, cpx_2 has a positive coefficient significant at 1%, and spx_3 has a positive coefficient significant at 5%. All the other predictors are not significant.

Table 3.3 provides hierarchical OLS regression for Radical Innovation regressed over controls, independent, and mediating variables. Overall, the R^2 and Adjusted R^2 obtain satisfying levels. Model 2 regresses the outcome variable over controls. In this model, Graduated has a positive coefficient that is significant at 5%. In Model 3, all proximity predictors are included, but only cpx_2 has a significant coefficient. Then, Model 4 introduces the squared term for geographical proximity in order to test for any non-linear effect for this variable. Both coefficients are significant, suggesting an inverted U-shaped relation between

[7]The Stata paramed package has been employed (Emsley and Liu, 2013).

[8]The code to produce bootstrapping analysis in Stata has been retrieved by UCLA: Statistical Consulting Group at `https://stats.idre.ucla.edu/stata/faq/how-can-i-analyze-multiple-mediators-in-stata/`

Table 3.2: Regression on BM Configurations with Robust Standard Errors

	Eff BM	Nov BM
Age(ln)	0.0143	-0.0579
	(0.0611)	(0.0913)
Size(ln)	0.0514	0.165***
	(0.0344)	(0.0439)
Graduated(%)	-0.000528	0.00569***
	(0.00129)	(0.00201)
Geo. Px.	-0.00171	-0.00387**
	(0.00128)	(0.00181)
cpx_1	0.0115	0.0105
	(0.0483)	(0.0782)
cpx_2	0.173***	0.257***
	(0.0462)	(0.0683)
spx_1	0.0467	0.00207
	(0.0340)	(0.0456)
spx_2	0.0679	0.0533
	(0.0428)	(0.0665)
spx_3	0.0982**	0.0844
	(0.0400)	(0.0615)
Constant	2.505***	1.884***
	(0.353)	(0.499)
Observations	198	198
R^2	0.178	0.227
Adjusted R^2	0.138	0.190

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

geographical proximity and radical innovation. With the inclusion of this non-linear effect, both coefficients for spx_1 and spx_3 become positive and significant. Cpx_2 maintains a positive and significant effect. In Model 5, both mediating variables are introduced while the non-linear effect for geographic proximity is left out. The two mediators show positive and significant coefficients, while the other independent variables are not significant. In Model 6, the non-linear effect for geographic proximity is reintroduced. Here, efficiency and novelty business models have both a positive and significant effect, geographical proximity has a non-linear effect over radical innovation, and spx_1 has a positive and significant coefficient. This last model will be the target model for mediation analysis.

Table 3.4 provides hierarchical OLS regression for Incremental Innovation regressed over controls, independent, and mediating variables. Here, the R^2 and Adjusted R^2 show non-satisfactory levels; thus, predictors cannot explain the observed variance. As first, Model 7 regresses the outcome variable over controls, of which none is significant. In Model 8, all proximity predictors are included, but only spx_3 has a significant coefficient. Then, Model 9 introduces the squared term for geographical proximity in order to test for any non-linear effect for this variable. Both coefficients are significant, suggesting an inverted U-shaped relation between geographical proximity and incremental innovation. With the inclusion of this non-linear effect, the coefficient of spx_3 gains in terms of p-value. In Model 10, both mediating variables are introduced while the non-linear effect for geographic proximity is left out. Here, only Novelty business models and spx_3 show positive and significant coefficients, while the other independent variables are not significant. In Model 11, the non-linear effect for geographic proximity is introduced again. Both the novelty business model and spx_3 maintain a positive and significant effect. The squared geographical proximity term shows a p-value of 0.104; thus, we can consider the U-shaped relationship also confirmed in model 11. However, given the low values for R^2 and Adjusted R^2, no one of these models will be included in the following analysis.

In line with the two fundamental criteria highlighted by the mediation literature, four possible indirect effects emerge. Firstly, cpx_2 influences both efficiency and novelty business models, which in turn affect radical innovation. In this case, also the other two stringent conditions described by Baron and Kenny (1986) are met. Indeed, cpx_2 has a direct effect on radical innovation when mediators are not considered (see M 3 and M 4), and it has a not significant effect when these are included. Hence, we have a candidate for full mediation operated by two mediators, thus two indirect effects to be tested. Second, spx_3 has a significant effect on the efficiency business model, which in turn influences radical innovation. Looking at M 4 and 6, also this variable is a candidate for full mediation. Lastly, geographic proximity has a significant effect

Table 3.3: Regression on Radical Innovation with Robust Standard Errors

	M 2	M 3	M 4	M 5	M 6
Age(ln)	-2.697	-3.718	-4.152	-3.487	-3.877
	(2.690)	(2.791)	(2.679)	(2.665)	(2.593)
Size(ln)	-0.990	-0.413	-1.136	-1.643	-2.155
	(1.346)	(1.333)	(1.453)	(1.354)	(1.420)
Graduated(%)	0.161**	0.147**	0.149**	0.119*	0.123*
	(0.0695)	(0.0719)	(0.0734)	(0.0706)	(0.0727)
Geo. Px.		-0.0644	0.564**	-0.0323	0.501**
		(0.0474)	(0.242)	(0.0470)	(0.240)
cpx_1		-2.652	-1.892	-2.782	-2.123
		(1.971)	(2.062)	(1.826)	(1.894)
cpx_2		3.553**	3.431**	1.055	1.159
		(1.752)	(1.737)	(1.670)	(1.601)
spx_1		1.993	2.619**	1.690	2.248*
		(1.238)	(1.319)	(1.269)	(1.339)
spx_2		0.381	-0.363	-0.336	-0.911
		(1.698)	(1.702)	(1.550)	(1.566)
spx_3		2.184	2.554*	1.106	1.510
		(1.446)	(1.394)	(1.415)	(1.410)
Geo. Px. $\times$ Geo. Px.			-0.00614**		-0.00524**
			(0.00239)		(0.00237)
Efficiency BM				6.236**	5.730**
				(2.412)	(2.441)
Novelty BM				5.525***	5.054**
				(2.068)	(2.036)
Constant	31.90***	18.24	8.985	-7.788	-13.52
	(8.679)	(15.00)	(15.71)	(17.12)	(17.23)
Observations	198	198	198	198	198
R^2	0.048	0.101	0.144	0.172	0.202
Adjusted R^2	0.033	0.058	0.098	0.123	0.151

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4: Regression on Incremental Innovation with Robust Standard Errors

	M 7	M 8	M 9	M 10	M 11
Age(ln)	1.054	1.108	0.796	1.435	1.133
	(2.729)	(2.780)	(2.777)	(2.779)	(2.787)
Size(ln)	-0.157	0.386	-0.134	-0.364	-0.761
	(1.322)	(1.378)	(1.430)	(1.430)	(1.462)
Graduated(%)	0.0895	0.0808	0.0818	0.0504	0.0531
	(0.0708)	(0.0731)	(0.0746)	(0.0750)	(0.0767)
Geo. Px.		0.0471	0.500**	0.0638	0.477*
		(0.0572)	(0.249)	(0.0572)	(0.250)
cpx_1		0.700	1.248	0.668	1.178
		(2.246)	(2.283)	(2.170)	(2.201)
cpx_2		1.390	1.302	0.397	0.478
		(2.194)	(2.185)	(2.247)	(2.244)
spx_1		1.640	2.091	1.720	2.152
		(1.491)	(1.507)	(1.497)	(1.515)
spx_2		0.578	0.0416	0.435	-0.0107
		(2.017)	(2.046)	(1.952)	(1.978)
spx_3		3.020*	3.286**	2.775*	3.088*
		(1.621)	(1.584)	(1.642)	(1.624)
Geo. Px. × Geo. Px.			-0.00443*		-0.00406
			(0.00245)		(0.00248)
Efficiency BM				-1.944	-2.335
				(3.147)	(3.148)
Novelty BM				5.163**	4.798**
				(2.423)	(2.410)
Constant	30.26***	2.128	-4.543	-2.731	-7.176
	(8.356)	(14.44)	(14.36)	(17.03)	(16.76)
Observations	198	198	198	198	198
R^2	0.009	0.040	0.059	0.063	0.079
Adjusted R^2	-0.007	-0.006	0.009	0.008	0.019

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

on the novelty business model, which influences radical innovation. However, geographical proximity is a candidate for partial mediation given M 6.

Table 3.5 provides estimates for the indirect effects of cpx_2 via Efficiency and Novelty. In both cases, all estimates support the existence of a mediation. For what concerns the indirect path via Efficiency, the simple product test shows a positive coefficient of 0,989, which is significant at 10%. The bootstrapping method provides further positive evidence. Indeed, both percentile, bias-corrected, and bias-corrected and accelerated confidence intervals do not contain the value zero, thus confirming the significance of the effect. Further, the paramed package provides a positive and highly significant coefficient. For what concerns the indirect path via Novelty, the product test highlights a strong indirect effect of 1.3, which is significant at 5%. Also, the absence of zeros from the confidence intervals obtained through bootstrapping assures the significance of this effect. Even here, paramed results show a positive and highly significant coefficient. Both product, bootstrapping, and SEM show that the overall total indirect effect is 2.29, with a p-value lower than 0.01. Hence, I can conclude than both Efficiency and Novelty mediate the effect of cpx_2 over Radical Innovation.

Table 3.6 provides estimates for the indirect effect of spx_3 via Efficiency and geographical proximity via Novelty. In the former case, the product test highlights a positive effect of 0,562 with a slightly not significant p-value of 0,118. On the other hand, the confidence intervals obtained through bootstrapping does not contain zeros; thus, I can conclude that this indirect effect is significant. Both paramed and SEM analyses confirm the significance and positive sign of the indirect effect. Therefore, I can confirm the existence of an indirect effect of spx_3 via Efficiency. For what concerns the indirect effect of geographical proximity via Novelty, the product test shows a negative effect of -0,019 with a p-value of 0,1. Here, the bootstrapping method provides two confidence intervals that do not contain zeros and one that does. On the other hand, both paramed and SEM confirm the negative effect and its significance. Hence, I can conclude that an indirect effect exists, but with an almost negligible magnitude (see Figure C.1 in the appendix).

3.4.1 Robustness test

To test the robustness of our initial results, additional analyses have been performed. In particular, three variables were included: *Provinces*, *Sector*, and *R&D*. The former is a categorical variable that identifies the location of each firm within a Tuscan province, namely: Firenze, Massa e Carrara, Lucca, Arezzo, Prato, Pisa, Grosseto, Pistoia, Siena, and Livorno. Indeed, this research assumed that firms do not face enough institutional heterogeneity since they are located within the same Region, thus supporting the exclusion of

Table 3.5: Mediation Analysis for cpx_2

		Indirect via Efficiency		*Indirect via Novelty*		*Total Indirect effect*		
product	coefficient	0,989271		1,300347		2,289618		
	p-value	0,071		0,028		0,002		
	95% Conf. Interval	-0,0860525	2,064594	0,1373879	2,463305	0,8346693	3,744566	
bootstrap	coefficient	0,98927098		1,3003466		2,2896176		
	p-value	0,068		0,049		0,006		
	95% Conf. Interval	0,1194735	2,225699	0,2051221	2,801245	0,8667869	4,124855	(P)
		0,196727	2,413275	0,2844948	2,985255	0,9512948	4,305936	(BC)
		0,2064738	2,436702	0,2618587	2,948115	0,9596151	4,343775	(BCa)
paramed	coefficient	1,477191		1,7437369				
	p-value	0,017		0,009				
	95% Conf. Interval	0,25902907	2,695353	0,44488508	3,0425887			
sem	coefficient					2,289618		
	p-value					0,005		
	95% Conf. Interval					0,6860751	3,89316	

(P): percentile confidence interval; (BC): bias-corrected confidence interval; (BCa): bias-corrected and accelerated confidence interval

Note: Bootstrapping and SEM are performed through 5,000 repetitions. Also, the paramed Stata package does not allow to run different models on mediators and outcome variables. That is why the coefficients and p-values may diverge.

Table 3.6: Mediation Analysis for Social Proximity 3 and Geographical Proximity

		SPX_3 Indirect via Efficiency		*GPX Indirect via Novelty*		
product	coefficient	0,5624348		-0,0195734		
	p-value	0,118		0,1		
	95% Conf. Interval	-0,1421436	1,267013	-0,042866	0,0037192	
bootstrap	coefficient	0,56243484		-0,0195734		
	p-value	0,106		0,14		
	95% Conf. Interval	0,0210968	1,363964	-0,0516091	0,0003563	(P)
		0,0804842	1,546082	-0,0573997	-0,0018336	(BC)
		0,0804842	1,546082	-0,0571242	-0,0016522	(BCa)
paramed	coefficient	0,83983431		-0,0262475		
	p-value	0,06		0,074		
	95% Conf. Interval	-0,03636587	1,7160345	-0,05504695	0,00255194	
sem	coefficient	0,9890649		-0,0293738		
	p-value	0,067		0,062		
	95% Conf. Interval	-0,0685069	2,046637	-0,0601646	0,001417	

(P): percentile confidence interval; (BC): bias-corrected confidence interval; (BCa): bias-corrected and accelerated confidence interval

Note: Bootstrapping and SEM are performed through 5,000 repetitions. Also, the paramed Stata package does not allow to run different models on mediators and outcome variables. That is why the coefficients and p-values may diverge.

an institutional proximity variable. Hence, the inclusion of this regressor enables to identify any local unobserved heterogeneity. The second variable distinguishes firms between manufacturers and service providers (respectively 97 and 101 in our sample). Even though the choice between novelty- and efficiency-centered designs as per Zott and Amit (2007) classification should transcend firms' sectors, it is worth testing if any unobserved heterogeneity exists. Lastly, *R&D* is measured as the percentage of total turnover invested in research and development activities. This variable is not included in our design given the large heterogeneity of economic activities and firms' sizes; instead the *graduated* regressor has been used.

After the inclusion of Provinces' dummies, Sector dummy, and the replacement of Graduated with R&D, the results do not diverge critically from Tables 3.2, 3.3, and 3.4. In particular, in the case of efficiency centered business model, cpx_2 and spx_3 both maintained their positive signs and significant p-values (0.001 and 0.017, respectively). Noteworthy, this model shows a positive and significant p-value (0.045) of spx_2. For what concerns novelty design, cpx_2, and geographical proximity both maintained their signs and significant p-values (0.022 and 0.001 respectively). Then, the same variables were included into M 6 (radical innovation) and 11 (incremental innovation) of Tables 3.3 and 3.4. In the former model, both the curvilinear effect of geographical proximity and the positive and significant effect of novelty designs were replicated, but the variable for efficiency design is barely not significant with a p-value of 0.118. In the case of incremental innovation, the inclusion of these variables is not able to predict any variance, as in M 11.

3.5 Discussion

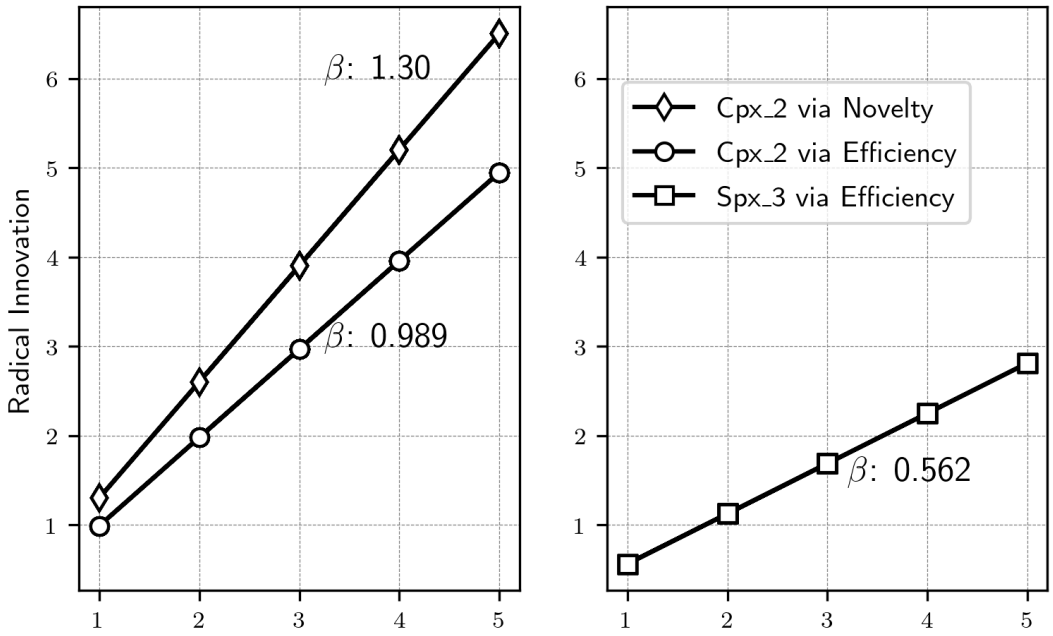
This research's empirical analysis provides initial insights on the relationship between external partner configurations, the focal firm's business model, and innovative outcome. In particular, the empirical exploration offers novel hints concerning the influence of proximity dimensions over business model designs. This alone contributes to the literature, which largely overlooked this relation (with some exceptions, such as: Burt, Johansson, and Dawson, 2016, and Mason and Chakrabarti, 2017). Results of Table 3.2 show that the ability to have partners from which the focal firm can get knowledge and integrate that resource with its endowments is a likely determinant of both configurations. Still, the relationship seems more substantial with the novelty business model. Hence, the focal firm decisions on how to do business appear related to proximity to external sources, translating into a competitive advantage. Furthermore, the design of an efficiency centered business model emerges to be related to bonds that go beyond those limited to business. Hence, those firms oriented to decreasing costs, lower errors, and — in general — to gain control over multifaceted business situations seem to rely on friendship. Indeed, the proximity literature has largely suggested

that social proximity may lower both communication and transaction costs, with a clear benefit in terms of efficiency (Crescenzi, Nathan, and Rodríguez-Pose, 2016; Ter Wal, 2014). This result is also extremely important in light of the recent pandemic crisis, where efficient management of transactions under uncertainty conditions has been a matter of resilience. As a last, the choice of a novelty oriented business model is negatively related to a high level of local partners. This insight recalls the over-embeddedness discussed by Uzzi (1997). Even though this is not necessarily detrimental in terms of innovative outcome (see Table 3.3, 3.4), it highlights a possible lock-in effect guided by an excessive regional partnership, that in the long run, may decrease the competitive ability of firms (Boschma, 2005).

The key contribution of this chapter comes in terms of indirect effects coming to light. Indeed, evidence emerges to sustain an influence of the external relational structure on radical innovation via business model configurations. Hence, not only proximity dimensions seem to influence firms' actions, but this effect appears to pass on the innovative outcomes of the focal firm, thus affecting its market strength. In particular, three main indirect paths emerged, as shown in Figure 3.4. Here, the cognitive proximity — measured in terms of the ability to integrate external knowledge — shows a strong indirect influence over radical innovation. Indeed, its effect flows through both Efficiency and Novelty business models leading to an increase of 2.29 percentage points per unit. The main driver is Novelty, which mediates 56,8% of the total indirect effect. In the case of friendship relations, Efficiency is the sole driver for the indirect effect that brings an increase of 0.56 percentage points per unit. However, taken together, these effects highlights interesting trends. Figure 3.5 shows the indirect effects for cpx_2 at different levels of spx_3 (mean and ± 1 standard deviation from the mean). In particular, the total indirect effect via Efficiency is larger than the one mediated by Novelty for medium and high levels of spx_3. In the case of low levels of spx_3, the indirect effect via Efficiency is still larger for values of cpx_2 equal to or lower than 3. This suggests two take-home points: (i) building personal relationships with partners from which the focal firm can not only receive but integrate knowledge, has a great influence on structuring efficiency oriented business models; (ii) this influence passes on the odds of focal firm to produce radical innovation more than what a novelty design can. This second point does not mean that efficiency configuration performs better than novelty in absolute terms, but that one scope should not exclude the other to fully take advantage of external relational structure. On this point, Figure 3.6 provides further insights combining on a three-dimensional graph efficiency, novelty, and radical innovation with the size of markers conditional to the number of regional partners. Here, low levels of radical innovation are excluded for clarity purposes. Overall, the graph is mainly populated by firms aiming at both novelty and efficiency (orange circles), followed by novelty oriented (violet circles) or

efficiency oriented (maroon circles), and as a last by firms with low attention to both novelty and innovation (green circles). Noteworthy, the top part of the graph (high radical innovation) is mainly populated by firms aiming at both novelty and efficiency, and by two efficiency oriented actors. In terms of circle sizes, firms with low attention to both novelty and innovation that still score high in terms of radical innovation, show a greater preference for regional partners. For other configurations, the non-linear effect described by Table 3.3 seems to emerge. In conclusion, firms performing better seem to take advantage of personal bonds in business relations and external knowledge, pushing them to design business models oriented at Efficiency and Novelty.

