Combining geographical and semantic proximity to measure spillovers. The case of Sweden

Alessandro Marra[1,2,3] and Andrea D'Isidoro[1]

[1] Department of Economics, University d'Annunzio of Chieti-Pescara, Italy
[2] GRIF research centre, LUISS University, Rome, Italy
[3] BLISS research centre, University Ca Foscari, Venice, Italy

This work is part of a research program funded by the Italian Ministry of University and Research as a Project of Relevant National Interest (PRIN 2022). The project aims to use textual data to define new measures of proximity through which to measure the spillover effects on firm performance (Grant Prot. 20224EATBN)

Abstract: To exchange knowledge, it is necessary to be physically close and share some expertise. We aim to combine a geographical adjacency matrix with a non-geographical one to investigate the existence of spillovers between firms. The latter matrix is designed to replicate semantic proximity and constructed using web-derived data, capturing firms' expertise about industrial specializations and adopted technologies. We maintain and test that startups generate knowledge spillovers, which positively impact the performance of their neighbours. Results show that firm's economic performance depends not only on its intrinsic characteristics such as its initial scale or growth stage but also, and more appreciably, on the spillover effects that arise from both geographical and semantic proximity. These effects were most pronounced when both forms of proximity were combined optimally.

1. Introduction

The advantages deriving from being physically nearby are unanimously accepted in the literature: geographical proximity facilitates informal interactions, collaboration, and the transfer of knowledge (Audretsch and Feldman, 1996). However, to exchange knowledge effectively, it is also essential to share a certain level of expertise (McCann and Ortega-Argilés, 2016).

Most of the research on proximity and knowledge exchange is grounded in robust theoretical frameworks, such as related variety (Frenken et al., 2007; Neffke and Henning, 2008), absorptive capacity (Cohen and Levinthal, 1990; Fritsch and Kublina, 2018), and recombinant innovation (Zhang et al., 2019; Li et al., 2021). These studies consistently highlight the importance of a common knowledge base to foster effective interactions (Nooteboom et al., 2007). That is why additional dimensions of non-geographical adjacency, such as cognitive, industrial and technological proximity, come into play (Cortinovis et al., 2020; Amoroso et al., 2023).

Scholars often incorporate such non-geographical dimensions of proximity alongside the geographical one, since their combinations may offer deeper insights into the investigation of knowledge flows (Liu and Ma, 2019; Lopolito et al., 2022; Panori at al., 2022). In this paper, we aim to combine a geographical adjacency matrix with a non-geographical one to investigate the existence of spillover effects between companies. The latter matrix is designed to replicate semantic proximity, capturing firms' expertise about specializations and technologies. Our contribution to the literature is twofold. Firstly, we develop a new semantic measure, based on web-derived data, to assess how similar companies are in terms of their industrial specializations and technological focus. This adds to measures built on industry and patent classification systems, which often fail to reflect firms' real and evolving activities, especially in technological sectors, due to their rigidity and slow updates. Relying on a single code or outdated taxonomy can misrepresent firms' profiles, limiting the accuracy of proximity and relatedness measures (Nathan and Rosso, 2015; Marra and Baldassari, 2022). Secondly, to the best of our knowledge, this is one of the few empirical attempts to apply a convex combination of a geographical adjacency matrix with a semantic proximity matrix to model spillover effects between firms (Sheng and LeSage, 2021; Debarsy and LeSage, 2022). The choice is justified by the ability to capture, more rigorously, interactions between firms (Parent and LeSage, 2008; Debarsy and LeSage, 2021).

By combining geographical and semantic proximity matrices, we posit that firms generate knowledge spillovers, which positively impact the performance of their neighbours (Ebert et al., 2019; Zhou et al., 2019a; Martin-Rios et al., 2022). We test our hypothesis using a parsimonious model of firm performance, with sales growth as the dependent variable and a few covariates: initial turnover (used as a proxy to identify scaleups, i.e., startups with rapid revenue growth), average number of employees (serving as a proxy for the firm's knowledge base, with each employee contributing distinct expertise), and growth stage (to distinguish between early-stage and more mature firms). The adopted framework is rooted in spatial econometrics, emphasizes local rather than global indirect effects, and avoids potentially misleading endogenous spatial lags (Elhorst, 2014; LeSage and Pace, 2009). By preventing endogeneity, the model enhances interpretability and reduces the risk of biases associated with omitted variables (Corrado and Fingleton, 2012).

Our results show that firm's economic performance depends not only on its intrinsic characteristics such as its initial scale or growth stage but also, and appreciably, on the spillover effects that arise from both geographical and semantic proximity. These effects were most pronounced when both forms of proximity were combined optimally.

The paper is organized as follows. Section 2 reviews the relevant literature, focusing on empirical studies that integrate both geographical and non-geographical adjacency matrices. Section 3 introduces the dataset, which includes information from the Dealroom database on Swedish tech companies, supplemented by additional text data gathered from the web. Section 4 details the methodology, including the construction of the semantic proximity measure and the integration of both geographical and semantic matrices. Section 5 presents and interprets the results, and provides some robustness checks. Finally, Section 6 concludes by addressing limitations and suggesting avenues for future research.

2. Literature

Companies located near one another may enhance their economic performance by enabling faster and more efficient resource sharing, reducing transaction costs, and improving access to specialized labour and suppliers (Nilsson, 2019; Sharma et al., 2024). Moreover, firms located in close geographical proximity are better positioned to benefit from knowledge spillovers (Breschi and Lissoni, 2001; Döring and Schnellenbach, 2006). As widely discussed across various theoretical frameworks, such as related variety, absorptive capacity, and recombinant innovation, knowledge exchange requires not only physical proximity but also a shared expertise between firms.

2.1. Geographical and non-geographical proximity

The literature on the related variety strand shows how firms benefit from a diversity of industries or knowledge bases that are similar or 'related' in nature, which contributes to regional growth and development (Frenken et al., 2007; Van Oort et al., 2015; Cortinovis and Van Oort, 2015). Absorptive capacity emphasizes that firms must possess a certain level of prior knowledge and internal expertise to effectively absorb and integrate external knowledge (Cohen and Levinthal, 1990). Recombinant innovation emphasizes how firms generate new innovations by recombining existing knowledge in novel ways, a process that occurs in the presence of cognitive, industrial, or technological proximity (Li et al., 2021; Zhang et al., 2019).

What emerges from these different strands of literature is that companies operating in the same or similar sectors are better positioned to collaborate and exchange knowledge, thereby fostering innovation and gaining competitive advantages (Quatraro, 2010; Frenken et al., 2007; Neffke and Henning, 2008). This proximity fosters synergies because firms with expertise on similar specializations or production processes often face comparable challenges and can jointly solve them more efficiently. Losurdo et al. (2019) propose an original measure of industrial specializations in digital sectors, moving away from activity codes and employing information on adopted technologies. Davids and Frenken (2018) argue about the role of different proximity dimensions, such as cognitive and organizational, depending on the stage of product development, distinguishing between research, development, and marketing phases. Literature on technological proximity highlights how firms that adopt similar technologies benefit from each other's advancements and innovations (Aldieri, 2013; Kogler et al., 2013; McCann, 2014). Such a shared technological foundation allows for easier knowledge transfer, enabling firms to leverage external innovations more quickly and effectively. Technological proximity can lead to more frequent and productive collaborations, as firms with overlapping capabilities are more likely to engage in cooperative research and development (R&D), share technological insights, or develop complementary products. Whittle (2020) shows that the production of technological knowledge exhibits strong path dependency, with firms more likely to diversify concentrically.

2.2. Combining geographical and non-geographical proximity

Researchers have increasingly explored the advantages of using multiple proximity dimensions to gain more insights about how firms exchange knowledge (Boschma et al., 2009). While geographical proximity traditionally facilitates face-to-face interactions and informal exchanges, it alone cannot fully capture the complexity of knowledge flows. Similarity in the firms' knowledge base and expertise enable to communicate effectively and absorb information reciprocally.

Liu and Ma (2019) investigate the interaction between geographical and technological proximity in recombinant innovation, finding that low technological proximity within dense R&D regions enhances the potential for innovation. Golra et al. (2024) study the role of informal networks in manufacturing clusters, showing that geographical and non-geographical proximities play distinct roles in product and process innovation networks. Li et al. (2021) explore the relationship between regional co-location and the combination of unrelated technologies in solar photovoltaics, suggesting that regional proximity fosters the recombination of diverse technologies.

Our work contributes to the literature by adopting a methodology that gives equal importance to the geographical and semantic dimensions, rather than reducing the former to a simple classification of firms within the same territorial unit or to geographic contiguity between administrative areas. By combining different layers of proximity, scholars argue whether and to what extent firms can leverage commonalities, enhance innovation, and perform successfully. Jespersen et al. (2018) show that technological proximity plays a key role in bridging geographical and market distances, enhancing the potential for collaboration and innovation. Cao et al. (2019) investigate the interactions between different forms of proximity, finding that some dimensions can offset the lack of geographical proximity, while others support scientific collaboration. Multiple integrated layers allow for a deeper analysis of knowledge spillovers (Lopolito et al., 2022; Panori et al., 2022). Marra et al. (2024) provide a new methodology to measure business proximity using text data, compare it with standard measures based on activity codes, and propose a spatial model to account simultaneously for geographical and business proximity.

2.3. Linking proximity and performance

There is an extensive literature linking proximity to firms' performance (Tubiana et al., 2022; Martin-Rios et al, 2022). Both industrial and technological proximity have been shown to enhance firm performance: firms with higher degrees of relatedness are often better positioned to innovate and maintain a competitive edge in their respective markets (Cortinovis et al., 2020; Content et al., 2022; Jespersen et al., 2018). Freitas et al. (2024) show that technological proximity between existing and new industries enhances the chance of economic success. Colombelli and Quatraro (2019) show that green start-ups benefit from diverse and heterogeneous knowledge sources, particularly in related and complementary technological fields. Aarstad et al. (2016) find that firms in related but not identical fields benefit positively in terms of innovation capacity, while firms from entirely different sectors suffer productivity losses due to insufficient knowledge exchange. Grillitsch and Nilsson (2019) study knowledge spillovers on high and low growth firms, finding that such externalities enable the former to surge ahead while helping low growth firms to catch up. Abbasiharofteh et al. (2023) show that strong cross-field connections are positively correlated with firms' innovation levels.

Sometimes, models employ both the geographical and non-geographical dimensions. Boschma et al. (2009) observe that knowledge spillovers between neighbouring firms enhance productivity, but they also point out that when firms are too similar, the effect can be detrimental. Similarly, Timmermans and Boschma (2014) highlight the need for an optimal balance to effectively leverage proximity and inter-firm relationships. Kaneva et al. (2023) explore the roles of spatial and non-spatial proximities in knowledge creation, finding that cognitive proximity boosts knowledge spillovers and innovation, whereas technological proximity does not. Shkolnykova (2023) examines how different proximity dimensions impact the innovation performance of biotechnology SMEs in Germany, showing a mixed impact of geographical and cognitive proximity on innovation. Marra et al. (2024) employ both geographical and non-geographical matrices, assigning the former to the error term and deriving the latter from textual data, to investigate spillovers on firms' sales growth.

Building on the existing literature, we adopt the convex combination framework, which enables the integration of multiple proximity structures, each weighted by its own parameter that captures its relative contribution (Debrasy and LeSage, 2021).

3. Data

We gathered data on Sweden's tech companies from Dealroom, a commercial database that integrates machine learning and data engineering with user-submitted information and verification processes.

Sweden has emerged as one of Europe's most successful tech ecosystems, particularly in the startup realm. The country's entrepreneurial landscape is supported by government funding and investments aimed at nurturing startups and scaleups, alongside broader initiatives designed to foster innovation. Sweden has solidified its status as a leading European tech hub, with its tech companies collectively valued at approximately $239 billion, including 41 unicorns (that is, companies valued at over $1 billion). In 2023, venture capital investments in the country reached around €4.7 billion (Dealroom, 2024).