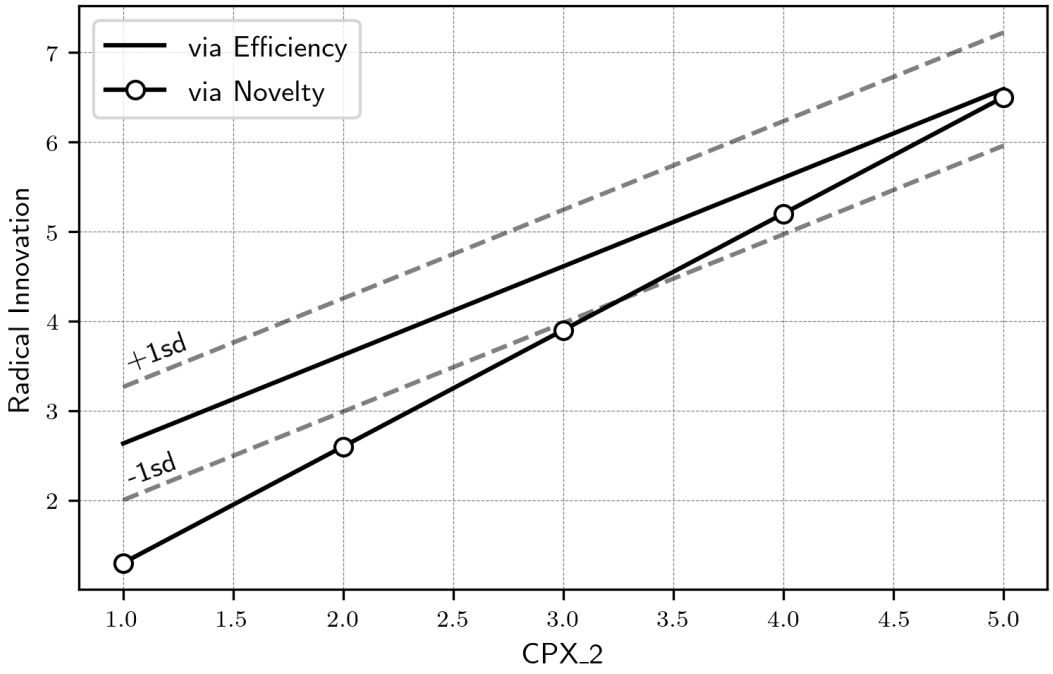
Figure 3.4: Indirect effects on Radical Innovation



Notes: the plot has been realized with Matplotlib predicting values of radical innovation by cognitive and social proximity. In so doing, boostrapped coefficients of Tables 3.5, 3.6 have been applied to observed values.

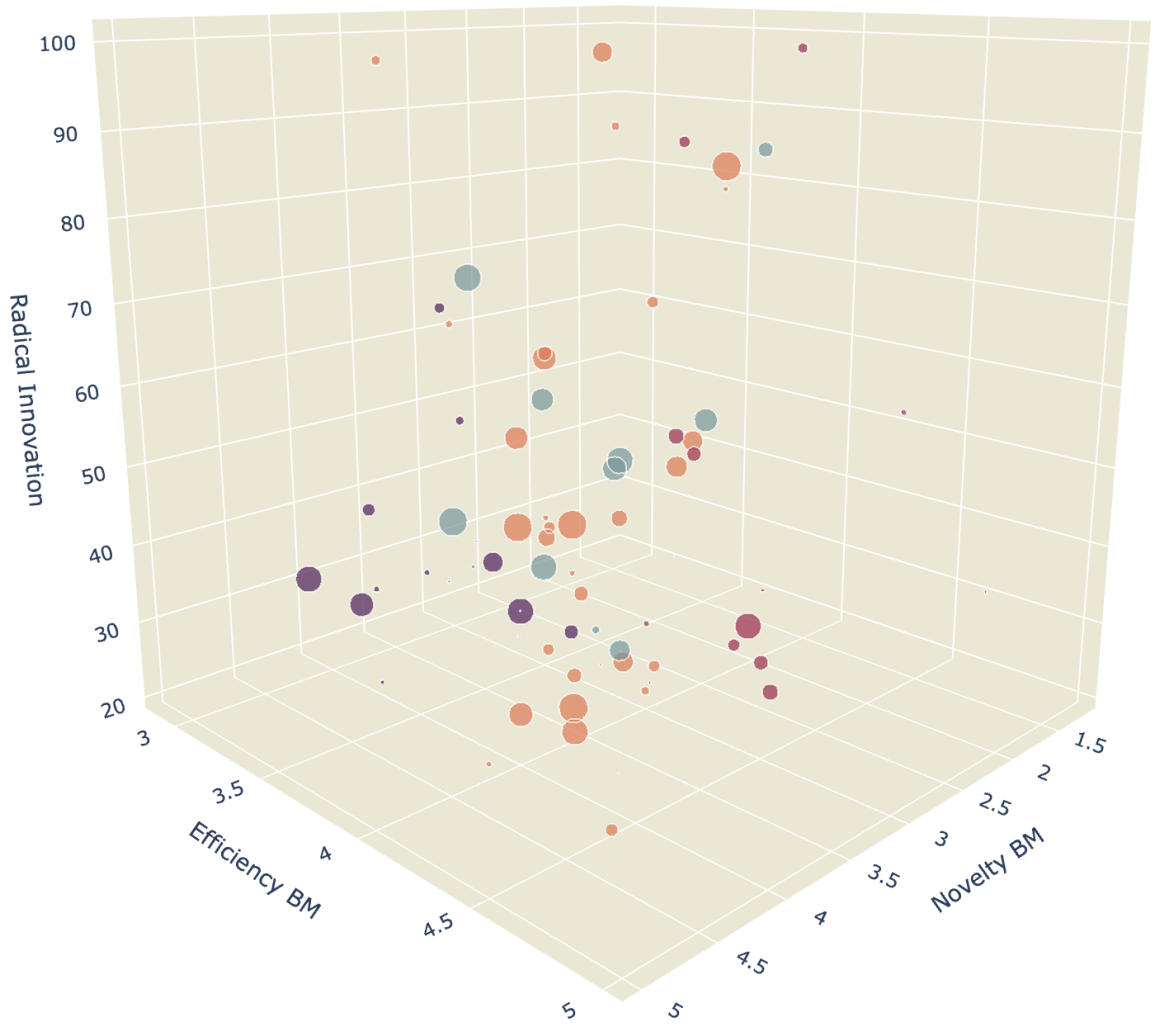
Before concluding, the geographical proximity effect deserves attention. The empirical insights have shown that relying on regional partners decreases the extent to which firms are oriented to novelty business models. Both too low and too high proximity are detrimental for radical and incremental innovation. Figure 3.7 helps to explain the non-linear effect on both innovation outcomes. In particular, two inverted

Figure 3.5: Indirect effects on Radical Innovation



Notes: the plot has been realized with Matplotlib predicting values of radical innovation by cognitive and social proximity. In so doing, boostrapped coefficients of Tables 3.5, 3.6 have been applied to observed values. To plot the indirect effect via efficiency, the following equations have been used: (1) $y = cpx_2 * \beta_{cpx_2} + [mean(spx_3) - sd(spx_3)] * \beta_{spx_3}$, (2) $y = cpx_2 * \beta_{cpx_2} + mean(spx_3) * \beta_{spx_3}$, (3) $y = cpx_2 * \beta_{cpx_2} + [mean(spx_3) + sd(spx_3)] * \beta_{spx_3}$.

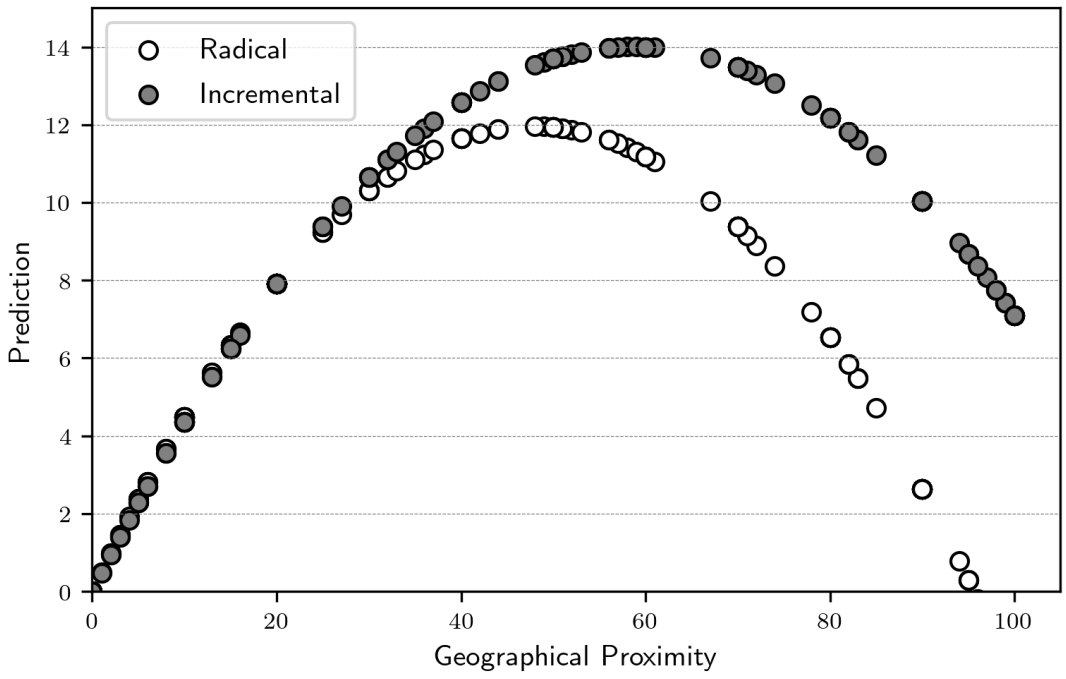
Figure 3.6: Radical innovation by Efficiency and Novelty Business Models



Notes: the 3D plot has been realized with Plotly projecting values of radical innovation by efficiency and novelty business model. The green circles stand for low novelty and efficiency business models, the maroon circles represent high efficiency but low novelty, the violet circles represent high novelty but low efficiency, and the orange circles represent high efficiency and novelty. Circle size depends on the percentage of local partners.

U-shaped curves describe the effect of geographical proximity. These have almost identical slopes for low levels of proximity, but in the case of incremental innovation, the maximum is achieved at higher levels of the predictor. Hence, incremental innovation benefits more of geographical proximity than radical, but excessive levels are detrimental for both. This curvilinear effect is in line with what theorized in other studies (e.g. Boschma, 2005), suggesting that sampled firms have similar behaviors to previously investigated ones. As a last, geographical proximity has an indirect effect on radical innovation, but its magnitude is shallow (see Figure C.1 in the appendix). Hence, local partners appear to play a role in modeling agents' actions and on their ability to innovate. So, if on one side geographical proximity can decrease communication and transaction costs, it can also undermine the exposure to what is novel, thus locking actors in crystallized routines. This has a detrimental effect on the competitive ability of actors.

Figure 3.7: Predicted innovation levels by Geographical Proximity



Notes: the scatter plot has been realized with Matplotlib predicting values for radical and incremental innovation. In so doing, coefficients of regression M 6 and M 11 have been applied to observed values of geographical proximity.

3.6 Conclusion

This research explored the relationship between proximity, business model, and the consequent effect on innovation. What emerges is a relation between actors' closeness to partners and the actions they design to create and capture value. This closeness is not merely geographical but social and cognitive. In particular, friendship and knowledge integration emerge as variables that can enhance the choice for an efficiency-oriented business model, aiming to lower transaction costs. Also, knowledge integration appears as a key driver of novelty-oriented models. Indeed, innovation is not a matter of a single node, but networks. Furthermore, the exploratory analysis shows how this translates into innovation performance. Hence, radical innovation appears to benefit from both novelty- and efficiency-oriented models, and of their ability to convey gains from closeness.

Three key managerial insights emerge from this analysis. First, the literature has shown several cases of incumbents or new entrants failing not for resource scarcity or innovation inability, but because they ignore peers. Looking at peers as sources of knowledge is not a weakness, but a huge opportunity. Indeed, the insights firms can gain from closeness to other actors translates into actions performed and profit. Second, pursuing innovation does not mean abandoning efficiency. As shown in figure 3.6, firms with a dual perspective gain the most. Third, it is not where you are based or who you know that determines your likelihood of profiting from innovation, but how you will take advantage of that resources. This is why experimenting with different business models with a focus on network resources available, it is crucial for success.

From a theoretical standpoint, this research highlighted the need to analyze the system of interdependencies raising between the product/service, the manufacturing/providing firm, and its environment. For proximity literature, the prompt is to consider how those proximities are translated into value by looking at firms' actions. For business model scholars, the suggestion is to consider social dynamics in which actors are embedded, and the consequent constraints searching for an optimal model. More generally, as suggested by a recent special issue (Nicholson, Gimmon, and Felzensztein, 2017), many theoretical and empirical gains can be obtained through the integration between management and economic geography. This chapter has chosen this direction of contribution.

3.6.1 Limiations

This exploratory study comes with several limitations that open up to further research development. First, data collection timing corresponds to one of the most challenging socio-economic crises, as shown by several recent contributions (e.g., Bonaccorsi et al., 2020; Hsiang et al., 2020), namely Covid-19. Government policies to contrast the pandemic largely affected individuals, and this translated into generalized social distress. Given the absence of any previous measure on the sampled firms, this exploratory research cannot take into account any variability due to this exogenous shock. Second, the sample of firms in this study does not guarantee any representativeness. Indeed, even though average trends emerged during the analysis do not highlight any particular issue, the considerable heterogeneity of business activities and the not-sufficient sample size may lead to biased interpretations. For this reason, the research team will continue to collect firms' answers to the questionnaire to be used in a follow up research to strengthen emerging direct and indirect relations. Third, given the static analysis provided, the results obtained cannot offer any evidence on underlying mechanisms. On the contrary, a careful mechanistic study as per Hedström and Bearman (2011), would have provided the chance to discern learning and coordination stages and the consequent effect of proximity, thus confirming or rejecting the theoretical explanation proposed in section two. Furthermore, concerning the third point, I cannot provide any causal link but only coefficients of relation, thus hoping for further research to provide such evidence with longitudinal analysis. Lastly, our model has not been tested against any competing explanation of the effects emerging, and it is also deficient of a robustness check over newly introduced digital technologies. On this last point, a great debate exists among scholars (Friedman, 2006; Morgan, 2004) that is worth to be considered in the future development of this research. In conclusion, this research opens up to neglected but potentially critical phenomena, without any claim for completeness, causality, or generalization.

3.6.2 Acknowledgments

The author is extremely grateful to professor Lorenzo Zanni who provided data for this research.

This page is intentionally left blank.

Conclusions

This section summarizes the main results, highlights some limitations, and offers future research advance considerations.

The thesis aims to deepen the relationship between *proximity and innovation*. In so doing, it leverages on novel methodological techniques and different theoretical angles to explore both *how and why* closeness may influence firms' innovation. The last debate has caught the attention of a growing audience since early 2000, not only in terms of magnitude but also of literature spanning contributions. Firstly, the astonishing diversity of approaches required the author to limit the viewpoint to a single discipline, namely, management. Then, the choice of a methodology able to sustain the author's qualitative inquiry ended up with topic modeling. After clarifying expectations and what constitutes a well-design procedure, the author applied a typical machine learning train-test design to uncover latent semantic structures. Together with qualitative means, this leads to the disentanglement of such a spanning and sometimes ambiguous literature theme, like proximity. Once theoretical claims, empirical evidence, and managerial attention are reconciled, the dissertation brings attention to a strategy topic to complement our understanding of innovation's closeness influence.

Chapter one offers three critical insights for what concerns researchers' scopes and evaluation strategies in topic modeling research. First, it disentangles four relevant units of analysis that populate management research: Classification, Qualitative Variables, Individual Topics, and Topology. Also, these categories are subdivided into two classes based on the 'substantial semantic interest' shown. This refinement of the authors' scope offered the chance to build up a simple and effective decision tree. Second, evaluation practices have been collected and grouped into six subgroups: Heuristic, Statistic, Eyeballing, Semantic, External, and Assessment of the statistical model. In so doing, an extensive collection of methods is offered to readers. Thirdly, comparing scopes and evaluation practices, the author details the emergence or lack of

connection between researchers' scopes and evaluation strategies. This is likely to offer a methodological blueprint for any scholar, and it certainly does here.

Deepening the relationship between proximity and innovation, chapters two leverages on a novel methodology and offers the following contributions. First, it clarifies to which managerial conversation proximity scholars took part. In so doing, it identifies the twelve most relevant topics learned from top tier management journals and projected into proximity and innovation debate. Second, it classifies all contributions based on the outcome variable considered, obtaining the following sub-classes: technological innovation, product & process innovation, innovation novelty, the economic value of innovation, network, and process. Thirdly, it identifies the most relevant theoretical claims for the effect of proximity on innovation, uncovering inconsistencies, or agreement. Lastly, it collects empirical evidences by outcome variable sub-class, intersecting these with managerial conversation focus. The research for clarity and relationship with management literature is meant to shed light on contributions offered and other development paths.

The last section introduces the business model in the proximity and innovation debate. This chapter offers a reflection of the role played by proximity in business model design. In so doing, it proposes learning and coordination as two phases where closeness may guide design. Both learning and coordination are recurrent topics of proximity and business model literature, which look at these from two different angles. Indeed, proximity dimensions are studied as possible ways to obtain better learning and coordination. For the business model literature, learning is the act of 'borrowing' good practices, and coordination is a component of the model design. At this intersection stands my theoretical claim. Then, the exploratory analysis results give evidence of both direct and indirect effects. The latter suggests that firms' business model design is affected by proximity and that, in turn, this influence may get to their ability to innovate. Limitations apart, this study bridges two literature streams often unrelated, showing possible gains for both.

This last chapter is still an exploratory blueprint both in terms of sample and methodology. While the former limitation has been extensively discussed, let me add on the second. Given the enormous amount of information offered by the Internet, a possible approach to measuring proximity dimension is topic modeling. For this, great examples gathered in chapter one provide several interesting future applications (e.g., Corritore, Goldberg, and Srivastava, 2020; Haans, 2019). In particular, learned topics can be used to generate measures of distinctiveness or business proximity (Shi, Lee, and Whinston, 2016) leveraging on entropy or cosine similarity. This may be especially useful to create closeness measures within and outside clusters. Other natural language processing techniques that may reveal novel insights are Name Entity Recognition

and Word Embedding. The first can detect 'real-world objects' (e.g. a person, an organization, or a country). While the second represents words as vectors of similarities to other words uncovering meanings. All these algorithms may push social sciences research boundaries far away. Unfortunately, these methodological leaps and intuitions could not be explored due to my sample of respondents. This is not only in terms of the number of collected answers but mostly of Italian SMEs attention over the Internet. Indeed, less than 50% of the firms sampled have a website, and those that do seem to not care (e.g. old or scarce content).

Overall, this thesis offers two contributions. Firstly, it shows how much cross-disciplinary dialogue among social science may benefit our understanding of social phenomena. Indeed, it tries to bring together both regional economists and management scholars to unpack how distance matters in firms' activities. Second, it highlights the need for a cross-sciences dialogue to explore novel ways of looking at phenomena. In particular, following the lead of great scholars such as DiMaggio, Nag, and Blei (2013), it demonstrates how the advancements made by computer scientists may help social disciplines to develop further. Nonetheless, the thesis bears limitations that transcend single chapters. First, the general attention over the Management literature may decrease the ability to understand the *proximity & innovation* topic. Even though this is considered a necessary condition not to lose track, many more insights would have been acquired from a 'holistic' approach. For this reason, an inter-communities effort may generate significant momentum around theoretical and empirical claims. Second, the thesis has not considered closeness dynamics within online spaces of interaction. Studies such as Lanzolla and Frankort (2016) prove how much this is worth. Hence, this dissertation cannot disentangle the two integrated dimensions — online and offline — of actors' business activities. Third, focusing on the sole 'proximity' stream to uncover the influence of cognitive and social dimensions may be reductive. An extensive analysis to reconcile different theoretical viewpoints — such as embeddedness, institutional theory, and homophily — is another great missing. Lastly, this thesis goes through a mixed approach to empirical research since topic modeling cannot be considered merely quantitative in social sciences. However, some pure qualitative analysis would provide an even greater appreciation of phenomena investigated. In so doing, research for face validity would have been helpful for both academic and practitioner audiences.

This page is intentionally left blank.

Appendix A

Topic modeling in management research

A.1 Articles Search and Retrieval

The articles search and retrieval process has been performed the 2^{nd} of January 2020. The query performed on Scopus was structured as:

TITLE-ABS-KEY ("topic modeling" OR "topic model*" OR "natural language processing" OR "nlp" OR "latent dirichlet" OR "LDA") AND (LIMIT-TO (EXACTSRCTITLE , "Academy Of Management Journal") OR LIMIT-TO (EXACTSRCTITLE , "Administrative Science Quarterly") OR LIMIT-TO (EXACTSRCTITLE , "Entrepreneurship Theory And Practice") OR LIMIT-TO (EXACTSRCTITLE , "Industrial And Corporate Change") OR LIMIT-TO (EXACTSRCTITLE , "Information Systems Research") OR LIMIT-TO (EXACTSRCTITLE , "Journal Of Business Venturing") OR LIMIT-TO (EXACTSRCTITLE , "Journal Of Management") OR LIMIT-TO (EXACTSRCTITLE , "Journal Of Management Studies") OR LIMIT-TO (EXACTSRCTITLE , "Journal Of Product Innovation Management") OR LIMIT-TO (EXACTSRCTITLE , "Leadership Quarterly") OR LIMIT-TO (EXACTSRCTITLE , "Management Science") OR LIMIT-TO (EXACTSRCTITLE , "MIS Quarterly Management Information Systems") OR LIMIT-TO (EXACTSRCTITLE , "Organization Science") OR LIMIT-TO (EXACTSRCTITLE , "Organization Studies") OR LIMIT-TO (EXACTSRCTITLE , "Research Policy") OR LIMIT-TO (EXACTSRCTITLE , "Strategic Entrepreneurship Journal") OR LIMIT-TO (EXACTSRCTITLE , "Strategic Management Journal") OR LIMIT-TO (EXACTSRCTITLE , "Strategic Organization"))

From this search, 30 articles have been retrieved. Additionally, the query on ISI - Web of Knowledge was structured as:

ALL=("Topic modeling" OR "topic model*" OR "natural language processing" OR "nlp" OR "latent dirichlet" OR "LDA") AND SO=(ACADEMY OF MANAGEMENT JOURNAL OR ADMINISTRATIVE SCIENCE QUARTERLY OR ENTREPRENEURSHIP THEORY "AND" PRACTICE OR INDUSTRIAL "AND" CORPORATE CHANGE OR INFORMATION SYSTEMS RESEARCH OR JOURNAL OF BUSINESS VENTURING OR JOURNAL OF MANAGEMENT OR JOURNAL OF MANAGEMENT STUDIES OR JOURNAL OF PRODUCT INNOVATION MANAGEMENT OR LEADERSHIP QUARTERLY OR MANAGEMENT SCIENCE OR MIS QUARTERLY OR ORGANIZATION SCIENCE OR ORGANIZATION STUDIES OR RESEARCH POLICY OR STRATEGIC ENTREPRENEURSHIP JOURNAL OR STRATEGIC MANAGEMENT JOURNAL OR STRATEGIC ORGANIZATION)

Here, 28 articles have been retrieved. Merging the results of these two searches, we obtained 33 unique articles. Of these, 24 were included.

Furthermore, authors performed a search on Google Scholar by source (the same journals of the previous queries) and within each Journal Website, with the following keywords: topic modeling, topic model, natural language processing, nlp, latent dirichlet, lda. This search provided 15 articles more. Lastly, a further check on each journal website has been performed. Tab A.1 shows count of articles included by journal.

A.2 Evaluation Metrics

A.2.1 Arun et al. 2010

Arun et al. (2010) obtained their measure leveraging on the Topic-Word matrix (M_1), and the Document-Topic matrix (M_2). In particular, employing the Symmetric Kullback-Leibler (KL) divergence measure, the optimal number of topics is obtained when the following equation reaches its minimum:

$$Arun2010(M_1, M_2) = KL(C_{M_1} \parallel C_{M_2}) + KL(C_{M_2} \parallel C_{M_1}) \tag{A.1}$$

Table A.1: Number of articles by journal

Journal	Count
Academy of Management Journal	1
Administrative Science Quarterly	3
Industrial and Corporate Change	1
Information Systems Research	8
Journal of Management	2
Journal of Product Innovation Management	2
Leadership Quarterly	2
MIS Quarterly	10
Management Science	3
Organization Science	2
Research Policy	1
Strategic Management Journal	3

Journals are listed in alphabetical order.

Where, C_{M_1} is the distribution of singular values obtained applying Singular Value Decomposition (SVD) to the matrix M_1, and C_{M_2} is the distribution obtained normailizing the vector $L \times M_2$ (L is a vector of documents lenghts)Arun, Suresh, Madhavan, and Narasimha Murty, 2010.

A.2.2 Cao et al. 2009

Cao et al. (2009) measure computes the average cosine distance among topics, to measure the topic structure stability:

$$ave\_dis(structure) = \frac{\sum_{i=0}^{K}\sum_{j=i+1}^{K} corre(T_i, T_j)}{K \times (K-1)/2} \tag{A.2}$$

Where K is the number of topics, and T_i and T_j represent two topics. The correlation is measured as:

$$corre(T_i, T_j) = \frac{\sum_{v=0}^{V} T_{iv} T_{jv}}{\sqrt{\sum_{v=0}^{V}(T_{iv})^2}\sqrt{\sum_{v=0}^{V}(T_{jv})^2}} \tag{A.3}$$

A lower distance corresponds to a better structure, therefore the optimal topic number is the one with minimum $ave_dis(structure)$Cao et al., 2009.

A.2.3 Deveaud et al. 2014

Deveaud et al. (2014) method is based on the following:

$$\hat{K} = arg\max_{K} \frac{1}{K(K-1)} \sum_{k,k' \in T_k} D(k \parallel k') \tag{A.4}$$

Where K is the number of topics given as parameters, T_K is the set of K topics modeled, and $D(k \parallel k')$ is the Jensen-Shannon divergence between pairs of topicsDeveaud, SanJuan, and Bellot, 2014:

$$D(k \parallel k') = \frac{1}{2} \sum_{w \in W_k \cap W_{k'}} P_{TM}(w|k) \log \frac{P_{TM}(w|k)}{P_{TM}(w|k')} + \frac{1}{2} \sum_{w \in W_k \cap W_{k'}} P_{TM}(w|k') \log \frac{P_{TM}(w|k')}{P_{TM}(w|k)} \tag{A.5}$$

Therefore, $\hat{K}$ is the number for which the model produces the best topics (or most scattered).

A.2.4 Dispersion of Residuals

This method considers the linkage between number of topics and model fitTaddy, 2012. In particular, since the theoretical multinomial dispersion of σ^2 should be equal to one, this method consists testing for overdispersion of the variance. Therefore, if the model σ^2 is higher than 1, the true K is larger than what estimated.

A.2.5 Document-completion Held-out Likelihood

The model predictive performance can be assessed estimating the probability of a slice of the document (words or a half) on the base of another slice of the same documentRoberts, Stewart, Tingley, et al., 2014; Wallach, Murray, Salakhutdinov, and Mimno, 2009. In particular, Wallach et al. (2009) formalized this estimation as follows:

$$P(w^{(2)} \mid w^{(1)}, \Phi, \alpha m) = \frac{P(w^{(2)}, w^{(1)} \mid \Phi, \alpha m)}{P(w^{(1)} \mid \Phi, \alpha m)} \tag{A.6}$$

Where $w^{(1)}$ is the first half and $w^{(2)}$ is the second half of the document w, $\Phi = \phi_1, ...\phi_T$ and ϕ_t is a probability vector for topic t over words, α is a concentration parameter, and m is a base measure.

A.2.6 Frequency and Exclusivity – FREX

Airoldi and Bischof (2016) built a composite measure that consider both words to topic frequency and words to topic exclusivity, trying to avoid compensation effects among the two. In particular, the following measure is an harmonic mean of bothAiroldi and Bischof, 2016:

$$FREX_{fk} = \left(\frac{\omega}{ECDF_{\phi_{.,k}}(\phi_{f,k})} + \frac{1-\omega}{ECDF_{\mu_{.,k}}(\mu_{f,k})} \right)^{-1} \tag{A.7}$$

Where ω is a weight to favour exclusivity over frequency (or vice-versa), $ECDF_{x,k}$ is the empirical comulative distribution function for x, $\phi_{f,k} = \frac{\beta_{f,k}}{\sum_{j=1}^{K} \beta_{j,v}}$ represents the exclusivity, and $\mu_{f,k} \equiv \beta_{f,k}$ represents the frequency (where $\beta_{f,k}$ is the rate of occurrence for word f in topic k).

A.2.7 Griffiths and Steyvers 2004

The Griffiths and Steyvers (2004) method consist in estimating $P(w \mid T)$ (where w are words in the corpus and T the number of topics) for different numbers of topics. In particular, authors suggest employing samples of the posterior distribution obtained throug Gibbs samplingGriffiths and Steyvers, 2004. In their example, for almost all T values, eight Markov chains were run (discarding the first 1,000 iterations) and 10 samples were taken from each chain (with a step of 100). Therefore, the best T correspond to the maximum value of $P(w \mid T)$.

A.2.8 Perplexity

Perplexity score is computed as followBlei, Ng, and Jordan, 2003:

$$perplexity(D_{test}) = \exp\left[-\frac{\sum_{d=1}^{M} \log p(w_d)}{\sum_{d=1}^{M} N_d} \right] \tag{A.8}$$

Where D_{test} is a test set composed by M documents d, w is a sequence of words for document d and N_d is the number of words in document d. Perplexity decreases monotonically.

A.2.9 Semantic Coherence

This metric has been introduced by Mimno et al. (2011) and it is maximized when words with higher probability in a topic tend to frequently co-occur togetherRoberts, Stewart, Tingley, et al., 2014. Topic

coherence is defined asMimno, Wallach, Talley, Leenders, and Mccallum, 2011:

$$C(t;V^{(t)}) = \sum_{m=2}^{M} \sum_{l=1}^{m-1} \log \frac{D(v_m^{(t)}, v_l^{(t)}) + 1}{D(v_l^{(t)})} \tag{A.9}$$

Where $D(v)$ is the number of documents with the word v, $D(v,v')$ is the number of documents containing one or more v and at least one v', and $V^{(t)}$ is a list of the M most probable words per topic tMimno, Wallach, Talley, Leenders, and Mccallum, 2011.

A.2.10 Silhoutte Coefficient Metric

This metric aims to assess the quality of clusters (topics) produced by LDA, looking at similarity and dissimilarity between themPanichella et al., 2013. The coefficient for a document d_i is:

$$s(d_i) = \frac{b(d_i) - a(d_i)}{max(a(d_i), b(d_i))} \tag{A.10}$$

Where $a(d_i)$ is the maximum distance of d_i from other documents in the same cluster, and $b(d_i)$ is the minimum distance from the centroids ($Centroid(C) = \sum_{d_i \in C} d_i / |C|$) of other clusters C. This metric ranges between -1 (bad clustering) and +1 (optimal clustering). The mean Silhouete coefficient can also be computed as:

$$s(C) = \frac{1}{n} \sum_{i=1}^{n} s(d_i) \tag{A.11}$$

A.2.11 Word and Topic Intrusion

In the word intrusion task, a human evaluator is provided with a set of high probability words (e.g. 5) for a topicChang et al., 2009. Beside these words, an intruder is randomly included from a set of words with low probability with respect to that topic. For example

Set of high probability words: dog, cat, horse, pig, cow

Inclusion of the intruder word: dog, cat, horse, apple, pig, cow

As the coherence of this illustrative topic example is high, apple is easily identified as the intruder. However,

there are cases where coherence is not so evident, e.g.: car, teacher, platypus, agile, blue, Zaire. Therefore, the model precision (MP) is evaluated as:

$$MP_k^m = \frac{\sum_s 1(i_{k,s}^m = \omega_k^m)}{S} \tag{A.12}$$

Where ω_k^m is the index of the intruder word for the k_{th} topic and model m, $i_{k,s}^m$ represent the intruder selected by individual s among the words generated for topic k_{th}, and S is the sum of individuals.
Similarly, in the topic intrusion task, a human evaluator is provided with a set of high probability topics for a document, then an intruder topic is randomly addedChang et al., 2009. The coherence evaluation task follows the structure of word intrusion one. The results of this task is then employed to generate the topic log odds (TLO):

$$TLO_d^m = \frac{\sum_s \log \hat{\theta}_{d,j_{d,\star}^m}^m - \log \hat{\theta}_{d,j_{d,s}^m}^m}{S} \tag{A.13}$$

Where j_d^m is the true intruder for document d in model m, $j_{d,s}^m$ is the intruder selected by individual s, and θ is the probability assigned.

A.2.12 Accuracy, Precision, Recall, and F-Measure

These concept are required for the operationalization of the following metrics:

True Positive: both the value predicted and the true value are 1 (true).

True Negative: both the value predicted and the true value are 0 (false).

False Positive: the value predicted is 1 (true), but the true value is 0 (false).

False Negative: the value predicted is 0 (false), but the true value is 1 (true).

The Accuracy score can be therefore calculated as:

$$Accuracy = \frac{Correct\ Predictions}{Total\ Predictions} \tag{A.14}$$

OR,

$$Accuracy = \frac{True\,Positive + True\,Negative}{True\,Positive + True\,Negative + False\,Positive + False\,Negative} \tag{A.15}$$

Precision, Recall, and F-Measure are calculated as followsHong and Davison, 2010:

$$Precision = \frac{True\,Positives}{True\,Positives + False\,Positives} \tag{A.16}$$

$$Recall = \frac{True\,Positives}{True\,Positives + False\,Negatives} \tag{A.17}$$

$$F - Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{A.18}$$

A.2.13 Area Under the ROC Curve (AUC)

The receiver operating characteristic (ROC) curve is employed in machine learning to compare classifiers on the basis of the relationship between true positive rate (tpr) and false positive rate (fpr)Powers, 2011. The best classifier is the closer to the coordinates ($fpr = 0$,$tpr = 1$), and distant from $tpr = fpr$. In order to choose the optimal model, the condition applied is to minimize the AUC. The AUC metrics varies between 0 and 1, where 0.5 indicates an uninformative classifier (almost random) and 1 a perfect classification performance.

A.3 Evaluation Practices by Unit of Analysis

Table A.2: Percentages of Evaluation Practices by Unit of Analysis.

	Classification	Qualitative Variables	Individual Topics	Topology
Heuristic		23		25
Statistical		41	84	100
Arun et al. 2010		19		25
Cao et al. 2009		14		
Devedeaud et al. 2014		14		25
Dispersion of Residuals			17	
Document-completion Held-out Likelihood			17	
Frequency and Exclusivity			17	
Griffiths and Steyvers 2004		28		75
Perplexity		10	50	
Semantic Coherence			34	
Silhoutte Coefficient			17	
Eyeballing		69	84	75
Keywords Inspection		64	84	75
Visual inspection		5	17	
Semantic		28	100	75
Word Intrusion		5	34	
Topic Intrusion				
Polysemy Inspection		5		
Topic to document inspection		19	100	25
Human coder agreement		10	50	50
External		5	17	
Assessment	100			
Accuracy				
Area Under the ROC curve	17			
Precision	50			
Recall	67			
F-measure	67			

Note: numbers refer to inter-group percentages. Aim (number of articles): Classification (6), Qualitative Variables (22), Individual Topics (6), Topology (4).

Appendix B

Proximity and Innovation

B.1 Articles sources

Table B.1 and Table B.2 shows articles retrieved by journal (in alphabetic order).

Table B.1: Number of proximity and innovation articles by journal

	Journal	Count of Studies
1	Academy of Management Journal	2
2	Administrative Science Quarterly	3
3	American Journal of Sociology	1
4	Annals of Tourism Research	2
5	British Journal of Management	1
6	Business Strategy and the Environment	1
7	Economic Geography	3
8	Entrepreneurship and Regional Development	3
9	Environment and Planning A	3
10	European Management Review	2
11	European Urban and Regional Studies	2
12	Industrial Marketing Management	3
13	International Business Review	1
14	International Journal of Industrial Organization	1

15	Journal of Business Research	6
16	Journal of Economic Geography	6
17	Journal of International Marketing	1
18	Journal of Product Innovation Management	1
19	Journal of Regional Science	2
20	Journal of Small Business Management	3
21	Journal of Urban Economics	1
22	Management Science	2
23	Organization Science	1
24	Production Planning and Control	1
25	R and D Management	1
26	Regional Studies	15
27	Research Policy	16
28	Small Business Economics	4
29	Strategic Management Journal	4
30	Technological Forecasting and Social Change	2
31	Technovation	3
32	Urban Studies	1

Table B.2: Number of training set articles by journal

Journal	Count of studies
Academy of Management Journal	1601
Academy of Management Review	865
Administrative Science Quarterly	542
Journal of Management	1974
Journal of Management Studies	1723
Organization Science	1467
Organization Studies	1844
Strategic Management Journal	2746

B.2 Topic Modeling

Topic modeling is a methodology offered to social scientists by the machine learning community (DiMaggio, Nag, and Blei, 2013). Under this umbrella term are grouped a series of algorithms that enable scientists to uncover themes diffused in a collection of documents, and to consequently characterize those documents with the discovered themes (Blei, 2012). This article leverages on a particular topic modeling algorithm, Latent Dirichlet Allocation, firstly introduced by Blei, Ng, and Jordan (2003).

B.2.1 Topic modeling analysis

The topic modeling analysis has been performed through Python programming language. In particular, we leveraged on SpaCy library for text pre-processing, Gensim and Mallet for topic modeling. The python script has been written following the example provided by Sieweke and Santoni (2020).

As a first step, the 12,762 abstracts from organization and management theory journals have been pre-processed. This step is necessary to prepare the raw text to be analyzed. The processing pipeline adopted in this study is based on a language model provided by SpaCy [1], and it consist of: *tokenization*, each abstract has been segmented in single objects, thus words, numbers, punctuation; *lemmatization*, each word has been transformed in its base form, e.g. 'goes' turns into 'go'; *token removal*, common words (stop-words such as 'the', 'or', 'and' etc.) and numbers have been removed. Authors agreed on not including any set of customized stop-words.

Once the *natural language processing* pipeline has been concluded, each abstract (document hereafter) has been further analyzed with Gensim . Firstly, n-grams (multi-words expressions, e.g. 'New York') have been detected. This step allows to generate from two (bi-grams) or three (tri-grams) distinct tokens (words) a unique one (e.g. 'absorptive_capacity' or 'transaction_cost_theory'). Hence, the 'tri-grammed' documents have been employed to generate a *Dictionary* (set of unique tokens in the database to which an id is assigned) and *Corpus* (each document is transformed in a vector composed by a word id and the count of its occurrence in that document). In so doing, a *bag-of-words* vectorized representation of documents has been obtained, which constitutes the input for the topic modeling.

As a third step, Mallet was used to estimate our LDA. A part from *Corpus* and *Dictionary*, the algorithm requires the researcher to indicate the number of topics. In order to select the optimum number, a

[1] The language model employed (namely 'en_core_web_lg') is an English multi-task CNN trained on OntoNotes5, with GloVe vectors on Common Crawl.

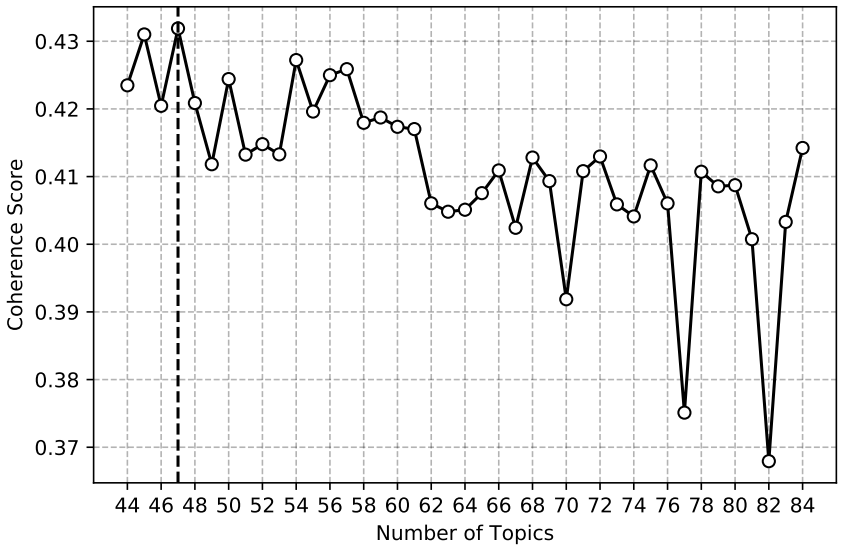
set of competing models have been estimated, in a range of ± 20 from 64. This is the number of keywords offered by the Organization and Management Division of the Academy of Management to their reviewers (Tab B.3 lists all keywords retrieved in January 2020). For each model, the *Coherence Score* metric has been estimated, thus retaining the model with the highest score (Mimno, Wallach, Talley, Leenders, and McCallum, 2011). As shown in Figure B.1, the number of topics for which the coherence score is maximized is 47.

Table B.3: Organization and management theory reviewers keywords

Collection of keywords			
Agency Theory	Careers & Mobility	Institutional Logics / Complexity	Social Responsibility & Ethics
Behavioral Theory & Decision Making	Categories & Categorization	International, Comparative & Global	Societal Impact
Complexity and Systems Theory	Change	Interorganizational Relations	Status and Reputation
Ecology (Organizational, Population, Community)	Communities	Knowledge Flows and Knowledge Management	Strategy and Strategizing
Evolutionary Theory, Path Dependence and Imprinting	Corporate Governance	Leadership	Technology
Institutional Theory	Culture	Learning, Adaptation, Routines, and Knowledge Management	Trust and Cooperation
Networks and Embeddedness	Deviance	Markets and Economies	Open Innovation
Power / Resource Dependence	Diffusion and Adoption	Materiality / Visuality	Archival or Historical
Practice Theory	Discourse, Rhetoric, Communication	Occupations, Professions and Work	Experimental (Lab or Field)
Sensemaking and Cognition	Diversity	Organizational Design, Structure and Control	Empirical, Big Data
Social Movement Theory	Economic Sociology	Organizational Identity	Empirical, Qualitative
Transaction Cost Economics	Entrepreneurship	Paradox Studies	Empirical, Quantitative
Upper Echelons Theory	Groups, Top Management Teams and Executives	Performance and Effectiveness	Empirical, Set-theoretic (including QCA)
Actor-Network Theory	Human Resource Management and Employment Relationship	Process Organization Studies	Mixed Methods
Critical Theory	Inequality/Stratification	Social Capital	Simulation
Capabilities and Competencies	Innovation and Creativity	Social Media	Theoretical/Conceptual (No Data)

As a fourth step, the LDA model with 47 topics has been trained on the sample of 12,762 abstracts from organization and management theory journals. To further assess the goodness of the topics retrieved, some eyeballing techniques have been employed. In particular, Figure B.2 shows a dynamic visualization of topics obtained through PyLDAvis (available in HTML format). The left side of the figure provides a visual representation of the relationship between topics through multidimensional scaling. On the right side, the thirty most salient terms per topic are reported. Additionally, Table B.4 provides the five most relevant

Figure B.1: Coherence score values



lemmas per each topic, thus with the highest posterior probabilities per that topic.

Lastly, the trained topic modeling was used to retrieve topics discussed (and the respective posterior probability) by each of the 98 studies. This step enables to uncover how proximity and innovation articles are nested into the broader managerial literature. Here, some topics dealing with methodological and technical issues have been excluded, such as: 11, 25, 26, 28, 29, 38. In particular, these excluded topics are not much relevant. Topic 38 shows high posterior probability for 4 articles, while topic 11, 25, 26, 28, 29 for one each.