The sample of firms was defined based on the dataset built by Dealroom, as part of a report produced in partnership with Startup Sweden, the Swedish Agency for Economic and Regional Growth, the Swedish Institute, Business Sweden, and Vinnova, on a 2024 survey (Dealroom, 2024).

The geographical distribution of the observed companies is concentrated in and around the largest cities in the south of Sweden such as Stockholm, Gothenburg, and Malmö (Figure 1).

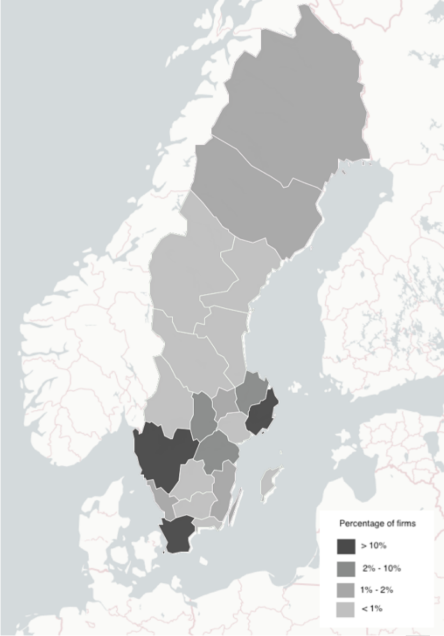

Figure 1: Geographical distribution of firms

It reflects the overall distribution of the 6094 funded companies of the Sweden's tech ecosystem. Table 1 shows the correspondence between the distributions of tech firms in Sweden by county (NUTS3 level).

Table 1. Distribution of tech firms in Sweden by county (NUTS3): population vs. sample.

County	Population	Sample
Stockholm	52.59%	55.68%
Västra Götaland	13.37%	12.99%
Skåne	12.50%	12.18%
Uppsala	3.23%	3.25%
Örebro	2.82%	2.44%
Östergötland	2.67%	2.44%
Västerbotten	1.84%	1.79%
Norrbotten	1.26%	1.14%
Kalmar	1.15%	0.97%
Halland	1.08%	0.97%
Västmanland	1.00%	0.81%
Jönköping	0.97%	0.81%
Blekinge	0.89%	0.49%
Västernorrland	0.84%	0.81%
Gävleborg	0.79%	0.81%
Värmland	0.77%	0.81%
Kronoberg	0.54%	0.32%
Jämtland	0.53%	0.32%
Dalecarlia	0.49%	0.32%
Södermanland	0.36%	0.32%
Gotland	0.31%	0.32%

Using hyperlinks to corporate websites, we generated text data for the observed units. As detailed below, we use this text data to compute a measure of semantic proximity, reflecting firms' expertise in industrial specializations and adopted technologies, to investigate spillover effects.

Observed firms are characterized by a strong technological focus and operate across a wide range of sectors. For the statistical model of firms' performance, we focus on a subset of 616 companies with available data on sales, number of employees, and geographical location. The sample is robust and representative of the entire population of tech firms in Sweden, ensuring comprehensive coverage not only in terms of geographical distribution but also in terms of company size and industrial sector, in line with standard statistical practices (Autant-Bernard and LeSage, 2011).

Table 2 reports the distribution of firms by size, while Table 3 presents the breakdown across different tech industries.

Table 2. Distribution of tech firms in Sweden by firm size: population vs. sample.

Size	Population	Sample
Micro (<10 employees)	25.01%	20.72%
Small (10-50 employees)	51.08%	52.92%
Medium (50-250 employees)	19.33%	22.47%
Large (> 250 employees)	4.58%	3.89%

Table 3. Distribution of tech firms in Sweden by industry: population vs. sample.

Industries	Population	Sample
Software enterprise	14.00%	17.09%
Health	14.74%	13.03%
Fintech	9.70%	11.19%
Energy	8.80%	5.78%
Media	4.78%	5.41%
Marketing	5.41%	5.41%
Real estate	4.30%	4.67%
Transportation	5.47%	3.93%
Gaming	3.33%	3.56%
Food	3.81%	3.44%
Security	2.29%	2.95%
Education	2.01%	2.70%
Fashion	2.91%	2.46%
Home living	2.49%	2.21%
Telecom	1.80%	2.09%
Jobs recruitment	1.87%	1.84%
Sports	1.87%	1.72%
Other	10.74%	10.83%

4. Methodology

This Section is divided into three subsections. The first subsection outlines the steps taken to construct the semantic matrix. The second subsection explains the methodology for combining the geographical and semantic matrices. The third subsection details the procedure for selecting the optimal model specification.

4.1. The semantic proximity matrix

Industry codes raise several issues. For example, when companies are established, they typically declare the activity code that most closely matches their business model and expertise. However, this declared code often fails to accurately reflect the firm's actual activities, especially as the company evolves expanding into new markets, developing new products and services, or building novel capabilities. Moreover, these codes are rarely updated as firms shift their specializations.

Using a single activity code to characterize a firm's activity is particularly problematic for technological companies, which often change strategies rapidly and operate across sectors. Even though classification systems are periodically revised, they struggle to keep up with emerging business trends. There is often a trade-off between capturing the novelty of rapidly changing industries and the need to recognize and formalize these activities through official classification. The same applies to technological classes, within which patents are registered and used to define a firm's technological profile and, consequently, to estimate their technological proximity. As a result, current industry and patent classification systems tend to lag behind real-world dynamics and are often too rigid to adequately capture modern industrial and technological complexity.

Accordingly, we use web data to profile companies with respect to their expertise about industrial specializations and adopted technologies (Marra et al., 2024).

The body of literature leveraging textual descriptions of industrial activities and technological advancements, such as those found on company websites and other online sources, is steadily growing (Nathan and Rosso, 2015; Papagiannidis et al., 2017; Cicerone et al., 2024). Beyond the technical aspects, what we aim to highlight here is the broad and diverse range of proposed applications.