Table B.4: Five most relevant lemmas per topic

Topic Id					
1	2	3	4	5	6
employee	team	technology	group	process	alliance
perceive	ceo	technological	member	innovation	activity
commitment	executive	production	task	dynamic	partner
job	pay	company	conflict	mechanism	benefit
perception	compensation	firm	individual	develop	increase
7	8	9	10	11	12
manager	type	corporate	relate	future	resource
decision	choice	board	study	review	capability
managerial	copyright	stakeholder	result	literature	human
cognitive	framework	governance	relationship	framework	competitive_advantage
decision_making	paper	director	associate	article	firm
13	14	15	16	17	18
work	job	structure	system	level	role
service	career	complexity	development	context	study
project	woman	formal	design	analysis	play
professional	worker	coordination	process	boundary	tension
study	gender	structural	paper	integration	highlight

19	20	21	22	23	24
relationship	literature	measure	institutional	perspective	identity
trust	provide	dimension	political	risk	form
exchange	question	study	logic	economic	process
relational	important	construct	policy	agency	identification
mechanism	search	assess	institution	view	core
25	26	27	28	29	30
collective	effect	investment	influence	analysis	study
action	relationship	ownership	factor	variable	leadership
communication	high	financial	examine	effect	leader
frame	positive	r&d	study	result	community
meaning	negative	incentive	impact	size	behaviour
31	32	33	34	35	36
approach	experience	business	time	practice	knowledge
culture	venture	opportunity	study	power	learning
paper	capital	success	unit	paper	external
concept	prior	entrepreneurial	diversity	relation	learn
cultural	event	entrepreneurship	pattern	discourse	internal
37	38	39	40	41	42
network	management	behavior	information	industry	environment
tie	article	implication	action	entry	response
social	field	individual	problem	rate	environmental
position	science	discuss	issue	growth	uncertainty
collaboration	author	propose	concern	incumbent	condition
43	44	45	46	47	
focus	social	market	acquisition	control	
attention	idea	product	country	support	
paper	study	competitive	international	test	
point	creativity	diversification	foreign	hypothesis	
argue	creative	competition	local	result	

B.2.2 Topic modeling results

To provide further supporting evidences for the LDA model trained, the following abstracts have been selected to show the ability to capture *'multivocality'*. In particular, Weterings and Koster (2007) abstract has a high posterior probability for topic 41 (*Entry, Growth & Survival*) of 0.098 and for topic 32 (*Experiences*) of 0.087:

> Previous studies showed that firms established by experienced founders have higher survival rates and employment growth, but the potential effect of pre-entry experiences on innovation remains unclear. Using an original dataset, we examine the effect of founder's experiences, the relationship with the founder's previous employer and spatial proximity to the previous workplace on the innovative performance of small software firms in the Netherlands. Apart from entrepreneurial experiences, the results suggest no effect of pre-entry experiences. Continued

> contacts with the founder's previous employer appear to limit the firm's innovative performance, but firms do benefit from being established near the previous workplace.

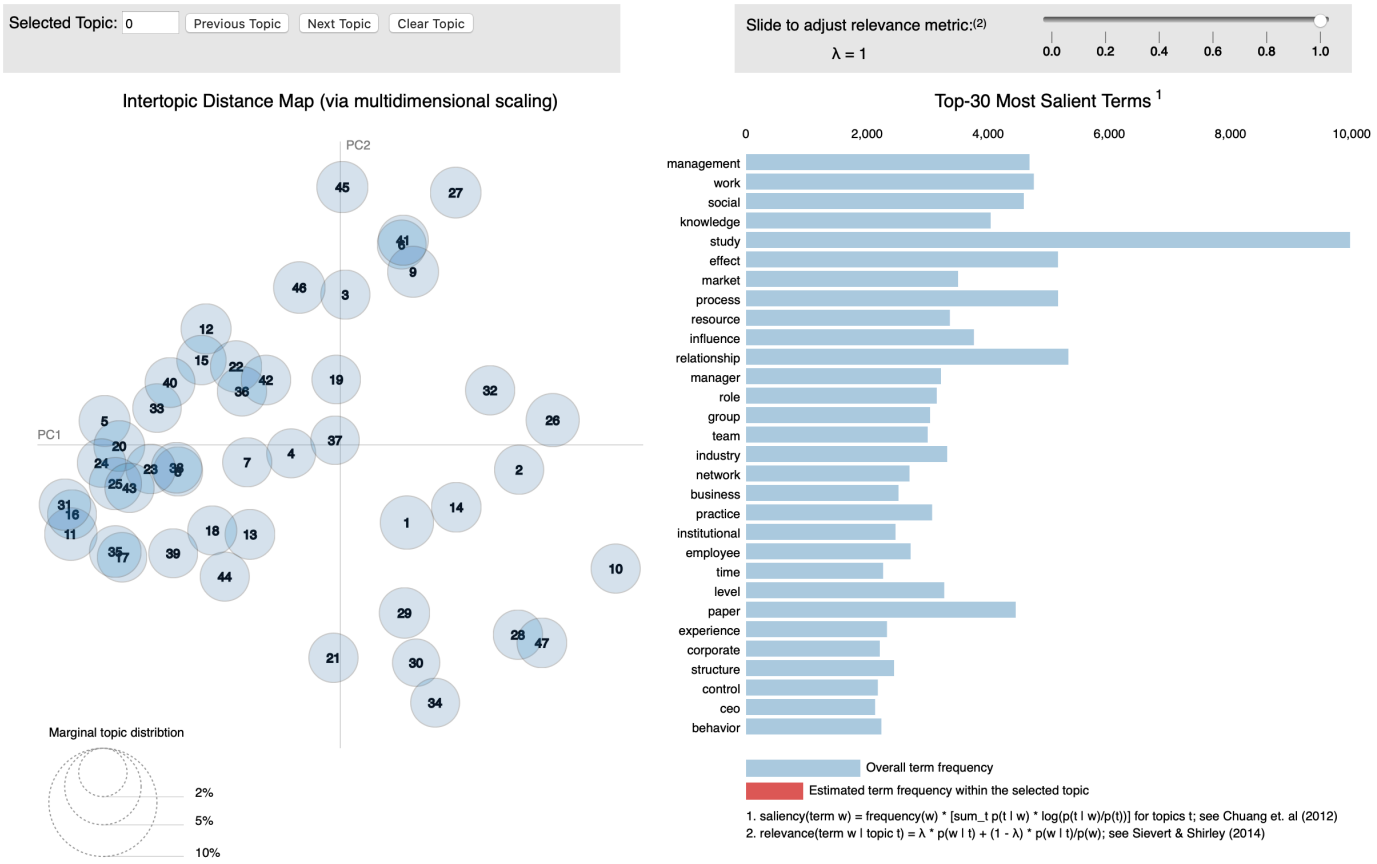
Another example is offered by Kapetaniou and Lee (2019) that shows a high posterior probability for topic 3 (*Technology & Production*) of 0.213 and for topic 46 (*Local & Distant*) of 0.105:

> Open innovation implies that geographical proximity is irrelevant. However, we posit that any potential innovation outcome depends on the spatial constraints on openness. In this paper, we add a geographical proximity dimension to open innovation by analysing how a domestic and international open innovation approach affects innovation outcomes. In particular, we hypothesise that domestic open innovation has positive effects on new-to-the-firm product innovation, due to easily accessible resources. We further posit that, through international open innovation, SMEs can access new and advanced knowledge which is not available locally, leading to more novel innovations. However, we expect that the relationship between openness, both domestic and international, and innovation is conditional on R&D activities. Our empirical analysis based on the Cyprus Community Innovation Survey supports these hypotheses. Our results underline the critical role of the spatial aspect on open innovation in SMEs, something which has remained surprisingly absent from the literature.

As last example, Li, Qiu, and Wang (2019) abstract has high posterior probability for topic 3 (*Technology & Production*) of 0.158 and for topic 6 (*Portfolio of Alliances*) of 0.101:

> We examine the organizational choice and innovative activity of technology conglomerates—firms that explore different technology fields with heated inventive activity. We develop a measure of firm-to-economy technological proximity to capture the extent of a firm's technology conglomeration. We show that technology conglomerates are more likely to form alliances and that these alliances lead to higher patent output. In terms of underlying mechanisms, we show that after alliance formation, there are significant knowledge pooling and cross-fertilization between technology conglomerates and their alliance partners. Moreover, technology conglomerates produce more patents that are novel and/or with greater impact. Our findings suggest that both synergy and tolerance for failure are important motives for technology conglomerates to use alliances to accelerate corporate innovation.

Figure B.2: Topic modeling representation with PyLDAvis



B.3 Proximity

B.3.1 Other Proximities

In our sampled articles, thirteen studies deal with proximity dimensions that are not included in the 'canonical' set. Table B.5 summarizes the investigated dimensions:

Table B.5: Other proximity dimensions

Authors	Dimensions
Brink, 2018	(i) temporal proximity: 'organisational proximity can also be framed by the visits of agents who facilitate cooperation in specific "time windows" that form "temporal proximity" for firm network innovation. ' (p. 69), (ii) virtual proximity: 'tools provided through the Internet, e.g. Skype and platforms for sharing documents facilitate "virtual proximity" of distant agents' (p. 69), (iii) vision proximity: 'in which actors possessing proximity in (. . .) in future vision' (p. 69)
Cabrer-Borras and Serrano-Domingo, 2007	commercial proximity: 'in an industrial context, the innovative contiguity between productive sectors, wij , is often set equal to 1 if the intensity in their commercial relationships is higher than the average. If we follow this idea, we can define the proximity between regions from a commercial perspective' (p. 1363)
Cantù, 2010	vision proximity: actors sharing a similar business vision
Crescenzi, Nathan, and Rodríguez-Pose, 2016	cultural–ethnic: 'whether co-patenting inventors share the same national, cultural, and/or ethnic background' (p.178)
Dolfsma and Van der Eijk, 2016	network distance: 'a focal actor may be in direct contact and can exchange knowledge directly' (p. 274)
Guan and Yan, 2016	cultural proximity: 'indicates differences between national cultures, such as social norms, ethnicities, and beliefs' (p. 1463)
Lundquist and Trippl, 2013	(i) physical proximity: 'it has less to do with pure distance measured in kilometers between different actors, than with the efforts it takes for them to interact in terms of time and costs' (p. 453), (ii) functional distance: 'refers to differences between regions in innovation performance' (p. 453), (iii) relational proximity: 'is an umbrella term for a number of non-tangible dimensions discussed in the literature, for instance cognitive, organizational, social, institutional, cultural and technological proximity' (p. 453)
Moodysson and Jonsson, 2007	(i) functional proximity: 'refers to physical distance affected by mobility. An alternative conception associated with functional proximity is therefore accessibility. It is hence not only bare Euclidean physical distance, but also includes time and cost dimensions' (p. 118), (ii) relational proximity: 'refers to a non-tangible dimension based on affinity and similarity' (p. 118)
Presutti, Boari, Majocchi, and Molina-Morales, 2019	relational proximity: 'this dimension captures both social and cognitive inter-organizational proximity ' (p. 344)
Torre, 2008	temporary geographical proximity: 'The mobility of individuals, which makes it possible to implement this mechanism, and which we shall call temporary geographical proximity, implies a strong relation to space but one that differs in nature from that described by the traditional approaches' (p.5?)
Weidenfeld, 2013	socio-cultural proximity: social and cultural similarities or dissimilarities
Zeller, 2004	(i) cultural proximity: 'is expressed by a common cultural background, which facilitates the understanding of information and the establishment of norms of behavior between innovative actors and researchers' (p. 88), (ii) internal/external proximity: 'refers to the internal relations of a firm that should enable the creation and transfer of knowledge and technologies among different units and locations of the organization' (p. 89), (iii) virtual proximity: 'can be produced by using communication and information technologies' (p. 88)

B.3.2 Proximity Measures

In the following section, a collection of measurements from our sample of article is reported. Tables are ordered on the base of the count of studies interested on that dimension, so: Geographical (Table B.6), Organizational (Table B.7), Cognitive (Table B.8), Technological (Table B.9), Social (Table B.10), and

Institutional (Table B.11).

Table B.6: Geographical proximity measurement

Authors	Research Question or Aim	Measurement
Ardito, Messeni Petruzzelli, Pascucci, and Peruffo, 2019	How do proximities moderate family firms involvement in interfirm R&D collaborations and with what consequences for innovation?	Distance in kilometers between the location sites
Baptista, 2001	Does the diffusion of new technological processes occur faster in geographical areas with higher density of knowledge sources?	'Number of previous adopters located in the firm's own region' (p.35)
Beise and Stahl, 1999	'Can the contribution of public research to industrial innovations be identified and traced back to their source by innovating companies?' (p.399)	Share of scientists from public research institutions located not more than 100 km away from the firm's country.
Bindroo, Mariadoss, and Pillai, 2012	What is the effect of customer clusters on firms' innovation?	Percentage of customers at different locations (10 choices, to which a discounting factor has been applied).
Broekel and Boschma, 2012	How do proximities influence the technical knowledge network and with what consequences for innovation performance?	'Average distance [km] between a firm and the organizations it is exchanging knowledge with' (p.424)
Broström, 2010	What is the role of geographical proximity in R&D projects?	'Link within county (=1) or outside (=0) county' (p. 1316)
Brunow, Hammer, and McCann, 2020	What is the relation of KIBS location choices and their innovation performance?	Minimum distance from cities (three types of cities distinguished on the number of inhabitants).
Capaldo and Petruzzelli, 2014	How do partner proximities influence performance in knowledge-creating alliances?	Distance (km) between partners location sites.
Chu, Tian, and Wang, 2019	How does supplier–customer proximity affect the supplier's innovation outputs?	Distance (spherical geometry: latitude and longitude) among headquarters
Crescenzi, Nathan, and Rodríguez-Pose, 2016	How do proximities influence collaborative knowledge creation at individual level?	'Inverse of linear distance between TTWA [Travel to Work Areas] centroids where each inventor is located' (p.182)
Czarnitzki and Hottenrott, 2009	How do regional characteristics influence innovative performance of firms?	location characteristics, availability of: skilled personnel, university, main suppliers, firms for networking, the infrastructure, fiscal incentives.
Davis and Greve, 1997	How do proximities influence the corporation as a governance structure?	Distance among headquarters location of prior adopters and focal firm.

Divella, 2017	To what extent heterogeneity in cooperation for innovation conduce to generation in-house, in cooperation, or adoption of innovation?	Set of dummies (local/regional [within 100 miles], national, international) and a categorical variable (0, no cooperation; 4, cooperation with partners at all geographical locations)
Dolfsma and Van der Eijk, 2016	How does distance affect opportunity of interaction and actor's innovative performance?	'Co-location of designated workspaces on the same floor in the same building' (p. 278)
Doloreux and Shearmur, 2012	Does the distance from metropolitan regions influence the innovation performance?	Distance from major or minor metropolitan areas.
Doran, Jordan, and O'Leary, 2012	Does'business innovation benefit more from increased frequency of interaction with agents located regionally, nationally and internationally?' (p.706)	Location (regional, national, international) of several actors (suppliers, customer, competitors, higher education institutes, agencies).
Dornbusch and Neuhäusler, 2015	How does academic involvement affects innovative output of research teams?	Average distance among all inventors ('crow flies' distances').
Drejer and Ostergaard, 2017	What is the role of 'employee-driven relations in firms' collaboration with specific universities on innovation'? (p.1193)	'Road travel time between firm and university postcode areas' (p.1197)
Fernandes and Ferreira, 2013	How does knowledge transfer between universities and KIBS take place?	'Importance attributed to geographic proximity to the university as a catalyst for cooperation (1 – Not at all important; 5 – Very important)' (p.465-466)
Fitjar and Rodríguez-Pose, 2011	What is 'the geographical dimension of the sources of innovation'? (p.1248)	Number of partners at three distinct locations (regional, national, international)
Freel, 2003	What are the factors influencing cooperation for innovation?	'Groupings based on the highest spatial level of innovation-related links (i.e. 'local', 'regional', 'UK' and 'overseas')' (p. 764)
Funk, 2014	'How [does] firms' innovative performance relate to the makeup of their local environments'? (p. 194)	' $FP_{it} = \Sigma_{j \neq i} \frac{x_j}{1+d_{ij}}$ where x_j is a weight, d_{ij} is the distance between firm i and firm j, t is an index for time,and j is an index for all firms other than i' (p. 203)
Gaba and Meyer, 2008	How do organizational practices spread and diffuse?	Distance of firm i at time t is $d_{ij} = \Sigma_{j=SV,128,NY} d_{ij}\theta_{jt-1}$, where d_{ij} is the distance of firm i from cluster j (Silicon Valley, New York, and Route 128) , and θ_{jt-1} is the proportion of funds for IT startups lagged by one year in that cluster.
Geerts, Leten, Belderbos, and Van Looy, 2018	How should firm organize for exploration and exploitation?	Location of inventors is used to calculate 'spatial ambidexterity (a higher degree of spatial proximity between a firm's technology exploration and exploitation activities)' (p.158)
Ghio, Guerini, and Rossi-Lamastra, 2016	How do university knowledge influence the creation of innovative start-ups?	Ratio of the academic staff of province j, 'specialized in the scientific fields that constitute the knowledge base of the industry i, and the population of the province j as in 2011' (p.297)

Giuliani, 2007	Can difference in knowledge bases influence the transfer and absorption of innovation-related knowledge?	Location in a cluster, ties with local firms for knowledge or business reasons
Guan and Yan, 2016	What are the antecedents of recombinative innovation and how can this type of innovation be measured?	Distance (spherical geometry: latitude and longitude) among country's capital.
Huggins and Johnston, 2010	What are the features of firms' networks to access knowledge and generate innovation?	Location inside/outside the region of network actors
Jespersen, Rigamonti, Jensen, and Bysted, 2018	What is the role of proximities to a partner when 'focal firm wants to initiate process innovation'? (p.880)	Spatial distance between partners
Kapetaniou and Lee, 2019	How does geographic proximity influence product innovation in open innovation?	Domestic open innovation looks at location of six different partners and generate a score that goes from 0 (no national partners) 6 (all national partners)
Knoben and Oerlemans, 2012	What is the differential impact on innovation performance of different configurations of firm's ego-network?	Number of local partners (same town/city) divided by the number of non-local partners (province, country, outside the country)
Laursen, Reichstein, and Salter, 2011	Is the propensity of university-firm collaboration influenced by distance and university research quality?	Distance ('crow flies') from the closest university and set of distance measures from first-, second-, and third-tier university
Lazzeretti and Capone, 2016	'How do the different forms of proximity influence the formation of innovation networks?' (p. 5855)	Co-location (same region/country)
Leten, Landoni, and Van Looy, 2014	What are the inter-organizational practices that enhance knowledge and innovation access in order to innovate?	Adjacent firms (location in different provinces)
Li, Xia, and Zajac, 2018	What is the role of external stakeholders in firm's innovation performance?	External innovativeness in the firm region
Li, Qiu, and Wang, 2019	How do technological conglomerates get access to new knowledge and with what consequences for innovation?	'Inventor Geographic Proximity, takes the value of one if the inventor resides less than 200 km away from their partner's headquarters, and zero otherwise' (p. 19)
Liang and Liu, 2018	How do changes in network structural and attributes proximity effects influence innovation performance?	'The proportion of direct ties between focal actor and partners in the same province to focal actor's total number of direct ties' (p. 1303)
Link and Scott, 2003	What is the influence of science parks on university missions?	Distance (miles) between an university and its associated science park
MacPherson, 1998	What is 'the role of academic linkages in the product development efforts' (p. 261) of SMEs?	'Time-distance by car' (p. 265)
Maietta, 2015	How do university–firm collaborations impact on firm product/process innovation?	Distances from the three closest faculties (km) and one distance dummy (>150 km)

Marrocu, Paci, and Usai, 2013	'What is the balance of internal and external factors in shaping regional innovative performance' (p. 1485) and with what consequences for policies?	'Distance in km between the centroids of any two regions' (p. 1487)
Mohliver, 2019	What is 'the role of of professional experts in the diffusion of innovative practices'? (p. 310)	'The proportion of backdating by proximate firms: $\frac{\Sigma_{t-2}^{t-1} Backdating_{zt}}{\Sigma_{t-2}^{t-1} Companies_{zt}}$', where z is the city where the headquarter is located, t is the year, ' Backdating denotes the firms that backdate their option grants, and Companies represents the city's population of public firms' (p. 321)
Molina-Morales and Martínez-Fernández, 2010	What is the role of social interactions, trust, shared vision, and the involvement of local institution in firms' innovation?	Affiliation to industrial district ('perceptual' identification)
Morescalchi et al., 2015	What is the influence of distance and country borders on inter-regional links?	'Distance, in kilometers, between the centroids of the NUTS3 regions' (p. 656)
Oerlemans and Meeus, 2005	What are 'the effects of proximity in innovation networks on innovative and economic performance?' (p. 90)	'Whether or not a firm's most innovative ties with buyers and suppliers were (1) intra-regional only, (2) interregional only, (3) both intra- and interregional and (4) no significant innovative ties (control group)' (p. 98)
Owen-Smith and Powell, 2004	'We contend that integrating considerations of the geographic propinquity of network structures and the institutional demography of network nodes offers new insights into the relationship between social structural position and firm-level outcome' (p. 5)	Membership in a geographically collocated network: 'Dummy variable, 1 = connected to the main network component in Boston network' (p. 15)
Parent and Riou, 2005	What is the effect of knowledge spillovers on patents' growth?	'Transportation time between the main administrative city of region i and the main administrative city of region j' (p. 759)
Phene, Fladmoe-Lindquist, and Marsh, 2006	'How [do] different sources of external knowledge influence a firm's ability to generate breakthrough innovation'? (p. 370)	National or non-national origin of a patent
Pillai and Bindroo, 2019	What is the influence of supplier cluster characteristics on firm innovation performance?	Percentage of suppliers located at different geographical distances (city, region, country etc.) discounted as distance increase.
Ponds, Oort, and Frenken, 2010	What is the effect of 'collaboration networks and geographical proximity for academic knowledge spillovers and their effect on regional innovation'? (p. 232)	Travel time between regions i and j

Presutti, Boari, Majocchi, and Molina-Morales, 2019	What is the direct and indirect (interaction with absorptive capacity) influence of proximities on innovative performance?	'We defined a dummy variable equal to 1 for customers located inside the cluster and 0 otherwise' (p. 348)
Rammer, Kinne, and Blind, 2020	How are 'different knowledge sources (...) geographically related to different type of innovation'? (p. 997)	'Distance thresholds: 50, 100, 250, 500, 1000 and 2500 m, measured as direct distance from the building in which a firm is located to the location of other firms, universities and research institutes' (p. 1006)
Schwartz, Peglow, Fritsch, and Günther, 2012	Deepen the relationship between projects characteristics and innovation output	'Mean distance between the respective ZIP-code areas of all project partners' (p. 363)
Shearmur, 2011	Is there 'any reason to suppose that local characteristics have an influence over the propensity of firms to innovate'? (p. 1226)	Euclidean distances (straight-line distance) from major and minor metropolitan areas.
Shearmur and Doloreux, 2015	Do firms engage with local KIBS more and with what consequences for innovation?	Distance between customer and supplier
Sonn and Storper, 2008	Do inventors cite locally?	Total local citations (citing and cited patents come from the same geographical unit) divided by total number of citations.
Stephan, 2014	Do research spin-offs 'have greater innovation capabilities than comparable knowledge-intensive firms'? (p. 353)	Relevance of several location factors (location in or nearby high-density region, agglomerations, university etc.)
Sternberg, 1999	What is 'the role that spatial proximity of partners plays in the establishment of innovative linkages between manufacturing SMEs'? (p. 529)	'Three spatial units where co-operation partners can be located, namely, in the same federal Land, in another federal Land or abroad' (p. 533)
Still and Strang, 2009	What is the structure of emulation (diffusion of innovation)?	Dummy variable for co-location in the same region
Taura and Radicic, 2019	'Why do some digital firms innovate more frequently than others?' (p. 351)	'DV = 1 if a firm responded "Yes" to the question "We tend to share and exchange knowledge only with recognized organizations in the local region"' (p. 359)
Ter Wal, 2014	What is the role of distance and triadic closure in the evolution of collaboration networks?	'Distance between two inventors is expressed in distance 'as the crow flies' between their places of residence, calculated on the basis of city geographical coordinates' (p.602)
Wang and Wu, 2016	What are the 'geographical knowledge spillover effects of foreign-invested firms on product innovation of local firms in a cluster'? (p. 896)	'Geographical FDI knowledge spillover as the aggregation of output value share of new products generated by foreign-invested firms in a county-level rather than provincial-level region' (p. 898)
Weterings and Boschma, 2009	Does spatial proximity facilitate face-to-face interactions and with what consequences for the innovative performance?	'Percentage of customers located within a 50-km range' (p. 750)

Weterings and Koster, 2007	What is the influence of pre-entry experience of the founders on the innovative performance of firms?	'Location near founder's previous workplace measures whether the new firm is located within a 50 km range surrounding the founder's previous workplace' (p. 327)
Whittington, Owen-Smith, and Powell, 2009	What are 'the contingent effects that network centrality and geographic propinquity exert on innovation'? (p.92)	Proximity to other firms and public research organizations was measured computing local density: $LD_{it} = \sum_j \frac{x_j}{(1+d_{ij})}$ 'where x is the weighting variable (set to one for this analysis), j indexes all firms (public research organizations) except for firm i, and d_{ij} is the distance [spherical distance in miles] between firm i and firm (public research organizations) j' (p. 103)

Table B.7: Organizational proximity measurement

Authors	Research Aim	Measurement
Broekel and Boschma, 2012	How do proximities influence the technical knowledge network and with what consequences for innovation performance?	*PUB* is a dummy with value one 'when both organizations are universities, research institutes, trade organizations, or associations'; *Private* is a similar variable for 'interactions between firms' (p. 420);
Capaldo and Petruzzelli, 2014	How do partner proximities influence performance in knowledge-creating alliances?	*SameGroup* 'taking value one if the co-assignees of the corresponding joint patent belonged to the same business group at the joint patent issue date' (p. 73)
Crescenzi, Nathan, and Rodríguez-Pose, 2016	How do proximities influence collaborative knowledge creation at individual level?	'Dummy taking the value 1 if pair belong to the same applicant, 0 if not' (p. 189)
Dolfsma and Van der Eijk, 2016	How does distance affect opportunity of interaction and actor's innovative performance?	A categorical variable for'hierarchy measure of organizational distance' (p. 278) is designed where : 0 = scientist, 1 = senior scientist, 2 = manager
Jespersen, Rigamonti, Jensen, and Bysted, 2018	What is the role of proximities to a partner when 'focal firm wants to initiate process innovation'? (p.880)	Please state if it is important to collaborate with partners that (no, partially, yes): Have similar organizational mechanisms; Have complementary managerial knowledge (p. 890)
Lazzeretti and Capone, 2016	'How do the different forms of proximity influence the formation of innovation networks?' (p. 5855)	'Prior collaboration experiences (Years of experience between ego and alter)' (p.5859)
Marrocu, Paci, and Usai, 2013	'What is the balance of internal and external factors in shaping regional innovative performance' (p. 1485) and with what consequences for policies?	'we are not considering the case in which the applicant and the inventor are the same as much as the case in which they are different but located in the same region' (p. 1488)

Oerlemans and Meeus, 2005	What are 'the effects of proximity in innovation networks on innovative and economic performance?' (p. 90)	'R&D cooperation is a count of the number of research collaborations of the innovating firm with a variety of external actors' ; External contributions to the innovation process is collected asking 'how often in the last 5 years external organizations [chamber of commerce, universities, buyers etc.] thought up ideas for, or made important contributions to, the realization of innovation' (p. 98)

Table B.8: Cognitive proximity measurement

Authors	Research Aim	Measurement
Broekel and Boschma, 2012	How do proximities influence the technical knowledge network and with what consequences for innovation performance?	TEC_{NAG} has a value of one if both organizations are active in the same technology; SIM_{NACE} is a continuous measure of technological similarity based on NACE codes and their co-occurrence at organization level.
Crescenzi, Nathan, and Rodríguez-Pose, 2016	How do proximities influence collaborative knowledge creation at individual level?	'Dummy for multiple patent inventor pairs, set as 1 if both have previously patented in the same 6-digit IPC technology field, 0 if not' (p. 189)
Dolfsma and Van der Eijk, 2016	How does distance affect opportunity of interaction and actor's innovative performance?	'How similar or different is your knowledge from your contact's knowledge? Scale 1–4; 1 = very similar, 2 = similar, 3 = different, 4 = very different' (p. 279)
Enkel and Heil, 2014	'How to build potential absorptive capacity for distant collaboration beyond established industry boundaries to gain radical rather than incremental results' (p. 242)	'Knowledge redundancy between firms based on an industry level analysis of structural equivalence' [correlation] (p.247)
Jespersen, Rigamonti, Jensen, and Bysted, 2018	What is the role of proximities to a partner when 'focal firm wants to initiate process innovation'? (p.880)	Please state if it is important to collaborate with partners that (no, partially, yes): $Cognitive-technology$ Work with similar or related technology; Are active in related technological disciplines; $Cognitive-marke$ Can support access to a new market; Operate in markets where we are not represented; Are located in new markets; Know and understand the local culture in a new market (p. 890)
Lazzeretti and Capone, 2016	'How do the different forms of proximity influence the formation of innovation networks?' (p. 5855)	'Same scientific domains' (p. 5859), actors are classified on the basis of role in the network and to nine classes

Parra-Requena, Ruiz-Ortega, García-Villaverde, and Rodrigo-Alarcon, 2015	What is the influence of external social capital on firms' innovativeness as mediated by knowledge acquisition?	Cognitive proximity (six-item Likert scale): We share the same ambition and vision as our contacts; The firm is enthusiastic about pursuing the collective goals and missions of our relationships; We share our goals and objectives with our contacts; We understand our contacts' strategy and needs; My firm's employees and my contacts' employees have positive attitudes toward a cooperative relationship; My firm and my contacts tend to agree on how to make the relationship work.
Taura and Radicic, 2019	'Why do some digital firms innovate more frequently than others?' (p. 351)	'DV = 1 if a firm responded "Yes" to the question "We search for innovations only from firms with presumably the same knowledge"' (p. 359)

Table B.9: Technological proximity measurement

Authors	Research Aim	Measurement
Ardito, Messeni Petruzzelli, Pascucci, and Peruffo, 2019	How do proximities moderate family firms involvement in interfirm R&D collaborations and with what consequences for innovation?	Measure of 'the extent to which coassignees had patented with the USPTO in the same technology classes [...] $TechRelatedness = \frac{f_i f_j'}{\sqrt{(f_i f_i')(f_j f_j')}}$ where f_i and f_j are multidimensional vectors assessing the distribution of all the patents filed by assignee i and assignee j' (pp. 189-190)
Balsmeier, Buchwald, and Stiebale, 2014	'How [do] outside directors on supervisory boards influence innovative activities of the firms they advise and monitor?' (p. 1800)	'We calculated the technological proximity of the home firm j of a particular external manager and the appointing firm i at time t, using Jaffe's measure of technological proximity (Jaffe, 1989): $p_{ij} = \frac{\Sigma_{k=1}^{K} f_{ikt} f_{jkt}}{\sqrt{\Sigma_{k=1}^{K} f_{jkt}^2 \Sigma_{k=1}^{K} f_{ikt}^2}}$ where f_{ikt} is the fraction of firm i's patents that belong to patent class k at time t' (p. 1805)
Capaldo and Petruzzelli, 2014.	How do partner proximities influence performance in knowledge-creating alliances?	'$TechProximity_{ij} = \frac{f_i f_j'}{\sqrt{(f_i f_i')(f_j f_j')}}$ where f_i and f_j are multidimensional vectors capturing the distribution of all the patents filed by the focal company (i) and by its partner (j) across the n (n = 1,..., 129) three-digit patent classes' (p. 74)
Chu, Tian, and Wang, 2019	How does supplier–customer proximity affect the supplier's innovation outputs?	'$Technology\ Proximity = \frac{(S' \Omega C)^2}{(S' S)(C' C)'}$ where [...] S is the ratio of the number of supplier's patents granted in the last three year in a patent class to the total number of supplier's patents granted in the last three years. The column vector C is similarly defined for customer's patents. Ω is a weighting matrix' (p. 17)

Guan and Yan, 2016	What are the antecedents of recombinative innovation and how can this type of innovation be measured?	'The multidimensional vector, $f_i = (f_i^1, f_i^2 \dots f_i^N)$ can be used to capture the distribution, where f_i^N indicates the ratio of patents assigned to country i in technology classification N(N=1, ...4958) [...] technological proximity between country i and j is then: $Technological\ Proximity_{ij} = \frac{f_i f_j^{'}}{\sqrt{(f_i f_i^{'})(f_j f_j^{'})}}$' (p. 1464)
Isaksson, Simeth, and Seifert, 2016	'How supplier innovation is impacted by buyer innovation' (p. 699)	'is measured by determining whether the two firms are in the same industry on different levels of SIC-code aggregation (using 2, 3, and 4 digit SIC-codes)' (p. 702)
Li, Qiu, and Wang, 2019	How do technological conglomerates get access to new knowledge and with what consequences for innovation?	Measured the 'patent output of firm i using the technology vector $S_{i,t}$ $(s_{i,1,t}, \dots, s_{i,K,t})$, and the scope of innovative activity through patent output of all other public firms in the economy using the aggregate technology vector $S_{-i,t}$ $(s_{-i,1,t}, \dots, s_{-i,K,t})$. The subscript $k \in (1,K)$ is the technology class index. [...] Firm-to-Economy Technological Proximity = $\langle \frac{S_{it}}{\|S_{it}\|}, \frac{S_{-it}}{\|S_{-it}\|} \rangle$
Liang and Liu, 2018	How do changes in network structural and attributes proximity effects influence innovation performance?	'The proportion of direct ties between focal actor and partners of the same knowledge base (PV-related inventor) to focal actor's total number of direct ties' (p. 1303)
Marrocu, Paci, and Usai, 2013	'What is the balance of internal and external factors in shaping regional innovative performance' (p. 1485) and with what consequences for policies?	'Similarity index between region i and region j, based on the distribution of patenting activity among 44 sectors, defined as: $t_{ij} = 1 - \left(\frac{1}{2}\sum_{K=1}^{K=44} |l_{ik} - l_{jk}|\right)$ where l_{ik} is the sectoral share of sector k in region i. The index tij is defined between zero (perfect dissimilarity of the sectoral distribution) and one (perfect similarity)' (p. 1487)
Phene, Fladmoe-Lindquist, and Marsh, 2006	'How [do] different sources of external knowledge influence a firm's ability to generate breakthrough innovation'? (p. 370)	'Four mutually exclusive categories: external knowledge that is (a) technologically distant and of national origin; (b) technologically proximate, of international origin; (c) technologically distant, of inter- national origin; and (d) technologically proximate, of national origin' (p. 379)
Van de Vrande, 2013	What are the effects of portfolio diversity on performance outcomes?	'The technological proximity between two firms (i and j) is computed as the uncentered correlation between their respective vectors of technological capital (measured as the number of patent applications in technology class k), P_{ik} and P_{jk}, respectively: $T_{ij} = \frac{\sum k P_{ik} P_{jk}}{\sqrt{\sum k P_{ik}^2 \sum k P_{jk}^2}}$

Table B.10: Social proximity measurement

Authors	Research Aim	Measurement
Broekel and Boschma, 2012.	How do proximities influence the technical knowledge network and with what consequences for innovation performance?	FOK is a dummy equal to one when 'former employees of Fokker B.V. are members of the top management of both firms' (p. 419)
Crescenzi, Nathan, and Rodríguez-Pose, 2016	How do proximities influence collaborative knowledge creation at individual level?	'Inverse social distance between inventors in a pair. For a given year, social distance is defined as the number of steps between pair members in the previous five years, from 0 (collaboration) to minus infinity (no connection)' (p. 189)
Dolfsma and Van der Eijk, 2016	How does distance affect opportunity of interaction and actor's innovative performance?	'How close is your working relationship with the person in question? Scale 1–5; 1 = very strong, 2 = strong, 3 = neutral, 4 = weak, 5 = very weak' (p. 279)
Jespersen, Rigamonti, Jensen, and Bysted, 2018	What is the role of proximities to a partner when 'focal firm wants to initiate process innovation'? (p.880)	Please state if it is important to collaborate with partners that (no, partially, yes): With whom we have established long- term relations. (p. 890)
Lazzeretti and Capone, 2016.	'How do the different forms of proximity influence the formation of innovation networks?' (p. 5855)	Opposite of 'number of actor pairs at distance 2' (p. 5859)
Marrocu, Paci, and Usai, 2013	'What is the balance of internal and external factors in shaping regional innovative performance' (p. 1485) and with what consequences for policies?	'Co-inventorship relations among multiple inventors of the same patent in case they are resident in different regions' (p. 1488)

Table B.11: Institutional proximity measurement

Authors	Research Aim	Measurement
Lazzeretti and Capone, 2016	'How do the different forms of proximity influence the formation of innovation networks?' (p. 5855)	'Same typology (firms, cultural organisation, research centres/universities)' (p. 5859)
Liang and Liu, 2018	How do changes in network structural and attributes proximity effects influence innovation performance?	'The proportion of direct ties between focal actor and partners of the same organizational type (e.g., university, enterprise) to focal actor's total number of direct ties' (p. 1303)
Marrocu, Paci, and Usai, 2013	'What is the balance of internal and external factors in shaping regional innovative performance' (p. 1485) and with what consequences for policies?	'a full set of country dummies' and 'a weight matrix, whose elements take value 1 if two regions belong to the same country and zero otherwise' (p. 1487)

Thomas III, 2004	'How [do] local contexts shape the capabilities of firms and the evolution of those capabilities over time'? (p. 866)	'We argue that Southern European nations are more 'proximate' to Japan than the more stringently regulated Anglo-Nordic nations. Thus, the cumulative experience by Japanese firms in launching their discoveries in Southern Europe are more likely to be complementary to the ongoing activity system than experience in the Anglo-Nordic nations' (p. 875)

B.4 Innovation

In the following section, a collection of innovation measurements from our sample of article is reported. Tables are presented with the same order of appearance in the article, so: Technological innovation (Table B.12), Product & process innovation (Table B.13), Innovation novelty (Table B.14), Economic value of innovation (Table B.15), Network (Table B.16), and Process (Table B.17).

Table B.12: Technological innovation measurement

Authors	Research Aim	Measurement
Ardito, Messeni Petruzzelli, Pascucci, and Peruffo, 2019	How do proximities moderate family firms involvement in interfirm R&D collaborations and with what consequences for innovation?	'Green innovation value (Value) was computed by counting the number of citations received by a joint patent in the 7 years after its application, hence capturing the value of the green innovation output of the collaboration' (p. 189)
Balsmeier, Buchwald, and Stiebale, 2014	'How [do] outside directors on supervisory boards influence innovative activities of the firms they advise and monitor?' (p. 1800)	'P_{it} denotes the number of patent applications of firm i in year t.' (p. 1806)
Cabrer-Borras and Serrano-Domingo, 2007	What is the role of geographical proximity in the 'dissemination of technological knowledge, both inside and between regions'? (p. 1357)	'Number of patent applications over gross added value (GAV) in 1995 constant euro for each region and year.' (p. 1364)
Capaldo and Petruzzelli, 2014	How do partner proximities influence performance in knowledge-creating alliances?	'we measured the alliance innovative performance by the number of citations each joint patent had received within five years of the issue date from subsequent patents, excluding self- citations of the coassignees.' (p. 73)
Chu, Tian, and Wang, 2019	How does supplier–customer proximity affect the supplier's innovation outputs?	'The first measure is the number of patent applications filed in a year that are eventually granted. This measure captures the quantity of innovation output. To capture the quality of innovation output, we construct a second measure by counting the total number of future citations a patent receives in subsequent years.' (p. 5)
Dolfsma and Van der Eijk, 2016	How does distance affect opportunity of interaction and actor's innovative performance?	'The number of patents per researcher was used as an admittedly less than perfect proxy for innovative output' (p. 278)

Funk, 2014	'How [does] firms' innovative performance relate to the makeup of their local environments'? (p. 194)	'I measured impact as the citation-weighted sum of nanotechnology patents applied for by firm i at times $t+1$ and $t+2$ [...] The second dependent variable measures new combinations as the sum of nanotechnology patents applied for by firm i at $t+1$ and $t+2$ that bridge previously uncombined technological domains.' (p. 202)
Geerts, Leten, Belderbos, and Van Looy, 2018	How should firm organize for exploration and exploitation?	'(Technological performance), measured by the number of patent applications weighted by their forward citations' (p. 156)
Isaksson, Simeth, and Seifert, 2016	'How supplier innovation is impacted by buyer innovation' (p. 699)	'Log Supp Pat Prod (number of patents, scaled by R&D expenditure to account for differences in input), as our core measure of supplier innovation' (p. 702)
Leten, Landoni, and Van Looy, 2014	What are the inter-organizational practices that enhance knowledge and innovation access in order to innovate?	'The dependent variable in our study is the number of firm patents in an industry and province, weighted by the number of forward patent citations received over a fixed five-year time window' (p. 1402)
Li, Qiu, and Wang, 2019	How do technological conglomerates get access to new knowledge and with what consequences for innovation?	'Log(1 + No. of an inventor's $patents_{t+1,t+3}$); Log(1 + No. of an inventor's patents in overlapping technology $classes_{t+1,t+3}$)' (p. 19)
Liang and Liu, 2018	How do changes in network structural and attributes proximity effects influence innovation performance	'the number of patent applications by the actors in a year.' (p. 1301)
Link and Scott, 2003	What is the influence of science parks on university missions?	'overall research output, measured in terms of publications, by faculty has increased; overall research output, measured in terms of patents, by faculty has increased ' (p. 1341)
Marrocu, Paci, and Usai, 2013	'What is the balance of internal and external factors in shaping regional innovative performance' (p. 1485) and with what consequences for policies?	'the innovation output inn is proxied by the yearly average of patents per-capita in 2005–2007' (p. 1489)
Owen-Smith and Powell, 2004	'We contend that integrating considerations of the geographic propinquity of network structures and the institutional demography of network nodes offers new insights into the relationship between social structural position and firm-level outcome' (p. 5)	'Yearly count of successful patent applications' (p. 15)
Parent and Riou, 2005	What is the effect of knowledge spillovers on patents' growth?	'Because the data on patents correspond to granted patents whose application date is between 1989 and 1999, the variable Y is the stock of knowledge i generated in region i at the initial period 1989, and $\dot{Y}_i$ is its time derivative over the period 1989–1999.' (p. 752)

Ponds, Oort, and Frenken, 2010	What is the effect of 'collaboration networks and geographical proximity for academic knowledge spillovers and their effect on regional innovation'? (p. 232)	'$P_{i,k,t}$ stands for economically valuable knowledge as measured by patent applications of firms in region i in technology k in the period of 1999–2001' (p. 241)
Schwartz, Peglow, Fritsch, and Günther, 2012	Deepen the relationship between projects characteristics and innovation output	'The innovation output of subsidized R&D cooperation projects is measured as the number of patent applications and as the number of publications that directly emerged from an R&D project.' (p.) 362
Van de Vrande, 2013	What are the effects of portfolio diversity on performance outcomes?	'Weighted patent counts (WPC) is a count variable, where each patent i is weighed according to the subsequent citations C_i it receives, assuming that more important patents receive more citations and vice versa.' (p .614)
Whittington, Owen-Smith, and Powell, 2009	What are 'the contingent effects that network centrality and geographic propinquity exert on innovation'? (p.92)	'The outcome of interest is a yearly count of patents assigned to the DBFs, categorized by application date rather than issue date' (p. 101)

Table B.13: Product & process innovation measurement

Authors	Research Aim	Measurement
Beise and Stahl, 1999	'Can the contribution of public research to industrial innovations be identified and traced back to their source by innovating companies?' (p.399)	'The number of innovations IPUBi introduced by firm i in a 3-year period derived with findings of public research' (p. 412)
Brunow, Hammer, and McCann, 2020	What is the relation of KIBS location choices and their innovation performance?	'Innovation data are recorded in the IAB EP as a binary variable and surveys whether the establishment has undertaken each kind of innovation within the last year [...] While technological forms of innovation are represented by product improvement and product introduction as well as process innovations, non-technological innovation is indicated by organizational innovation, which is a summary variable encompassing various organizational changes.' (p. 4)