Nathan and Rosso (2015) illustrate how text data can deepen our understanding of digital industries. Kinne and Lenz (2021) demonstrate the ability of text analysis and big data to uncover collaboration networks, supply chains, and innovative outcomes. Qin et al. (2021) use topic modelling for measuring cognitive proximity by mining patent description texts. Peng et al. (2023) identify critical technologies through text analysis to assist firms in discovering technology opportunities. Zhou et al. (2019b) detect typical research patterns to identify technologies for effective technological recombination. Qi et al. (2022) investigate partner selection for collaborative innovation by mining the content of patent documents. Marra et al. (2020) use text data provided by Crunchbase to rationalize merger and acquisition strategies in high-tech industries. Russo et al. (2022) map the potential application of Internet of Things technologies by using textual analysis to identify NACE codes associated with five technological domains. Similarly, with respect to artificial intelligence, Kinne and Axenbeck (2020) employ text data from over two million company websites to discover an emerging innovation ecosystem. Petralia (2020) develops a complex indicator to capture the key features of general-purpose technologies in patent data. Marra and Baldassari (2022) classify firms and identify technological trajectories across industries using text data from company websites. Dahlke et al. (2024) train a transformer language model on text data from over one million websites to identify firm-level AI adoption and its relation to firms' performance.

The methodology for the construction of the semantic proximity matrix consists of a few steps (Marra and Baldassari, 2022).

We initiate the profiling process by retrieving indexed textual content from company websites using structured search engine queries. This step ensures that we gather firm-level information that is up-to-date and self-described, thus reflecting the company's current positioning. Then, keyword extraction allows to isolate a first set of specific terms ('entities') that capture the firm's core industrial specializations and technologies.

To enable comparability across firms, we apply Latent Dirichlet Allocation (LDA), a well-established probabilistic topic modeling technique. LDA effectively identifies latent topics providing a structured representation of firms' technological and industrial domains. LDA allows us to assign broader thematic categories ('topics'), facilitating higher-level generalization while preserving relevant detail. This approach enables scalable analysis across large textual datasets and reduces dimensionality while capturing meaningful patterns in firms' language use. Moreover, by mapping firms onto a common set of topics, LDA increases informational redundancy across otherwise heterogeneous textual descriptions, thereby enhancing comparability and enabling the identification of proximity and spillovers based on shared semantic content.

The semantic matrix is inherently sensitive to the quality of textual data available for each firm. Poor or sparse firm-level information can result in less accurate representations. Where entity-level data is insufficient (specifically, when the profile contained two or fewer keywords, typically due to limited publicly available textual content), we conduct a semantic enrichment process. This is justified to avoid the exclusion of otherwise relevant firms and to prevent bias due to missing data. The enrichment is performed using Generative Pre-trained Transformer or GPT. These tools allow us to augment the original profiles with semantically coherent terms, enhancing both the internal consistency and the external validity of firm characterizations. This enrichment was applied in a limited and careful manner to minimize artificial manipulation and preserve the integrity of the original data.

Then, keywords are pre-processed using standard natural language processing (NLP) techniques, such as tokenization, lemmatization, and stop-word removal.

Table 4 provides a couple of examples.

Firm X is assigned a set of entities that describe its industrial specialization and technological focus, as well as a set of more general topics capturing its broader profile. In contrast, Firm Y shows a limited number of entities due to the scarce information on its website, which also results in a smaller number of extracted topics. In such cases, the semantic enrichment step allows to complete the profile by integrating keywords.

Table 4. Example of firm-level profiling.

Firm	Entities	Topics	Semantics
X	financial services, fraud detection, cloud computing, real time processing, SaaS	enterprise software solutions, fraud analytics, digital banking, real-time data	
Y	neural networks, credit scoring	financial services, data analytics, artificial intelligence	risk assessment, credit risk modeling, machine learning, predictive modeling

Each firm is assigned a final vector of tokens, which is used to calculate the cosine similarity for each pair of firms. Cosine similarity is well-suited as it captures similarity in orientation rather than scale, making it robust to variations in vectors length. In addition, it allows for a refined comparison of semantic content, even when firms differ significantly in the quantity of textual information available. Its computational efficiency makes it particularly suitable for large-scale pairwise comparisons across extensive firm datasets.

Lastly, in line with Boschma (2005) and Nooteboom et al. (2007), we convert the cosine similarity's linear relationships into an inverted-U shaped curve, that is our semantic proximity matrix, meant to replicate the chance of knowledge exchange. This transformation aligns with a broad body of empirical evidence suggesting that knowledge exchange tends to follow a non-linear pattern with respect to proximity. Specifically, the benefits of proximity are maximized at intermediate levels, when firms are sufficiently close to facilitate mutual understanding and interaction, but not so similar as to limit the diversity of ideas or induce redundancy (Kok et al., 2020; Marra et al., 2019, 2024). When proximity is too low, cognitive gaps hinder effective absorption of new ideas and information. Conversely, when proximity is too high, the overlap in capabilities and knowledge bases may reduce opportunities for novel combinations and learning. The inverted-U transformation thus captures this dynamic, reflecting how moderate levels of proximity are most conducive to innovation and performance gains.

Accordingly, we estimate a spatial weight matrix (W), assumed to be exogenous and constructed using a hyperbolic function. The diagonal elements are set to zero. Moreover, if two firms *i* and *j* exhibit either very low or very high cosine similarity, the corresponding weight $w_{ij}$is set to zero. Following the rationale of Marra et al. (2024), since intermediate levels of cosine similarity are likely to have the strongest impact on adjacency and spillover effects, w_{ij} increases toward one as similarity approaches the third quartile of its distribution. Beyond this point—both below and above—the adjacency weight decreases. In practical terms, the conversion from semantic proximity to adjacency follows a Gaussian kernel centered around the third quartile of the cosine similarity distribution:

$$w_{ij} = exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{\gamma}\right)^2\right) \qquad [1]$$

where distance d_{ij} is the difference between the cosine similarity value and the third quartile value (which corresponds to the maximum adjacency), γ is the bandwidth, settled to reach zero adjacency in correspondence of maximum value of cosine similarity (one) and for the values lower than the second quartile of the distribution.