Doran, Jordan, and O'Leary, 2012	Does'business innovation benefit more from increased frequency of interaction with agents located regionally, nationally and internationally?' (p.706)	'Product innovation is defined as the introduction of new or improved goods/services, which may be either new to the market or to the business during the reference period, which is 2004-2006. Process innovation includes any of the other four types of innovation identified by Schumpeter (1934): (i) the introduction of a new method of production, (ii) the opening of a new market, (iii) the acquisition of a new source of supply or (iv) the re-organization of management or distribution channels.' (p. 712)
Fernandes and Ferreira, 2013	How does knowledge transfer between universities and KIBS take place?	'We applied the number of product/service innovations, the number of process innovations and the total number of innovations (products/services, processes, organisational, introduction of already existing products into new markets, patents, registering brands and new designs for products and processes) as the dependent variables' (p. 465)
Fitjar and Rodríguez-Pose, 2011	What is 'the geographical dimension of the sources of innovation'? (p.1248)	'managers were asked if their business had introduced any new or significantly improved products ('product innovation') and/or methods or processes for production or delivery of products ('process innovation') during the last three years' (p. 1253)
Huggins and Johnston, 2010	What are the features of firms' networks to access knowledge and generate innovation?	'The innovation measure is based on how many new products or services or adaptations to existing products or services firms had introduced during the previous 3 years' (p. 467)
Jespersen, Rigamonti, Jensen, and Bysted, 2018	What is the role of proximities to a partner when 'focal firm wants to initiate process innovation'? (p.880)	'Please state if the following organizational changes are important goals for your company for the next 3 years: Implementation of modern models to report costs; Withdrawal from unprofitable fields of activity; Moving some functions to other firms; Internationalization of the company's activities Implementation of logistics systems; Building strategic alliances; Implementation of quality systems (ISO,TQM, HACCP, etc.); Implementation of resource planning system (ERP); Implementation of systems for recording the company's activities and resources; Implementation of IT systems; Enhancement of the company's market position; Focus the company's strategy on innovation' (p. 896)
Maietta, 2015	How do university–firm collaborations impact on firm product/process innovation?	'firms are asked whether process, product and/or other innovations were introduced during the previous three years. ' (p.1346)

Molina-Morales and Martínez-Fernández, 2010	What is the role of social interactions, trust, shared vision, and the involvement of local institution in firms' innovation?	'We asked for the number of product and process innovations that had been produced in their firm.' (p. 269)
Oerlemans and Meeus, 2005	What are 'the effects of proximity in innovation networks on innovative and economic performance?' (p. 90)	'Firms were asked to indicate which percentage of the processes and products was new to the firm in a 5-year period' (p. 96)
Presutti, Boari, Majocchi, and Molina-Morales, 2019	What is the direct and indirect (interaction with absorptive capacity) influence of proximities on innovative performance?	'We measured innovation activity by asking the firms how many new products or services they had developed during the previous three years as a result of the relationship with their key customers. ' (p. 349)
Rammer, Kinne, and Blind, 2020		'we consider five types k of innovative firms: 1) innovator (product and/or process); (2) product innovator (goods and/or services); (3) new-to-market innovator (i.e. novel product innovation); (4) process innovator; (5) firms with continuous in-house R&D activity' (p. 1001-1002)
Shearmur and Doloreux, 2015	Do firms engage with local KIBS more and with what consequences for innovation?	'Whether or not the establishment has, over the three years prior to the survey, introduced a new or improved product, process, internal management or marketing/client contact method' (p. 1659)
Wang and Wu, 2016	What are the 'geographical knowledge spillover effects of foreign-invested firms on product innovation of local firms in a cluster'? (p. 896)	'Product innovation of domestic firms, the dependent variable in this study, is measured as the share of turnover generated by new products' (p. 898)
Weterings and Boschma, 2009	Does spatial proximity facilitate face-to-face interactions and with what consequences for the innovative performance?	'A dummy 0–1 coded variable that indicates whether a firm has brought new products or services to the market between 2000 and 2003 or not' (p. 750)

Table B.14: Innovation novelty measurement

Authors	Research Aim	Measurement
Bindroo, Mariadoss, and Pillai, 2012	What is the effect of customer clusters on firms' innovation?	'We captured radical innovation using one item that measured novelty of the innovation, where "3" represented the category indicating new-to-world innovations, "2" represented the category indicating new-to-the-sector/industry innovations, and "1" represented the category indicating new-to-the-firm innovations' (pp. 23-24)

Doloreux and Shearmur, 2012	Does the distance from metropolitan regions influence the innovation performance?	‘ For each of these innovation types, firms that have only introduced innovations new to the firm are analysed separately from those that have introduced an innovation new amongst their competitors: so, within each innovation category, incremental innovation and major innovation are independent of each other.’ (p. 88)
Fitjar and Rodríguez-Pose, 2011	What is ‘the geographical dimension of the sources of innovation’? (p.1248)	‘In order to analyse whether different forms of collaboration lead to different forms of innovation, the successful innovators were then asked whether the products were new to the market (‘radical product innovation’) or, in the case of process innovation, whether the processes were new to the industry (‘radical process innovation’).’ (p. 1253)
Guan and Yan, 2016	What are the antecedents of recombinative innovation and how can this type of innovation be measured?	‘Two countries are involved in technical and recombinant innovation if they create a new portfolio of technological combinations. That is to say, two countries’ co-patents involve a pair of classifications that had not been used by any patent during 1976–2007.’ (p. 1464)
Kapetaniou and Lee, 2019	How does geographic proximity influence product innovation in open innovation?	‘Fraction of the firm’s turnover relating to products new-to-the-firm; Fraction of the firm’s turnover relating to products new-to-the-market’ (p. 268)
Knoben and Oerlemans, 2012	What is the differential impact on innovation performance of different configurations of firm’s ego-network?	‘The novelty of the innovations was determined by differentiating between three types of innovation sales, that is by turnover generated by products or services that were improved versions of existing ones, new for the firm or new to the market’ (pp. 1010-1011)
Phene, Fladmoe-Lindquist, and Marsh, 2006	‘How [do] different sources of external knowledge influence a firm’s ability to generate breakthrough innovation’? (p. 370)	‘An analysis of citations indicated that the top 1 percent received an average of 59 cites, the top 2 percent received an average of 50 cites, the top 5 percent received an average of 36 cites, and the top 10 percent received an average of 27 cites. Since we found evidence of a significant drop-off at 2 percent, we define breakthrough innovations as the top 2 percent of our sample.’ (p. 379)
Pillai and Bindroo, 2019	What is the influence of supplier cluster characteristics on firm innovation performance?	‘A single item measuring novelty of the innovation captured radical innovation, with 1 representing new-to-the-firm innovations, 2 representing new-to-the-sector/new-to-the-industry innovations, and 3 representing new-to-the-world innovations.’ (p. 4)
Rammer, Kinne, and Blind, 2020	How are ‘different knowledge sources (...) geographically related to different type of innovation’? (p. 997)	‘New-to-market innovator (i.e. novel product innovation)’ (p. 1001)

Shearmur, 2011	Is there 'any reason to suppose that local characteristics have an influence over the propensity of firms to innovate'? (p. 1226)	' Two types of innovation are studied: product innovation and process innovation. Each is subdivided into innovations new to the firm, and major innovations new to the firm's market (for products) or new to Quebec (for process' (p. 1231)

Table B.15: Economic value of innovation measurement

Authors	Research Aim	Measurement
Broekel and Boschma, 2012	How do proximities influence the technical knowledge network and with what consequences for innovation performance?	'We approximate the innovative performance of firms by means of the share of significantly improved products/process on a firm's turnover' (p. 423)
Czarnitzki and Hottenrott, 2009	How do regional characteristics influence innovative performance of firms?	'The measure of innovation performance, is the share of total sales due to new products (NEWSALES)' (p. 90)
Knoben and Oerlemans, 2012	What is the differential impact on innovation performance of different configurations of firm's ego-network?	'Innovativeness was determined by asking what percentage of the firm's turnover in 2000 was generated by these innovative products and services' (p. 1010)
Li, Xia, and Zajac, 2018	What is the role of external stakeholders in firm's innovation performance?	'a firm's new product output in a specific year is the amount that a firm would receive at the market price for the new products it produces in that year if such sales were realized. New product output may not coincide with new product sales' (p. 202)
MacPherson, 1998	What is 'the role of academic linkages in the product development efforts' (p. 261) of SMEs?	'The term 'innovation performance' refers to the proportion of a firm's 1994 output (total sales) represented by products that had been introduced over the last 5 years (giving a crude measure of technological creativity)' (p. 265)
Weterings and Boschma, 2009	Does spatial proximity facilitate face-to-face interactions and with what consequences for the innovative performance?	'The percentage of turnover due to the sales of those new products or services (innovation output)' (p. 750)
Weterings and Koster, 2007	What is the influence of pre-entry experience of the founders on the innovative performance of firms?	'What percentage of the total turnover is due to the sales of the new products or services developed since 2000 (including services required for integration and implementation)?' (p. 326)

Table B.16: Network-related measurement

Authors	Research Aim	Measurement
Crescenzi, Nathan, and Rodríguez-Pose, 2016	How do proximities influence collaborative knowledge creation at individual level?	'Y is either a collaboration dummy (taking the value 1, if ij is an actual pair), or a continuous variable giving the count of collaborations for ij' (p. 182)
Divella, 2017	To what extent heterogeneity in cooperation for innovation conduce to generation in-house, in cooperation, or adoption of innovation?	'New products and/or processes are developed 'mainly by your enterprise or enterprise group' (Generation in-house) or 'mainly by your enterprise together with other enterprises or institutions', which means at least partially within the firm (Generation in cooperation)' (p. 1498)
Drejer and Ostergaard, 2017	What is the role of 'employee-driven relations in firms' collaboration with specific universities on innovation'? (p.1193)	'Dummy variable indicates whether firms had collaborated on innovation with a specific university between 2011 and 2013' (p. 1196)
Fernandes and Ferreira, 2013	How does knowledge transfer between universities and KIBS take place?	'we deployed the binary cooperation variable of cooperation with higher education institutions (0 – No; 1 – Yes), technology transfer cooperation with higher education institutions (0 – No; 1 – Yes) and R&D cooperation with higher education institutions (0 – No; 1 – Yes)' (p. 465)
Freel, 2003	What are the factors influencing cooperation for innovation?	"Did your firm co-operate with other firms or organisations for innovation related activity (including marketing, training, etc.) and/or technology transfer during the last 3 years?' (p. 755)
Giuliani, 2007	Can difference in knowledge bases influence the transfer and absorption of innovation-related knowledge?	'If you are in a critical situation and need technical advice, to which of the local firms mentioned in the roster do you turn?' (p. 149)
Laursen, Reichstein, and Salter, 2011	Is the propensity of university-firm collaboration influenced by distance and university research quality?	'whether a firm collaborates with a local university ('local' being defined as within a radius of approximately 100 miles from the firm), a non-local university (more than 100 miles from the firm), or does not collaborate with a university [...] The survey also asked firms whether they collaborated with a range of external partners, and to indicate the location of each of these partners.' (p. 512)
Lazzeretti and Capone, 2016	'How do the different forms of proximity influence the formation of innovation networks?' (p. 5855)	'data was gathered from 42 networks, related to projects funded through regional, national and international donors, covering a time-span of over 15 years (1995–2012)' (p. 5858)
Morescalchi et al., 2015	What is the influence of distance and country borders on inter-regional links?	'number of links $(y_i \equiv y(m,n))$ between NUTS3 regions (m and n) and we model its probability distribution with a count density' (p. 655)
Sternberg, 1999	What is 'the role that spatial proximity of partners plays in the establishment of innovative linkages between manufacturing SMEs'? (p. 529)	presence or absence of linkages of innovative SMEs
Ter Wal, 2014	What is the role of distance and triadic closure in the evolution of collaboration networks?	'Whether or not a tie exists between any pair of inventors at time t, given the set of inventors that is part of the co-invention network in German biotechnology at time $t-5$' (p. 601)

Table B.17: Process-related measurement

Authors	Research Aim	Measurement
Baptista, 2001	Does the diffusion of new technological processes occur faster in geographical areas with higher density of knowledge sources?	'Year of adoption (1969–80)' (p. 39)
Broström, 2010	What is the role of geographical proximity in R&D projects?	Link focusing on learning ('the link is assessed to have contributed more to impulses for innovation than to implementation of existing R&D projects with a long-term perspective (=1)', p. 1316) or on short-term ('the link is assessed to have contributed more to the execution of short-term R&D projects than to R&D projects with a long-term perspective (=1)', p. 1316)
Davis and Greve, 1997	How do proximities influence the corporation as a governance structure?	'Dates that U.S. firms initially adopted a "poison pills" or a "golden parachute"' (p. 16)
Divella, 2017	To what extent heterogeneity in cooperation for innovation conduce to generation in-house, in cooperation, or adoption of innovation?	'innovation Adoption is done by firms whose products and/or processes are developed 'mainly by other enterprises or institutions" (p. 1498)
Gaba and Meyer, 2008	How do organizational practices spread and diffuse?	'probability that a firm will adopt a CVC program at time t' (p. 986)
Mohliver, 2019	What is 'the role of of professional experts in the diffusion of innovative practices'? (p. 310)	Adoption of backdated stock-option grants
Still and Strang, 2009	What is the structure of emulation (diffusion of innovation)?	'we defined a risk set of organizations that could have become the bank's benchmarking partners and compared the characteristics of those that were and were not visited' (p. 67)

Appendix C

Business Model, Proximity, and Innovation

C.1 Additional Material

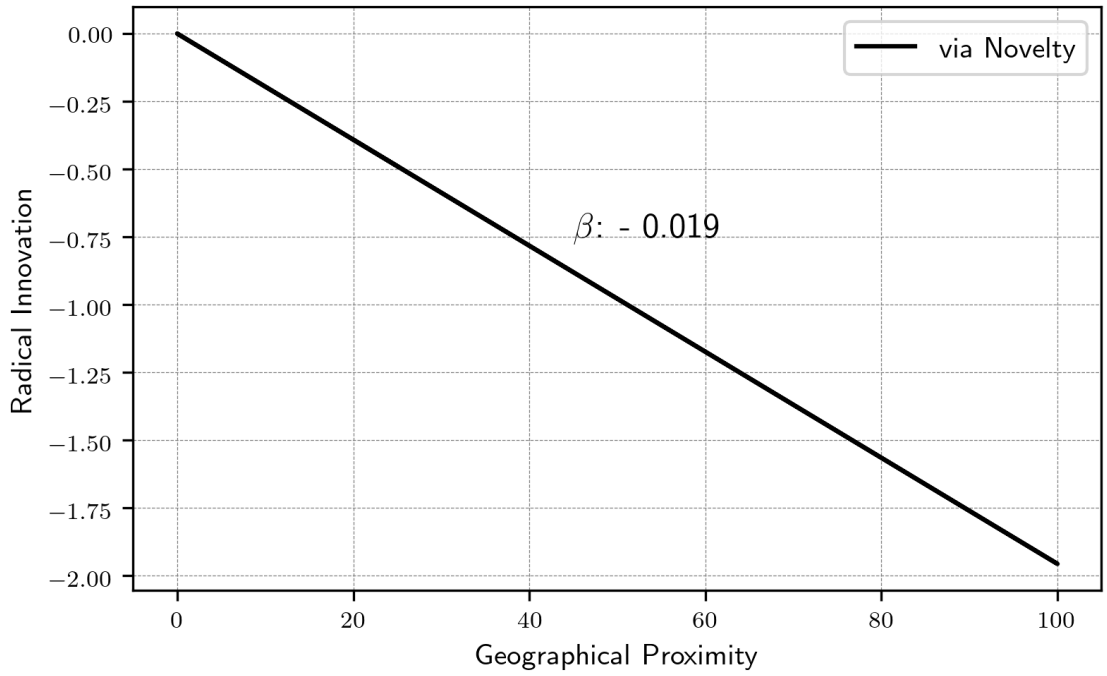
Table C.1: Factors and alpha

Variable	Item	Factor1	alpha
Efficiency BM			
	ebm_0	0.6808	0.8034
	ebm_1	0.6764	0.8030
	ebm_2	0.5585	0.8196
	ebm_3	0.5507	0.8236
	ebm_4	0.7597	0.7882
	ebm_5	0.6185	0.8096
	ebm_6	0.7686	0.7885
	ebm_7	0.7925	0.7845
	Test Scale:		**0.8231**
Novelty BM			
	ibm_0	0.8396	0.8289
	ibm_1	0.6943	0.8732
	ibm_2	0.828	0.8301
	ibm_3	0.8478	0.8278
	ibm_4	0.8371	0.8305
	Test Scale:		**0.8664**

Table C.2: Descriptive statistics

	mean	sd	min	max
Radical Inn.	24	24.12594	0	100
Incr. Inn.	34.69192	26.16447	0	95
Geo. Px.	52.30303	33.19775	0	100
Cog. Px. 1	3.409091	.9555173	1	5
Cog. Px. 2	3.722222	1.041578	1	5
Soc. Px. 1	3.656566	1.279765	1	5
Soc. Px. 2	3.681818	.9845366	1	5
Soc. Px. 3	2.924242	1.126207	1	5
Efficiency BM	3.948232	.6165058	1.75	5
Novelty BM	3.375758	.9196215	1	5
Observations	198			

Figure C.1: Indirect effect of Geographical Proximity on Radical Innovation



Notes: the plot has been realized with Matplotlib predicting values of radical innovation by geographical proximity. In so doing, boostrapped coefficient of Table 3.6 has been applied to observed values.

Table C.3: Correlation Analysis

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) Radical Inn.	1.000												
(2) Incr. Inn.	0.305*	1.000											
(3) Age(ln)	-0.116	0.015	1.000										
(4) Size(ln)	-0.073	-0.002	0.223*	1.000									
(5) Graduated(%)	0.191*	0.089	-0.131	-0.002	1.000								
(6) Geo. Px.	-0.106	0.053	-0.112	-0.032	-0.085	1.000							
(7) Cog. Px. 1	-0.023	0.072	-0.070	0.023	0.046	0.026	1.000						
(8) Cog. Px. 2	0.172*	0.088	-0.040	-0.066	0.066	-0.117	0.314*	1.000					
(9) Soc. Px. 1	0.031	0.035	0.204*	0.057	-0.203*	-0.088	0.053	0.050	1.000				
(10) Soc. Px. 2	0.035	0.055	0.007	-0.034	0.101	0.000	0.198*	0.082	-0.015	1.000			
(11) Soc. Px. 3	0.125	0.140*	-0.113	-0.208*	0.181*	0.134	0.128	0.172*	-0.261*	0.129	1.000		
(12) Efficiency BM	0.254*	0.059	0.042	0.064	0.026	-0.114	0.157*	0.343*	0.087	0.151*	0.179*	1.000	
(13) Novelty BM	0.299*	0.180*	-0.021	0.191*	0.229*	-0.177*	0.140*	0.331*	-0.027	0.105	0.129	0.394*	1.000

*** *p<0.01*, ** *p<0.05*, * *p<0.1*

Table C.4: VIF and Tolerance

	M 0		*M 1*		*M 2*		*M 3*	
Variable	VIF	Tolerance	VIF	Tolerance	VIF	Tolerance	VIF	Tolerance
Radical Inn.					1.21	0.8284		
Incremental Inn.							1.07	0.9367
Efficiency BM	1.22	0.8225			1.36	0.7361	1.33	0.7518
Novelty BM			1.29	0.7730	1.47	0.6820	1.45	0.6906
Age(ln)	1.13	0.8865	1.13	0.8847	1.14	0.8737	1.13	0.8824
Size(ln)	1.12	0.8930	1.18	0.8484	1.19	0.8397	1.18	0.8466
Graduated(%)	1.10	0.9067	1.14	0.8783	1.17	0.8567	1.15	0.8706
Geo.	1.08	0.9239	1.10	0.9113	1.10	0.9066	1.11	0.9028
Cog. 1	1.17	0.8525	1.17	0.8527	1.19	0.8420	1.17	0.8520
Cog. 2	1.27	0.7862	1.28	0.7823	1.34	0.7483	1.33	0.7496
Soc. 1	1.17	0.8545	1.16	0.8629	1.18	0.8468	1.18	0.8484
Soc. 2	1.08	0.9291	1.07	0.9379	1.08	0.9280	1.08	0.9280
Soc. 3	1.26	0.7924	1.24	0.8086	1.27	0.7880	1.28	0.7806

This page is intentionally left blank.

Bibliography

Abbasi, A., Y. Zhou, S. Deng, and P. Zhang. "Text analytics to support sense-making in social media: A language-action perspective". In: *MIS Quarterly: Management Information Systems* 42.2 (June 2018), pp. 427–464. ISSN: 21629730. DOI: 10.25300/MISQ/2018/13239.

Airoldi, E. M. and J. M. Bischof. "Improving and Evaluating Topic Models and Other Models of Text". In: *Journal of the American Statistical Association* 111.516 (Oct. 2016), pp. 1381–1403. ISSN: 1537274X. DOI: 10.1080/01621459.2015.1051182.

Amit, R. and C. Zott. "Crafting business architecture: The antecedents of business model design". In: *Strategic Entrepreneurship Journal* 9.4 (2015), pp. 331–350.

— "Value creation in e-business". In: *Strategic management journal* 22.6-7 (2001), pp. 493–520.

Antons, D., A. M. Joshi, and T. O. Salge. "Content, Contribution, and Knowledge Consumption: Uncovering Hidden Topic Structure and Rhetorical Signals in Scientific Texts". In: *Journal of Management* 45.7 (June 2019), pp. 3035–3076. ISSN: 15571211. DOI: 10.1177/0149206318774619.