This transformation aligns with a substantial body of empirical evidence indicating that knowledge exchange often follows a non-linear relationship with proximity. Specifically, the benefits of proximity are maximized at intermediate levels: when firms are close enough to enable mutual understanding and interaction, yet not so similar as to restrict diversity of perspectives or lead to redundancy (Kok et al., 2020; Marra et al., 2019, 2024). At low levels of proximity, cognitive and communicative gaps may impede knowledge absorption; at very high levels, excessive similarity can reduce opportunities for novel recombination and learning. The inverted-U transformation reflects this dynamic, capturing how moderate proximity tends to foster the most favorable conditions for innovation and performance improvement.

4.2. The combination of the geographical and semantic proximity matrices

The reliance on purely geographical proximity matrices is a widespread practice in spatial econometric modeling (Hazir and Autant-Bernard, 2014; Ter-Wal, 2013). Corrado and Fingleton (2012) emphasize that while such matrices have the significant advantage of being exogenous, researchers should always strive to incorporate more complexity into the spatial framework. This could include elements such as knowledge flows, social interactions, trade in goods and services, and other factors that enrich the understanding of proximity and its impact on economic and innovation performance.

In recent studies, researchers have developed various approaches to enhance spatial econometric models by building hybrid spatial weight matrices (W), combining multiple dimensions of proximity (Harris et al., 2011). Parent and LeSage (2008) argue that combining geographical proximity with other types of proximity can provide deeper insights into knowledge spillovers than using geographical proximity alone. More specifically, Autant-Bernard (2012) suggests that adding a semantic proximity matrix to the geographical matrix can shed light on the mechanisms through which knowledge flows occur.

One approach to incorporate multiple proximity dimensions is through 'higher-order' models, where multiple spatial lags of the dependent variable are used, each relying on a different W matrix (Lacombe, 2004; Li and Liu, 2010). However, these models can encounter estimation challenges due to the interaction of spatial parameters associated with different structures. This interplay complicates both accurate estimation and the interpretation of results (LeSage and Pace, 2011; Elhorst et al., 2011).

In this study, interpretability is considered a key criterion in selecting the most appropriate modeling approach. Accordingly, we adopt the convex combination framework, which allows for the integration of multiple proximity structures, each associated with its own parameter that reflects its relative contribution (Debrasy and LeSage, 2021).

The methodology follows Debarsy and LeSage (2021; 2022). They propose, drawing on Pace and LeSage (2010) and Hazir et al. (2018), a model that combines different W matrices through a convex combination. This approach avoids many of the complications found in higher-order models. The scalar weights assigned to each matrix must be positive and sum to one. In our case, both the geographical and semantic proximity matrices are treated as exogenous and row normalized. Since a convex combination of row normalized matrices remains row normalized, standard spatial model specifications and estimation methods can be applied, with the spatial parameters bounded between $1/\lambda_{min}$ and 1, where λ_{min} is the smallest eigenvalue of the matrix W (Debarsy and LeSage, 2018; 2022). In such a way, the parameters associated with each matrix in the convex combination indicate the relative importance of each proximity matrix (Debarsy and LeSage, 2021).

Accordingly, the combined W matrix is obtained as:

$$W = \varphi_1 B_1 + \varphi_2 B_2 \qquad [2]$$

$$with \sum_{i=1}^{2} \varphi_i = 1 \rightarrow \ \varphi_2 = (1 - \varphi_1)$$

where B_1 is the geographical proximity matrix and B_2 is the semantic proximity one.

The resulting matrix is then employed within spatial econometric specifications to model interactions between neighbouring units. The intensity of interaction between any two generic units, *i* and *j*, is represented by the corresponding element of the spatial weight matrix W, denoted as w_{ij}. As defined in Equation [2], this value is computed as follows:

$$w_{ij} = \varphi_1 * b1_{ij} + \varphi_2 * b2_{ij}$$

where $b1$ and $b2$ are the spatial structures corresponding to different proximity matrices (e.g., geographical and semantic), and φ_1 and φ_2 are their associated weights indicating relative importance.

To enhance understanding of the construction process for matrix W, we provide additional details, also through a numerical example, in Appendix.

To ensure a proper transition between the physical proximity and the combined one, in which the semantic structure is inserted, in our application we follow a two-step approach. First, we confirm that the geographical distribution of firms is significant according to a Moran's test on the OLS residuals (King, 1981). Once this was established, we proceeded to combine geographical and semantic proximities in a single spatial weight matrix (W), using a convex combination of the two.

While the combined use of geographical and semantic proximity matrices offers a richer understanding of inter-firm spillovers, it is important to acknowledge a few methodological limitations. First, the convex combination of the two proximity matrices introduces a level of complexity in interpretation. Although this method allows us to capture hybrid forms of proximity, the relative weights assigned to each matrix ($\varphi_1$and φ_2) must be interpreted with caution, as they may be influenced by the underlying structure of the data. Second, as with any spatial model, parameter sensitivity may be a concern when combining multiple adjacency structures. Although we perform robustness checks and validate our model against alternative specifications, these issues remain areas for further exploration in future research.

4.3. Model specification

To choose the best model specification, it is standard practice to start with an Ordinary Least Squares (OLS) regression and then assess whether spatial interaction effects need to be incorporated. In recent spatial econometrics, the Spatial Durbin Model (SDM) is widely recommended due to its ability to account for potential biases caused by omitted variables (Elhorst, 2010; LeSage, 2014; LeSage and Pace, 2009). The SDM effectively captures both endogenous spatial effects (represented as WY) and exogenous spatial effects (represented as WX), which allow for the identification of global and local knowledge spillovers, respectively.

Following the methodologies of Debarsy and LeSage (2018) and Hazir et al. (2018), we adopt the SDM as a starting point. This model provides a robust framework for analyzing the interplay of spatial factors, and its flexibility in handling multiple proximity dimensions makes it suitable for investigating knowledge spillovers between firms.

In matrix notation, we consider the SDM as:

$$y= \rho Wy+\beta X+\theta WX+\varepsilon \qquad [3]$$
$$with \;\; W = \varphi_1 B_1 + \varphi_2 B_2$$

where y is the vector of sales growth rates (Δ_Sales) for the each firm between 2022 and 2023 (Lu et al., 2021); β is the parameters' vector related to each of the covariates; ρ is the autocorrelation parameter for the dependent variable, indicating the magnitude of the mutual influence between neighbours, while the vector of parameters θ measures the influence of the covariates over the neighbours' dependent variable; ε is the error term. The matrix X includes 616 units and 3 variables, namely: initial turnover ($Sales\_t0$), used as a proxy to identify scaleups (Lindelöf and Löfsten, 2004), average number of employees ($Emplo$), calculated as the average number of employees in the observed period and serving as a proxy for the firm's knowledge base, with each employee contributing distinct expertise (Balsmeier et al., 2014; Tubiana et al., 2022), and growth stage ($Startup$) to distinguish between early-stage and more mature firms (Rydehell et al., 2019; Guerrero et al., 2023).

The correlation between these variables reaches its maximum for the couple $Sales\_t0$ and $Emplo$, namely 0.38, circumstance that allows to exclude possible problems of multicollinearity.

The SDM nests other spatial models like the Spatial Lag Model (SLM) and the Spatial Lag of X (SLX). Notably, the SLX model focuses on exogenous spatial interaction effects (WX), which are particularly relevant to our study. There is a strong case for considering the SLX model, as Gibbons and Overman (2012) argue that the reduced form of the SDM can hardly be distinguished from a model that only includes first-order exogenous interaction effects, like the SLX model. Moreover, Elhorst (2017) advocates for models that prioritize exogenous interaction effects over endogenous ones, while Corrado and Fingleton (2012) suggest using exogenous effects because endogenous interactions (WY) may obscure the true impact of omitted spatially dependent variables, potentially driving to misleading interpretations.

In this work, we employ the SLX model as:

$$y= \beta X+\theta WX+\varepsilon \qquad [4]$$
$$with \;\; W = \varphi_1 B_1 + \varphi_2 B_2$$

In terms of interpretation and computational efficiency, Halleck Vega and Elhorst (2015) argue that the SLX model offers a significant advantage due to its simplicity. Unlike the SDM, which requires the computation of both direct and indirect impacts derived from the partial derivatives matrix of the expected value of y concerning each explanatory variable (LeSage and Pace, 2009), the SLX model allows for immediate interpretation of its coefficients. The direct effects (β) and indirect effects (θ) can be directly understood without the need for further manipulation (Elhorst, 2014).

To determine the most appropriate model, Elhorst (2014) suggests a top-down approach, where the most general model, such as the SDM, is first estimated using maximum likelihood. From there, a likelihood ratio test (LR-test) can be conducted to assess whether a simpler nested model, such as the SLX, provides an adequate fit, potentially reducing model complexity without sacrificing explanatory power.

The LR-test takes the following form:

$$-2(LogL_{res} - LogL_{unres}) \qquad [5]$$

where $LogL_{res}$ represents the log-likelihood of the nested (restricted, in this case the SLX) model, and $LogL_{unres}$ represents the log-likelihood of the most general (unrestricted, here the SDM) model. This statistic follows a Chi-squared distribution, with the degrees of freedom corresponding to the number of restrictions applied. The null hypothesis of the test posits that the most general and complex model does not outperform the simpler and restricted one, which is than advisable.

To estimate the model parameters, we follow the approach outlined by Hazir et al. (2018). This process involves two steps. In the first step, we assess the likelihood function over a grid of values for the convex combination parameters φ_1 and φ_2. The optimal combination, which corresponds to the highest likelihood, is then used in the second step to estimate the remaining model parameters. Debarsy and LeSage (2018) propose using Bayesian methods for parameter estimation, suggesting that the Hazir et al. (2018) approach can lead to biases in the scalar summary measures of impacts developed by LeSage and Pace (2009). However, our case involves a combination of only two parameters, and the SLX model does not require calculating these impacts. Therefore, we opted for the Hazir et al. (2018) procedure due to its simplicity, as it allows us to employ conventional maximum likelihood estimation methods (Debarsy and LeSage, 2022).

Once the optimal convex combination is established according to the method described, we conduct an LR-test between the optimal SDM and the optimal SLX model to determine which is most suitable. Additionally, we aim to contribute to the literature on convex combinations by proposing the use of the LR-test to differentiate between models using a pure geographical proximity matrix and those with a combined adjacency matrix. It is important to note that LR-tests can only be applied when comparing nested models. Therefore, they cannot formally be used to test models with different weights matrices (Elhorst, 2010). However, with a convex combination matrix, the formula [4] can be rewritten as follows:

$$y= \beta X + \theta(\varphi_1 B_1 + \varphi_2 B_2)X + \varepsilon \qquad [6]$$

where the case of pure geographical proximity is a restricted case in which $\varphi_2 = 0$.

Therefore, the geographical and the combined models can be considered as nested, and a LR-test can be performed with one degree of freedom.

5. Discussion

The following subsections first present the results of the models under different specifications, and then provide a series of robustness checks conducted to reinforce the main findings.

5.1. Results

To clarify the role of geographical proximity, we first estimate an OLS regression using the variables introduced in the previous section. We then combine the geographical structure with the semantic dimension.

The importance of the physical structure is supported by a Moran's I test on the residuals (King, 1981), which indicates the presence of spatial autocorrelation.

The null hypothesis of no spatial autocorrelation is rejected at a 5% confidence level, with a p-value of 0.009. The tested geographical adjacency matrix is based on the nearest-neighbour method, with 40 neighbours per unit, as this matrix showed the highest significance according to Moran's *I* statistic and a density comparable to that proposed by Hazir et al. (2018).