Antons, D., R. Kleer, and T. O. Salge. "Mapping the Topic Landscape of JPIM, 1984–2013: In Search of Hidden Structures and Development Trajectories". In: *Journal of Product Innovation Management* 33.6 (Nov. 2016), pp. 726–749. ISSN: 15405885. DOI: 10.1111/jpim.12300.

Ardito, L., A. Messeni Petruzzelli, F. Pascucci, and E. Peruffo. "Inter-firm R&D collaborations and green innovation value: The role of family firms' involvement and the moderating effects of proximity dimensions". In: *Business Strategy and the Environment* 28.1 (2019), pp. 185–197.

Arun, R., V. Suresh, C. E. V. Madhavan, and M. Narasimha Murty. "On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations". In: *Pacific-Asia conference on knowledge discovery and data mining*. 2010, pp. 391–402.

Atkins, D. C., T. N. Rubin, M. Steyvers, M. A. Doeden, B. R. Baucom, and A. Christensen. "Topic models: A novel method for modeling couple and family text data." In: *Journal of family psychology* 26.5 (2012), pp. 816–827.

Aversa, P., S. Furnari, and S. Haefliger. "Business model configurations and performance: A qualitative comparative analysis in Formula One racing, 2005–2013". In: *Industrial and Corporate Change* 24.3 (2015), pp. 655–676.

Baden-Fuller, C. and S. Haefliger. "Business models and technological innovation". In: *Long range planning* 46.6 (2013), pp. 419–426.

Baden-Fuller, C. and V. Mangematin. "Business models: A challenging agenda". In: *Strategic Organization* 11.4 (2013), pp. 418–427.

Balboni, B., G. Bortoluzzi, R. Pugliese, and A. Tracogna. "Business model evolution, contextual ambidexterity and the growth performance of high-tech start-ups". In: *Journal of Business Research* 99 (2019), pp. 115–124.

Balland, P.-A., R. Boschma, and K. Frenken. "Proximity and innovation: From statics to dynamics". In: *Regional Studies* 49.6 (2015), pp. 907–920.

Balsmeier, B., A. Buchwald, and J. Stiebale. "Outside directors on the board and innovative firm performance". In: *Research Policy* 43.10 (2014), pp. 1800–1815.

Banks, G. C., H. M. Woznyj, R. S. Wesslen, K. A. Frear, G. Berka, E. D. Heggestad, and H. L. Gordon. "Strategic Recruitment Across Borders: An Investigation of Multinational Enterprises". In: *Journal of Management* 45.2 (Feb. 2019), pp. 476–509. ISSN: 15571211. DOI: 10.1177/0149206318764295.

Bao, Y. and A. Datta. "Simultaneously discovering and quantifying risk types from textual risk disclosures". In: *Management Science* 60.6 (2014), pp. 1351–1616. ISSN: 15265501. DOI: 10.1287/mnsc.2014.1930.

Bapna, S., M. J. Benner, and L. Qiu. "Nurturing Online Communities: An Empirical Investigation." In: *MIS Quarterly* 43.2 (2019).

Baptista, R. "Geographical clusters and innovation diffusion". In: *Technological Forecasting and Social Change* 66.1 (2001), pp. 31–46.

Barney, J. "Firm resources and sustained competitive advantage". In: *Journal of management* 17.1 (1991), pp. 99–120.

Baron, R. M. and D. A. Kenny. "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." In: *Journal of personality and social psychology* 51.6 (1986), pp. 1173–1182.

Becattini, G. “The Marshallian industrial district as a socio-economic notion”. In: *Industrial Districts and Inter-firm Co-operation in Italy*. 1990.

Beise, M. and H. Stahl. “Public research and industrial innovations in Germany”. In: *Research policy* 28.4 (1999), pp. 397–422.

Bindroo, V., B. J. Mariadoss, and R. G. Pillai. “Customer clusters as sources of innovation-based competitive advantage”. In: *Journal of International Marketing* 20.3 (2012), pp. 17–33.

Blei, D. “Probabilistic topic models”. In: *Communications of the ACM* 55.4 (2012), pp. 77–84.

Blei, D. and J. D. Lafferty. “A correlated topic model of Science”. In: *The Annals of Applied Statistics* 1.1 (June 2007), pp. 17–35. ISSN: 1932-6157. DOI: 10.1214/07-aoas114.

Blei, D., A. Y. Ng, and M. I. Jordan. “Latent dirichlet allocation”. In: *Journal of machine Learning research* 3.1 (2003), pp. 993–1022.

Bonaccorsi, G., F. Pierri, M. Cinelli, A. Flori, A. Galeazzi, F. Porcelli, A. L. Schmidt, C. M. Valensise, A. Scala, W. Quattrociocchi, et al. “Economic and social consequences of human mobility restrictions under COVID-19”. In: *Proceedings of the National Academy of Sciences* 117.27 (2020), pp. 15530–15535.

Boschma, R. “Proximity and innovation: a critical assessment”. In: *Regional studies* 39.1 (2005), pp. 61–74.

Brink, T. “Organising of dynamic proximities enables robustness, innovation and growth: The longitudinal case of small and medium-sized enterprises (SMEs) in food producing firm networks”. In: *Industrial Marketing Management* 75 (2018), pp. 66–79.

Broekel, T. and R. Boschma. “Knowledge networks in the Dutch aviation industry: the proximity paradox”. In: *Journal of Economic Geography* 12.2 (2012), pp. 409–433.

Broström, A. “Working with distant researchers—Distance and content in university–industry interaction”. In: *Research Policy* 39.10 (2010), pp. 1311–1320.

Brunow, S., A. Hammer, and P. McCann. “The impact of KIBS’location on their innovation behaviour”. In: *Regional Studies* 54.9 (2020), pp. 1289–1303.

Bunduchi, R. “Trust, partner selection and innovation outcome in collaborative new product development”. In: *Production planning & control* 24.2-3 (2013), pp. 145–157.

Burt, S., U. Johansson, and J. Dawson. “International retailing as embedded business models”. In: *Journal of Economic Geography* 16.3 (2016), pp. 715–747.

Cabrer-Borras, B. and G. Serrano-Domingo. “Innovation and R&D spillover effects in Spanish regions: A spatial approach”. In: *Research Policy* 36.9 (2007), pp. 1357–1371.

Callois, J.-M. “The two sides of proximity in industrial clusters: The trade-off between process and product innovation”. In: *Journal of Urban Economics* 63.1 (2008), pp. 146–162.

Cantù, C. “Exploring the role of spatial relationships to transform knowledge in a business idea—Beyond a geographic proximity”. In: *Industrial Marketing Management* 39.6 (2010), pp. 887–897.

Cao, J., T. Xia, J. Li, Y. Zhang, and S. Tang. “A density-based method for adaptive LDA model selection”. In: *Neurocomputing* 72.7-9 (Mar. 2009), pp. 1775–1781. ISSN: 09252312. DOI: 10.1016/j.neucom.2008.06.011.

Capaldo, A. and A. M. Petruzzelli. “Partner geographic and organizational proximity and the innovative performance of knowledge-creating alliances”. In: *European Management Review* 11.1 (2014), pp. 63–84.

Carbonell, P. and A. I. Rodriguez. “Designing teams for speedy product development: The moderating effect of technological complexity”. In: *Journal of Business Research* 59.2 (2006), pp. 225–232.

Chang, J., S. Gerrish, C. Wang, J. L. Boyd-graber, and D. Blei. “Reading Tea Leaves: How Humans Interpret Topic Models”. In: *Advances in Neural Information Processing Systems 22*. Ed. by Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta. Curran Associates, Inc., 2009, pp. 288–296. URL: http://papers.nips.cc/paper/3700-reading-tea-leaves-how-humans-interpret-topic-models.pdf.

Chesbrough, H. “Business model innovation: opportunities and barriers”. In: *Long range planning* 43.2-3 (2010), pp. 354–363.

Chesbrough, H. and R. S. Rosenbloom. “The role of the business model in capturing value from innovation: evidence from Xerox Corporation’s technology spin-off companies”. In: *Industrial and corporate change* 11.3 (2002), pp. 529–555.

Choudhury, P., D. Wang, N. A. Carlson, and T. Khanna. “Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles”. In: *Strategic Management Journal* 40.11 (2019), pp. 1705–1732. DOI: 10.1002/smj.3067. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/smj.3067.

Chu, Y., X. Tian, and W. Wang. “Corporate innovation along the supply chain”. In: *Management Science* 65.6 (2019), pp. 2445–2466.

Chuang, J., C. D. Manning, and J. Heer. “Termite: Visualization Techniques for Assessing Textual Topic Models”. In: *Advanced Visual Interfaces*. Capri Island, Italy, 2012. ISBN: 9781450312875. URL: http://vis.stanford.edu/papers/termite.

Cohen, W. M. and D. A. Levinthal. “Absorptive capacity: A new perspective on learning and innovation”. In: *Administrative science quarterly* 35.1 (1990), pp. 128–152.

Coleman, J. S. “Social Capital in the Creation of Human Capital”. In: *American Journal of Sociology* 94 (1988), S95–S120. ISSN: 00029602, 15375390. URL: http://www.jstor.org/stable/2780243.

Cornish, S. L. “Product innovation and the spatial dynamics of market intelligence: does proximity to markets matter?” In: *Economic Geography* 73.2 (1997), pp. 143–165.

Corritore, M., A. Goldberg, and S. B. Srivastava. “Duality in Diversity: How Intrapersonal and Interpersonal Cultural Heterogeneity Relate to Firm Performance”. In: *Administrative Science Quarterly* 65.2 (2020), pp. 359–394. ISSN: 19303815. DOI: 10.1177/0001839219844175.

Crescenzi, R., M. Nathan, and A. Rodríguez-Pose. “Do inventors talk to strangers? On proximity and collaborative knowledge creation”. In: *Research Policy* 45.1 (2016), pp. 177–194.

Croidieu, G. and P. H. Kim. “Labor of Love: Amateurs and Lay-expertise Legitimation in the Early U.S. Radio Field”. In: *Administrative Science Quarterly* 63.1 (Mar. 2018), pp. 1–42. ISSN: 19303815. DOI: 10.1177/0001839216686531.

Cucculelli, M. and C. Bettinelli. “Business models, intangibles and firm performance: evidence on corporate entrepreneurship from Italian manufacturing SMEs”. In: *Small Business Economics* 45.2 (2015), pp. 329–350.

Czarnitzki, D. and H. Hottenrott. “Are local milieus the key to innovation performance?” In: *Journal of regional science* 49.1 (2009), pp. 81–112.

Dahl, M. S. and O. Sorenson. “Home sweet home: Entrepreneurs’ location choices and the performance of their ventures”. In: *Management science* 58.6 (2012), pp. 1059–1071.

Davids, M. and K. Frenken. “Proximity, knowledge base and the innovation process: Towards an integrated framework”. In: *Regional Studies* 52.1 (2018), pp. 23–34.

Davidson, R. and N. Poor. “Location, location, location: how digital platforms reinforce the importance of spatial proximity”. In: *Information, Communication & Society* 22.10 (2019), pp. 1464–1478.

Davis, G. F. and H. R. Greve. “Corporate elite networks and governance changes in the 1980s”. In: *American journal of sociology* 103.1 (1997), pp. 1–37.

Delanty, G. and P. Strydom. “Philosophies of social science: The classic and contemporary readings”. In: (2003).

Deveaud, R., E. SanJuan, and P. Bellot. “Accurate and effective latent concept modeling for ad hoc information retrieval”. In: *Document numérique* 17.1 (2014), pp. 61–84.

DiMaggio, P. “Adapting computational text analysis to social science (and vice versa)”. In: *Big Data & Society* 2.2 (2015), pp. 1–5.

DiMaggio, P., M. Nag, and D. Blei. “Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding”. In: *Poetics* 41.6 (2013), pp. 570–606.

Divella, M. “Cooperation linkages and technological capabilities development across firms”. In: *Regional Studies* 51.10 (2017), pp. 1494–1506.

Doldor, E., M. Wyatt, and J. Silvester. “Statesmen or cheerleaders? Using topic modeling to examine gendered messages in narrative developmental feedback for leaders”. In: *The Leadership Quarterly* 30.5 (Aug. 2019), p. 101308. ISSN: 10489843. DOI: 10.1016/j.leaqua.2019.101308.

Dolfsma, W. and R. Van der Eijk. “Distances in organizations: innovation in an R&D lab”. In: *British Journal of Management* 27.2 (2016), pp. 271–286.

Doloreux, D. and R. Shearmur. “Collaboration, information and the geography of innovation in knowledge intensive business services”. In: *Journal of economic geography* 12.1 (2012), pp. 79–105.

Doran, J., D. Jordan, and E. O’Leary. “The effects of the frequency of spatially proximate and distant interaction on innovation by Irish SMEs”. In: *Entrepreneurship & Regional Development* 24.7-8 (2012), pp. 705–727.

Dornbusch, F. and P. Neuhäusler. “Composition of inventor teams and technological progress–The role of collaboration between academia and industry”. In: *Research Policy* 44.7 (2015), pp. 1360–1375.

Drejer, I. and C. R. Ostergaard. “Exploring determinants of firms’ collaboration with specific universities: Employee-driven relations and geographical proximity”. In: *Regional Studies* 51.8 (2017), pp. 1192–1205.

Emsley, R. and H. Liu. “PARAMED: Stata module to perform causal mediation analysis using parametric regression models”. In: (2013).

Enkel, E. and S. Heil. “Preparing for distant collaboration: Antecedents to potential absorptive capacity in cross-industry innovation”. In: *Technovation* 34.4 (2014), pp. 242–260.

Fernandes, C. I. and J. J. Ferreira. “Knowledge spillovers: cooperation between universities and KIBS”. In: *R&D Management* 43.5 (2013), pp. 461–472.

Fitjar, R. D. and A. Rodríguez-Pose. “When Local Interaction Does Not Suffice: Sources of Firm Innovation in Urban Norway”. In: *Environment and Planning A: Economy and Space* 43.6 (2011), pp. 1248–1267. DOI: 10.1068/a43516.

Foss, N. J. and T. Saebi. “Fifteen years of research on business model innovation: How far have we come, and where should we go?” In: *Journal of Management* 43.1 (2017), pp. 200–227.

Freel, M. S. “Sectoral patterns of small firm innovation, networking and proximity”. In: *Research policy* 32.5 (2003), pp. 751–770.

Friedman, T. L. *The world is flat: A brief history of the twenty-first century*. Macmillan, 2006.

Funk, R. J. “Making the most of where you are: Geography, networks, and innovation in organizations”. In: *Academy of Management Journal* 57.1 (2014), pp. 193–222.

Gaba, V. and A. D. Meyer. “Crossing the organizational species barrier: How venture capital practices infiltrated the information technology sector”. In: *Academy of Management Journal* 51.5 (2008), pp. 976–998.

Geerts, A., B. Leten, R. Belderbos, and B. Van Looy. “Does spatial ambidexterity pay off? On the benefits of geographic proximity between technology exploitation and exploration”. In: *Journal of product innovation management* 35.2 (2018), pp. 151–163.

Geva, H., G. Oestreicher-Singer, and M. Saar-Tsechansky. “Using retweets when shaping our online persona: Topic modeling approach”. In: *MIS Quarterly: Management Information Systems* 43.2 (June 2019), pp. 501–524. ISSN: 21629730. DOI: 10.25300/MISQ/2019/14346.

Ghio, N., M. Guerini, and C. Rossi-Lamastra. “University knowledge and the creation of innovative start-ups: An analysis of the Italian case”. In: *Small Business Economics* 47.2 (2016), pp. 293–311.

Giorgi, S., M. Maoret, and E. J. Zajac. “On the Relationship Between Firms and Their Legal Environment: The Role of Cultural Consonance”. In: *Organization Science* 30.4 (June 2019), pp. 803–830. ISSN: 1047-7039. DOI: 10.1287/orsc.2018.1250.

Giuliani, E. “The selective nature of knowledge networks in clusters: evidence from the wine industry”. In: *Journal of economic geography* 7.2 (2007), pp. 139–168.

Giuliani, E. and M. Bell. “The micro-determinants of meso-level learning and innovation: evidence from a Chilean wine cluster”. In: *Research policy* 34.1 (2005), pp. 47–68.

Gong, J., V. Abhishek, and B. Li. “Examining the impact of keyword ambiguity on search advertising performance: A topic model approach”. In: *MIS Quarterly: Management Information Systems* 42.3 (Sept. 2018), pp. 805–829. ISSN: 21629730. DOI: 10.25300/MISQ/2018/14042.

Granovetter, M. “Economic action and social structure: The problem of embeddedness”. In: *American journal of sociology* 91.3 (1985), pp. 481–510.

— “The Strength of Weak Ties”. In: *American Journal of Sociology* 78.6 (1973), pp. 1360–1380. ISSN: 00029602, 15375390. URL: http://www.jstor.org/stable/2776392.

Griffiths, T. L. and M. Steyvers. “Finding scientific topics”. In: *Proceedings of the National academy of Sciences* 101.suppl 1 (2004), pp. 5228–5235.

Grimmer, J. and B. M. Stewart. “Text as data: The promise and pitfalls of automatic content analysis methods for political texts”. In: *Political Analysis* 21.3 (2013), pp. 267–297. ISSN: 14764989. DOI: 10.1093/pan/mps028. URL: https://doi.org/10.1093/pan/mps028.

Guan, J. C. and Y. Yan. “Technological proximity and recombinative innovation in the alternative energy field”. In: *Research Policy* 45.7 (2016), pp. 1460–1473.

Guo, X., Q. Wei, G. Chen, J. Zhang, and D. Qiao. “Extracting representative information on intra-organizational blogging platforms”. In: *MIS Quarterly: Management Information Systems* 41.4 (2017), pp. 1105–1127. ISSN: 21629730. DOI: 10.25300/MISQ/2017/41.4.05. URL: http://www.misq.org.

Haans, R. F. “What’s the value of being different when everyone is? The effects of distinctiveness on performance in homogeneous versus heterogeneous categories”. In: *Strategic Management Journal* 40.1 (Jan. 2019), pp. 3–27. ISSN: 10970266. DOI: 10.1002/smj.2978.

Hair, J. F., W. C. Black, B. J. Babin, and R. E. Anderson. *Multivariate data analysis*. Pearson College, Div, 2009.

Hall, D., D. Jurafsky, and C. D. Manning. “Studying the History of Ideas Using Topic Models”. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Honolulu, Hawaii: Association for Computational Linguistics (ACL), 2008, pp. 363–371. URL: http://dl.acm.org/citation.cfm?id=1613715.1613763.

Hall, P. V. and W. Jacobs. “Shifting proximities: The maritime ports sector in an era of global supply chains”. In: *Regional Studies* 44.9 (2010), pp. 1103–1115.

Hannigan, T. R., R. F. Haans, K. Vakili, H. Tchalian, V. L. Glaser, M. S. Wang, S. Kaplan, and P. D. Jennings. “Topic modeling in management research: Rendering new theory from textual data”. In: *Academy of Management Annals* 13.2 (2019), pp. 586–632.

Hansen, T. and J. Mattes. “Proximity and power in collaborative innovation projects”. In: *Regional Studies* 52.1 (2018), pp. 35–46.

He, J., X. Fang, H. Liu, and X. Li. “Mobile App Recommendation: An Involvement-Enhanced Approach”. In: *MIS Quarterly* 43.3 (2019).

Hedström, P. and P. Bearman. “What is Analytical Sociology All About? An Introductory Essay”. In: *booktitle*. 2011.

Hess, M. "'Spatial'relationships? Towards a reconceptualization of embedded ness". In: *Progress in human geography* 28.2 (2004), pp. 165–186.

Hong, L. and B. D. Davison. "Empirical Study of Topic Modeling in Twitter". In: *Proceedings of the first workshop on social media analytics*. Association for Computing Machinery, 2010, pp. 80–88. ISBN: 9781450302173. DOI: 10.1145/1964858.1964870. URL: https://doi.org/10.1145/1964858.1964870.

Hopp, C., D. Antons, J. Kaminski, and T. O. Salge. "The Topic Landscape of Disruption Research—A Call for Consolidation, Reconciliation, and Generalization". In: *Journal of Product Innovation Management* 35.3 (2018), pp. 458–487.

Hsiang, S., D. Allen, S. Annan-Phan, K. Bell, I. Bolliger, T. Chong, H. Druckenmiller, L. Y. Huang, A. Hultgren, E. Krasovich, et al. "The effect of large-scale anti-contagion policies on the COVID-19 pandemic". In: *Nature* 584.7820 (2020), pp. 262–267.

Hu, B. and W. Chen. "Business model ambidexterity and technological innovation performance: evidence from China". In: *Technology Analysis & Strategic Management* 28.5 (2016), pp. 583–600.

Huang, A. H., R. Lehavy, A. Y. Zang, and R. Zheng. "Analyst information discovery and interpretation roles: A topic modeling approach". In: *Management Science* 64.6 (2018), pp. 2833–2855.

Huggins, R. and A. Johnston. "Knowledge flow and inter-firm networks: The influence of network resources, spatial proximity and firm size". In: *Entrepreneurship & regional development* 22.5 (2010), pp. 457–484.

InfoCamere. *Imprese, movimento anagrafico e unità locali in Toscana: dati InfoCamere 2019*. Chamber of Commerce, Italy, 2019. URL: https://www.regione.toscana.it/-/imprese-movimento-anagrafico-e-unit%C3%A0-locali-in-toscana-dati-infocamere-2019.

Isaksson, O. H., M. Simeth, and R. W. Seifert. "Knowledge spillovers in the supply chain: Evidence from the high tech sectors". In: *Research Policy* 45.3 (2016), pp. 699–706.

ISTAT. *Dati statistici per il territorio: Regione Toscana*. Central Statistics Institute, Italy, 2020. URL: https://www.istat.it/it/files//2020/05/09_Toscana_Scheda_DEF.pdf.

Jaffe, A. B. "Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers". In: *Research Policy* 18.2 (1989), pp. 87–97.

— "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value". In: *The American Economic Review* 76.5 (1986), pp. 984–1001. ISSN: 00028282. URL: http://www.jstor.org/stable/1816464.