Having established that the geographical distribution of firms is significant, we proceed with the estimation process for the convex combination parameters to construct the W matrix in which both proximities act together. Table 5 presents the grid of values of the weights assigned to geographical proximity (φ_1) and semantic proximity (φ_2). For each pair of weights, we estimate the model and calculate the log-likelihood. The optimal combination, based on the highest log-likelihood, is highlighted in bold. As a check, the Newton's algorithm performed in R confirms that there is no other maximum in the log-likelihood function.

Table 5. Loglikelihood of SLX and SDM for each convex combination varying φ_1 and φ_2.

φ_1 (geographical)	φ_2 (semantic)	Loglikelihood SLX	Loglikelihood SDM
1	0	625.2261	625.2265
0.9	0.1	626.5937	626.5974
0.8	0.2	627.4345	627.4368
0.79	0.21	627.4806	627.4825
0.78	0.22	627.5191	627.5206
0.77	0.23	627.5498	627.5509
0.76	0.24	627.5726	627.5733
0.75	0.25	627.5876	627.5879
0.74	**0.26**	**627.5946**	**627.5948**
0.73	0.27	627.5939	627.5939
0.72	0.28	627.5855	627.5855
0.71	0.29	627.5696	627.5697
0.7	0.3	627.5464	627.5468
0.69	0.31	627.5162	627.5171
0.68	0.32	627.4792	627.4808
0.67	0.33	627.4359	627.4383
0.66	0.34	627.3865	627.3900
0.6	0.4	626.9855	626.9995
0.5	0.5	626.0924	626.1331
0.4	0.6	625.1670	625.2279
0.3	0.7	624.3462	624.4104
0.2	0.8	623.5524	623.7166
0.1	0.9	623.1059	623.1448
0	1	622.6553	622.6796

For both the SDM and SLX models, the optimal convex combination remains the same, with φ_1 equal to 0.74 and, consequently, φ_2 equal to 0.26. In the procedure implemented, following the approach of Hazir et al. (2018), we define a grid of values, reducing φ_1 by 0.1 each time. Subsequently, we refine the grid using 0.01 intervals around the values that yield the highest likelihood (Figure 2).

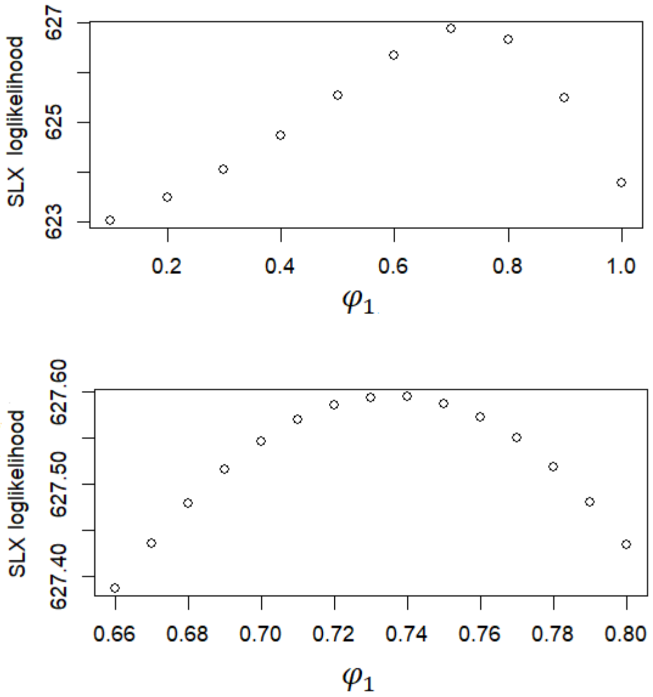

Figure 2: Loglikelihood of SLX combined model varying φ_1.

To determine the most appropriate spatial specification between the SDM and SLX models, we conduct an LR-test on the two models estimated via maximum likelihood using the optimal combination of φ_1 and φ_2. Since the difference between the loglikelihoods of SDM and SLX at the optimal combination is minimal (0.0002), with a test value is 0.0004, this means that the null hypothesis cannot be rejected at any confidence level. Since the SDM does not significantly outperform the SLX, the latter is preferable due to its simplicity and ease of interpretation. For completeness, the SLX model outperforms OLS, as shown by the LR-test, which takes a value of 17.208 with 3 degrees of freedom, exceeding the 5% critical value of 7.81.

As outlined in Section 4, we also aim to distinguish between the model with pure geographical proximity and the model using the combined W matrix, via the LR-test. To do this, we assess whether the SLX model estimated with the optimal combination of φ_1 and φ_2 outperforms the SLX model restricted to $\varphi_2 = 0$. The result yields a test statistic of 4.74, exceeding the 5% critical value of 3.84 for the Chi-squared distribution with 1 degree of freedom. Thus, the null hypothesis is rejected, indicating that the model with the combined W matrix outperforms the one based solely on geographical adjacency. This result confirms the importance of non-geographical proximity, which accounts for more than a quarter of the combined effect, while geographical proximity remains predominant, contributing over 70% of the weight.

Table 6. Estimation results of OLS, SLX_geo ($\varphi_2 = 0$), and SLX_combined.

Variables	OLS	SLX_geo	SLX_combined
$intercept$	0.2871***	-0.4874*	-0.6424**
$Sales\_t0$	-0.0374***	-0.0366***	-0.0373***
$Emplo$	0.000006*	0.000006*	0.000006*
$Startup$	-0.0535***	-0.05246***	-0.0512***
Lag_Sales_t0		0.1076***	0.1298***
Lag_Emplo		-0.00005	-0.00002
$Lag\_Startup$		0.1608**	0.1871***
R-squared	0.1067	0.1246	0.1313
Loglikelihood	618.9906	625.2261	627.5946

Note: *** p<0.005; **p<0.01; *p<0.05.