Jespersen, K., D. Rigamonti, M. B. Jensen, and R. Bysted. “Analysis of SMEs partner proximity preferences for process innovation”. In: *Small Business Economics* 51.4 (2018), pp. 879–904.

Johnson, S. L., H. Safadi, and S. Faraj. “The emergence of online community leadership”. In: *Information Systems Research* 26.1 (2015), pp. 165–187. ISSN: 15265536. DOI: 10.1287/isre.2014.0562.

Kapetaniou, C. and S. H. Lee. “Geographical proximity and open innovation of SMEs in Cyprus”. In: *Small Business Economics* 52.1 (2019), pp. 261–276.

Kaplan, S. and K. Vakili. “The double-edged sword of recombination in breakthrough innovation”. In: *Strategic Management Journal* 36.10 (Aug. 2015), pp. 1435–1457. ISSN: 10970266. DOI: 10.1002/smj.2294.

Kirat, T. and Y. Lung. “Innovation and proximity: territories as loci of collective learning processes”. In: *European urban and regional studies* 6.1 (1999), pp. 27–38.

Klang, D., M. Wallnöfer, and F. Hacklin. “The business model paradox: A systematic review and exploration of antecedents”. In: *International Journal of Management Reviews* 16.4 (2014), pp. 454–478.

Knoben, J. and L. A. Oerlemans. “Configurations of inter-organizational knowledge links: Does spatial embeddedness still matter?” In: *Regional Studies* 46.8 (2012), pp. 1005–1021.

— “Proximity and inter-organizational collaboration: A literature review”. In: *international Journal of management reviews* 8.2 (2006), pp. 71–89.

Kobayashi, V. B., S. T. Mol, H. A. Berkers, G. Kismihók, and D. N. Den Hartog. “Text classification for organizational researchers: A tutorial”. In: *Organizational research methods* 21.3 (2018), pp. 766–799.

Lanzolla, G. and H. T. Frankort. “The online shadow of offline signals: Which sellers get contacted in online B2B marketplaces?” In: *Academy of Management Journal* 59.1 (2016), pp. 207–231.

Lappas, T., G. Sabnis, and G. Valkanas. “The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry”. In: *Information Systems Research* 27.4 (2016), pp. 940–961. ISSN: 15265536. DOI: 10.1287/isre.2016.0674.

Larsen, K. R. and C. H. Bong. “A tool for addressing construct identity in literature reviews and meta-analyses”. In: *MIS Quarterly: Management Information Systems* 40.3 (June 2016), pp. 529–551. ISSN: 21629730. DOI: 10.25300/MISQ/2016/40.3.01.

Laursen, K., T. Reichstein, and A. Salter. “Exploring the effect of geographical proximity and university quality on university–industry collaboration in the United Kingdom”. In: *Regional studies* 45.4 (2011), pp. 507–523.

Lazzeretti, L. and F. Capone. “How proximity matters in innovation networks dynamics along the cluster evolution. A study of the high technology applied to cultural goods”. In: *Journal of Business Research* 69.12 (2016), pp. 5855–5865.

Letaifa, S. B. and K. Goglio-Primard. “How does institutional context shape entrepreneurship conceptualizations?” In: *Journal of Business Research* 69.11 (2016), pp. 5128–5134.

Letaifa, S. B. and Y. Rabeau. “Too close to collaborate? How geographic proximity could impede entrepreneurship and innovation”. In: *Journal of Business Research* 66.10 (2013), pp. 2071–2078.

Leten, B., P. Landoni, and B. Van Looy. “Science or graduates: How do firms benefit from the proximity of universities?” In: *Research Policy* 43.8 (2014), pp. 1398–1412.

Levitt, B. and J. G. March. “Organizational learning”. In: *Annual review of sociology* 14.1 (1988), pp. 319–338.

Lex, A., N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister. “UpSet: visualization of intersecting sets”. In: *IEEE transactions on visualization and computer graphics* 20.12 (2014), pp. 1983–1992.

Li, J., J. Xia, and E. J. Zajac. “On the duality of political and economic stakeholder influence on firm innovation performance: Theory and evidence from Chinese firms”. In: *Strategic Management Journal* 39.1 (2018), pp. 193–216.

Li, K., J. Qiu, and J. Wang. “Technology conglomeration, strategic alliances, and corporate innovation”. In: *Management Science* 65.11 (2019), pp. 5065–5090.

Liang, X. and A. M. Liu. “The evolution of government sponsored collaboration network and its impact on innovation: A bibliometric analysis in the Chinese solar PV sector”. In: *Research Policy* 47.7 (2018), pp. 1295–1308.

Link, A. N. and J. T. Scott. “U.S. science parks: the diffusion of an innovation and its effects on the academic missions of universities”. In: *International Journal of Industrial Organization* 21.9 (2003), pp. 1323–1356.

Lundquist, K.-J. and M. Trippl. “Distance, proximity and types of cross-border innovation systems: A conceptual analysis”. In: *Regional Studies* 47.3 (2013), pp. 450–460.

MacPherson, A. D. “Academic-industry linkages and small firm innovation: evidence from the scientific instruments sector”. In: *Entrepreneurship & Regional Development* 10.4 (1998), pp. 261–276.

Maietta, O. W. “Determinants of university–firm R&D collaboration and its impact on innovation: A perspective from a low-tech industry”. In: *Research Policy* 44.7 (2015), pp. 1341–1359.

Marrocu, E., R. Paci, and S. Usai. "Proximity, networking and knowledge production in Europe: What lessons for innovation policy?" In: *Technological Forecasting and Social Change* 80.8 (2013), pp. 1484–1498.

Marshall, A. *Principles of Economics*. 1920.

Martin, J. L. *Thinking through theory*. WW Norton, Incorporated, 2015.

Mason, K. and R. Chakrabarti. "The role of proximity in business model design: Making business models work for those at the bottom of the pyramid". In: *Industrial Marketing Management* 61 (2017), pp. 67–80.

Massa, L., C. L. Tucci, and A. Afuah. "A critical assessment of business model research". In: *Academy of Management Annals* 11.1 (2017), pp. 73–104.

McDonald, R. M. and K. M. Eisenhardt. "Parallel play: Startups, nascent markets, and effective business-model design". In: *Administrative Science Quarterly* 65.2 (2020), pp. 483–523.

McPherson, M., L. Smith-Lovin, and J. M. Cook. "Birds of a feather: Homophily in social networks". In: *Annual review of sociology* 27.1 (2001), pp. 415–444.

Mimno, D., H. M. Wallach, E. Talley, M. Leenders, and A. McCallum. "Optimizing semantic coherence in topic models". In: *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics. 2011, pp. 262–272.

Mimno, D., H. M. Wallach, E. Talley, M. Leenders, and A. Mccallum. "Optimizing Semantic Coherence in Topic Models". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2011, pp. 262–272.

Miric, M., K. J. Boudreau, and L. B. Jeppesen. "Protecting their digital assets: The use of formal & informal appropriability strategies by App developers". In: *Research Policy* 48.8 (Oct. 2019), p. 103738. ISSN: 00487333. DOI: 10.1016/j.respol.2019.01.012.

Moe, W. W. and D. A. Schweidel. "Opportunities for innovation in social media analytics". In: *Journal of Product Innovation Management* 34.5 (2017), pp. 697–702.

Mohliver, A. "How misconduct spreads: auditors' role in the diffusion of stock-option backdating". In: *Administrative Science Quarterly* 64.2 (2019), pp. 310–336.

Molina-Morales, F. X. and M. T. Martínez-Fernández. "Social networks: effects of social capital on firm innovation". In: *Journal of Small Business Management* 48.2 (2010), pp. 258–279.

Moodysson, J. "Principles and practices of knowledge creation: On the organization of "buzz" and "pipelines" in life science communities". In: *Economic Geography* 84.4 (2008), pp. 449–469.

Moodysson, J., L. Coenen, and B. Asheim. “Explaining spatial patterns of innovation: analytical and synthetic modes of knowledge creation in the Medicon Valley life-science cluster”. In: *Environment and planning A* 40.5 (2008), pp. 1040–1056.

Moodysson, J. and O. Jonsson. “Knowledge collaboration and proximity: The spatial organization of biotech innovation projects”. In: *European urban and regional studies* 14.2 (2007), pp. 115–131.

Morescalchi, A., F. Pammolli, O. Penner, A. M. Petersen, and M. Riccaboni. “The evolution of networks of innovators within and across borders: Evidence from patent data”. In: *Research Policy* 44.3 (2015), pp. 651–668.

Morgan, K. “The exaggerated death of geography: learning, proximity and territorial innovation systems”. In: *Journal of economic geography* 4.1 (2004), pp. 3–21.

Nelson, R. and S. Winter. *An evolutionary theory of economic change*. Harvard University Press, 1982.

Nicholson, J., E. Gimmon, and C. Felzensztein. *Economic geography and business networks: creating a dialogue between disciplines: An introduction to the special issue*. 2017.

Nooteboom, B. “Innovation and inter-firm linkages: new implications for policy”. In: *Research policy* 28.8 (1999), pp. 793–805.

Norbutas, L. “Offline constraints in online drug marketplaces: An exploratory analysis of a cryptomarket trade network”. In: *International Journal of Drug Policy* 56 (2018), pp. 92–100.

Ocasio, W., T. Laamanen, and E. Vaara. “Communication and attention dynamics: An attention-based view of strategic change”. In: *Strategic Management Journal* 39.1 (Jan. 2018), pp. 155–167. ISSN: 10970266. DOI: 10.1002/smj.2702.

Oerlemans, L. and M. Meeus. “Do organizational and spatial proximity impact on firm performance?” In: *Regional studies* 39.1 (2005), pp. 89–104.

Owen-Smith, J. and W. W. Powell. “Knowledge networks as channels and conduits: The effects of spillovers in the Boston biotechnology community”. In: *Organization science* 15.1 (2004), pp. 5–21.

Panichella, A., B. Dit, R. Oliveto, M. Di Penta, D. Poshyvanyk, and A. De Lucia. “How to effectively use topic models for software engineering tasks? an approach based on genetic algorithms”. In: *Proceedings of the 2013 International Conference on Software Engineering*. IEEE Press. 2013, pp. 522–531.

Parent, O. and S. Riou. “Bayesian analysis of knowledge spillovers in European regions”. In: *Journal of Regional Science* 45.4 (2005), pp. 747–775.

Parra-Requena, G., M. J. Ruiz-Ortega, P. M. García-Villaverde, and J. Rodrigo-Alarcon. “The mediating role of knowledge acquisition on the relationship between external social capital and innovativeness”. In: *European Management Review* 12.3 (2015), pp. 149–169.

Pati, R. K., M. Nandakumar, A. Ghobadian, R. D. Ireland, and N. O'Regan. "Business model design–performance relationship under external and internal contingencies: Evidence from SMEs in an emerging economy". In: *Long Range Planning* 51.5 (2018), pp. 750–769.

Phene, A., K. Fladmoe-Lindquist, and L. Marsh. "Breakthrough innovations in the US biotechnology industry: the effects of technological space and geographic origin". In: *Strategic management journal* 27.4 (2006), pp. 369–388.

Pillai, R. G. and V. Bindroo. "Supplier cluster characteristics and innovation outcomes". In: *Journal of Business Research* (2019).

Ponds, R., F. v. Oort, and K. Frenken. "Innovation, spillovers and university–industry collaboration: an extended knowledge production function approach". In: *Journal of Economic Geography* 10.2 (2010), pp. 231–255.

Porter, M. E. et al. "Clusters and competition: New agendas for companies, governments, and institutions". In: (1998).

Porter, M. E. "The competitive advantage of nations". In: *Harvard business review* 68.2 (1990), pp. 73–93.

Porter, M. E. and M. ilustraciones Gibbs. "Strategy and the Internet". In: (2001).

Powers, D. M. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation". In: *Journal of Machine Learning Technologies* 2.1 (2011), pp. 37–63.

Preacher, K. J. and A. F. Hayes. "Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models". In: *Behavior research methods* 40.3 (2008), pp. 879–891.

Presutti, M., C. Boari, A. Majocchi, and X. Molina-Morales. "Distance to Customers, Absorptive Capacity, and Innovation in High-Tech Firms: The Dark Face of Geographical Proximity". In: *Journal of Small Business Management* 57.2 (2019), pp. 343–361.

Pucci, T. "Il Modello di Business: Caratteri Strutturali e Dinamiche Evolutive". In: (2016).

Quinn, K. M., B. L. Monroe, M. Colaresi, M. H. Crespin, and D. R. Radev. "How to Analyze Political Attention with Minimal Assumptions and Costs". In: *American Journal of Political Science* 54.1 (2010), pp. 209–228. DOI: 10.1111/j.1540-5907.2009.00427.x. URL: https://www.jstor.org/stable/20647980.

Rammer, C., J. Kinne, and K. Blind. "Knowledge proximity and firm innovation: A microgeographic analysis for Berlin". In: *Urban Studies* 57.5 (2020), pp. 996–1014.

Roberts, M. E., B. M. Stewart, D. Tingley, et al. "stm: R package for structural topic models". In: *Journal of Statistical Software* 10.2 (2014), pp. 1–40.

Romijn, H. and M. Albaladejo. “Determinants of innovation capability in small electronics and software firms in southeast England”. In: *Research policy* 31.7 (2002), pp. 1053–1067.

Ruckman, K. and I. McCarthy. “Why do some patents get licensed while others do not?” In: *Industrial and Corporate Change* 26.4 (Aug. 2017), pp. 667–688. ISSN: 14643650. DOI: 10.1093/icc/dtw046.

Schmiedel, T., O. Müller, and J. vom Brocke. “Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture”. In: *Organizational Research Methods* 22.4 (2019), pp. 941–968.

Schumpeter, J. “The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle.” In: (1934).

Schwartz, M., F. Peglow, M. Fritsch, and J. Günther. “What drives innovation output from subsidized R&D cooperation?—Project-level evidence from Germany”. In: *Technovation* 32.6 (2012), pp. 358–369.

Shearmur, R. “Innovation, regions and proximity: from neo-regionalism to spatial analysis”. In: *Regional Studies* 45.9 (2011), pp. 1225–1243.

Shearmur, R. and D. Doloreux. “Knowledge-intensive business services (KIBS) use and user innovation: high-order services, geographic hierarchies and internet use in Quebec’s manufacturing sector”. In: *Regional Studies* 49.10 (2015), pp. 1654–1671.

Shi, Z., G. M. Lee, and A. B. Whinston. “Toward a Better Measure of Business Proximity: Topic Modeling for Industry Intelligence.” In: *MIS quarterly* 40.4 (2016).

Sievert, C. and K. E. Shirley. “LDAvis: A method for visualizing and interpreting topics”. In: *Proceedings of the workshop on interactive language learning, visualization, and interfaces*. 2014, pp. 63–70.

Sieweke, J. and S. Santoni. “Natural experiments in leadership research: An introduction, review, and guidelines”. In: *The Leadership Quarterly* 31.1 (2020), p. 101338.

Sonn, J. W. and M. Storper. “The increasing importance of geographical proximity in knowledge production: an analysis of US patent citations, 1975–1997”. In: *Environment and Planning A* 40.5 (2008), pp. 1020–1039.

Steinmo, M. and E. Rasmussen. “How firms collaborate with public research organizations: The evolution of proximity dimensions in successful innovation projects”. In: *Journal of Business Research* 69.3 (2016), pp. 1250–1259.

Stephan, A. “Are public research spin-offs more innovative?” In: *Small Business Economics* 43.2 (2014), pp. 353–368.

Sternberg, R. “Innovative linkages and proximity: empirical results from recent surveys of small and medium sized firms in German regions”. In: *Regional Studies* 33.6 (1999), pp. 529–540.

Still, M. C. and D. Strang. “Who does an elite organization emulate?” In: *Administrative Science Quarterly* 54.1 (2009), pp. 58–89.

Taddy, M. “On estimation and selection for topic models”. In: *Artificial Intelligence and Statistics*. 2012, pp. 1184–1193.

Takhteyev, Y., A. Gruzd, and B. Wellman. “Geography of Twitter networks”. In: *Social networks* 34.1 (2012), pp. 73–81.

Taura, N. and D. Radicic. “Intra-cluster knowledge exchange and frequency of product innovation in a digital cluster”. In: *Journal of Small Business Management* 57.sup2 (2019), pp. 350–373.

Teece, D. J. “Business models, business strategy and innovation”. In: *Long range planning* 43.2-3 (2010), pp. 172–194.

Teece, D. J., G. Pisano, and A. Shuen. “Dynamic capabilities and strategic management”. In: *Strategic management journal* 18.7 (1997), pp. 509–533.

Ter Wal, A. L. “The dynamics of the inventor network in German biotechnology: geographic proximity versus triadic closure”. In: *Journal of Economic Geography* 14.3 (2014), pp. 589–620.

Thomas III, L. “Are we all global now? Local vs. foreign sources of corporate competence: the case of the Japanese pharmaceutical industry”. In: *Strategic Management Journal* 25.8-9 (2004), pp. 865–886.

Torre, A. “On the role played by temporary geographical proximity in knowledge transmission”. In: *Regional studies* 42.6 (2008), pp. 869–889.

Torre, A. and J.-P. Gilly. “On the analytical dimension of proximity dynamics”. In: *Regional studies* 34.2 (2000), pp. 169–180.

Torre, A. and A. Rallet. “Proximity and localization”. In: *Regional studies* 39.1 (2005), pp. 47–59.

Uzzi, B. “Social structure and competition in interfirm networks: The paradox of embeddedness”. In: *Administrative science quarterly* (1997), pp. 35–67.

Valeri, L. and T. J. VanderWeele. “Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros.” In: *Psychological methods* 18.2 (2013), p. 137.

Van de Ven, A. H. and M. S. Poole. “Alternative approaches for studying organizational change”. In: *Organization studies* 26.9 (2005), pp. 1377–1404.

Van de Vrande, V. “Balancing your technology-sourcing portfolio: How sourcing mode diversity enhances innovative performance”. In: *Strategic Management Journal* 34.5 (2013), pp. 610–621.

Wallach, H. M., I. Murray, R. Salakhutdinov, and D. Mimno. “Evaluation Methods for Topic Models”. In: *Proceedings of the 26th annual international conference on machine learning*. 2009, pp. 1105–1112.

DOI: 10.1145/1553374.1553515. URL: https://doi.org/10.1145/1553374.1553515.

Wang, C. C. and A. Wu. "Geographical FDI knowledge spillover and innovation of indigenous firms in China". In: *International Business Review* 25.4 (2016), pp. 895–906.

Wang, Q., B. Li, and P. V. Singh. "Copycats vs. original mobile apps: A machine learning copycat-detection method and empirical analysis". In: *Information Systems Research* 29.2 (June 2018), pp. 273–291. ISSN: 15265536. DOI: 10.1287/isre.2017.0735.

Weidenfeld, A. "Tourism and cross border regional innovation systems". In: *Annals of Tourism Research* 42 (2013), pp. 191–213.

Weidenfeld, A., A. M. Williams, and R. W. Butler. "Knowledge transfer and innovation among attractions". In: *Annals of tourism research* 37.3 (2010), pp. 604–626.

Weterings, A. and R. Boschma. "Does spatial proximity to customers matter for innovative performance?: Evidence from the Dutch software sector". In: *Research Policy* 38.5 (2009), pp. 746–755.

Weterings, A. and S. Koster. "Inheriting knowledge and sustaining relationships: What stimulates the innovative performance of small software firms in the Netherlands?" In: *Research Policy* 36.3 (2007), pp. 320–335.

Whittington, K. B., J. Owen-Smith, and W. W. Powell. "Networks, propinquity, and innovation in knowledge-intensive industries". In: *Administrative science quarterly* 54.1 (2009), pp. 90–122.

Williamson, O. "Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Economics of Internal Organization". In: (1975).

Wirtz, B. W., A. Pistoia, S. Ullrich, and V. Göttel. "Business models: Origin, development and future research perspectives". In: *Long range planning* 49.1 (2016), pp. 36–54.

Wu, L. "Social network effects on productivity and job security: Evidence from the adoption of a social networking tool". In: *Information Systems Research* 24.1 (2013), pp. 30–51. ISSN: 15265536. DOI: 10.1287/isre.1120.0465.

Wullum Nielsen, M. and L. Börjeson. "Gender diversity in the management field: Does it matter for research outcomes?" In: *Research Policy* 48.7 (Sept. 2019), pp. 1617–1632. ISSN: 00487333. DOI: 10.1016/j.respol.2019.03.006.

Yue, W. T., Q. H. Wang, and K. L. Hui. "See no evil, hear no evil? Dissecting the impact of online hacker forums". In: *MIS Quarterly: Management Information Systems* 43.1 (Mar. 2019), pp. 73–95. ISSN: 21629730. DOI: 10.25300/MISQ/2019/13042.

Zeller, C. “North Atlantic innovative relations of Swiss pharmaceuticals and the proximities with regional biotech arenas”. In: *Economic geography* 80.1 (2004), pp. 83–111.

Zott, C. and R. Amit. “Business model design and the performance of entrepreneurial firms”. In: *Organization science* 18.2 (2007), pp. 181–199.

— “Business model design: an activity system perspective”. In: *Long range planning* 43.2-3 (2010), pp. 216–226.

— “The fit between product market strategy and business model: implications for firm performance”. In: *Strategic management journal* 29.1 (2008), pp. 1–26.

Zott, C., R. Amit, and L. Massa. “The business model: recent developments and future research”. In: *Journal of management* 37.4 (2011), pp. 1019–1042.

Zuckerman, E. W., T.-Y. Kim, K. Ukanwa, and J. Von Rittmann. “Robust identities or nonentities? Typecasting in the feature-film labor market”. In: *American journal of sociology* 108.5 (2003), pp. 1018–1074.

Zukin, S. and P. DiMaggio. *Structures of capital: The social organization of the economy*. CUP Archive, 1990.