The model estimations yield several noteworthy findings. In the combined specification, the coefficients show the expected signs and statistical significance, reinforcing previous insights into the moderating effects of initial firm size and growth stage on firm performance (Rydehell et al., 2019; Guerrero et al., 2023). Additionally, the number of employees exhibits a positive effect, consistent with the idea that workforce size reflects a firm's knowledge base and operational capacity (Balsmeier et al., 2014).

Although spatial lag coefficients primarily capture spatial autocorrelation, which may indicate, but does not conclusively prove, the presence of spillover effects, we emphasize, in what follows, the economic interpretation that spatial dependence suggests potential spillovers. We are aware that, from a technical standpoint, this is a somewhat overstated interpretation, as spatial dependence alone does not constitute definitive evidence of spillovers.

Interestingly, while being a startup is associated with a negative direct effect on performance, we observe positive spillover effects, supporting the idea that startups can generate valuable externalities that enhance neighbouring firms' growth (Lindelöf and Löfsten, 2004; Kaneva et al., 2023; Marra et al., 2024). This finding highlights the importance of considering startups not only as individual actors but also as contributors to the broader knowledge environment. Moreover, the positive spillover associated with firms' initial size supports the knowledge equilibrium argument (Grillitsch and Nilsson, 2019), suggesting that larger firms may help balance performance across the ecosystem over time through knowledge diffusion.

Given the distinct structure of Sweden's tech ecosystem, marked by both densely populated urban hubs such as Stockholm, Gothenburg, and Malmö, and more geographically dispersed innovation hotspots, these findings carry important policy implications. First, they underline the need for spatially differentiated innovation policies. In high-density areas, policies might focus on maximizing the externalities produced by startups by enhancing collaboration spaces, funding early-stage ventures, and supporting mentorship networks. In more peripheral areas, spillovers from larger firms could be leveraged through incentives for co-location, shared R&D infrastructure, and digital platforms to connect firms operating at cognitive, industrial or technological proximity.

Our framework, which combines geographical and semantic proximity, allows policymakers to identify potential spillover pathways that extend beyond physical closeness. This adaptability makes it particularly well-suited for informing targeted and place-sensitive innovation strategies. By incorporating both dimensions of proximity, our approach offers a preliminary understanding of how firm characteristics shape ecosystem dynamics, enabling evidence-based policies tailored to the specific spatial and structural features of Sweden's technology sector.

5.2. Robustness checks

In empirical contexts such as the one presented here, where the estimated coefficients are consistent in sign and magnitude across both OLS and geographic SLX specifications, and align with theoretical expectations, confidence in the validity of the results is generally well-founded (Lu and White, 2014).

Further validation of model reliability is performed through a robustness analysis aimed at assessing the stability of the core coefficients. To ensure their robustness, we check their sensitivity to different model specifications and to the inclusion of additional explanatory variables.

Table 7 presents the estimation results for three different spatial econometric models: the SDM (Model 1), the SLM (Model 2), and the SLX (Model 3). In Model 3, we include two additional control variables: the firm's age (Age), as recommended by several studies such as, for example, Coad et al. (2017), and a dummy variable indicating whether the firm operates in a business-to-business ($B2B$) context, given that firms engaging with other businesses are more likely to foster reciprocal exchanges of knowledge, ideas, and opportunities (Cappelli and Cucculelli, 2024).

Table 7. Robustness check. Direct and indirect impacts of model 1 (combined SDM), model 2 (combined SLM), and model 3 (combined SLX with 2 added variables).

	Model 1	**Model 2**	**Model 3**
Direct Impacts			
$Sales\_t0$	-0.0376***	-0.0313***	-0.0386***
$Emplo$	0.000006**	0.000006**	0.000004*
$Startup$	-0.00512***	-0.00521***	-0.00505***
Age			0.00043
$B2B$			0.00064
Indirect impacts			
$Sales\_t0$	0.122***	0.121***	0.123**
$Emplo$	-0.000015	-0.000011	-0.00001
$Startup$	0.170**	0.167**	0.173***
Age			-0.00007
$B2B$			-0.00192

Note: *** p<0.005; ** p<0.01; *p<0.05.

Across all models, the estimated coefficients maintain the same sign and exhibit very similar magnitudes, with no substantial differences observed. This consistency supports the robustness of our findings.

An additional issue that is often overlooked in spatial econometric analysis is endogeneity. As noted by Halleck-Vega and Elhorst (2015), a notable advantage of the SLX model is that it permits the use of conventional non-spatial econometric techniques, such as the Wu-Hausman test, for detecting endogeneity. We conduct the Wu-Hausman test, where the null hypothesis states that the regressors are exogenous.

Following standard recommendations in the spatial econometrics literature (Kelejian and Prucha, 1998; Baltagi et al., 2014), we use as instruments the first-order spatial lag of the regressors (WX, internal instrument), along with their second and third spatial lags ($W^2X$and W^3X, respectively). The test yields an F-statistic of 1.229 with 3 and 606 degrees of freedom, corresponding to a p-value of 0.298. This result indicates that the null hypothesis of exogeneity cannot be rejected, thereby supporting the validity of our model and excluding the presence of endogenous regressors.

The detection of spatial patterns in OLS residuals may indicate the presence of spatial heterogeneity in addition to spatial dependence. Spatial tests are known to capture both effects, making it essential to distinguish between them. According to Anselin (1988), heteroskedasticity in the residuals of spatial models may often arise from unobserved spatial heterogeneity: namely, variations in the data-generating process across spatial units that lead to non-constant coefficients.

To address this potential issue, we apply a scan test for spatial groupwise heteroscedasticity (SGWH), following the approach proposed by Chasco et al. (2018), which is based on the spatial scan methodology developed by Kulldorff et al. (2009). The null hypothesis of this test is that the residuals from the spatial model are independently and identically distributed and follow a normal distribution, while the alternative hypothesis allows for heterogeneity between regional clusters. The test is implemented using the SpatialScan function in R (Frévent et al., 2022). The result indicates that the null hypothesis cannot be rejected at the 1% significance level, thus ruling out significant spatial heterogeneity in the residuals and reinforcing the robustness of the estimated results.

Although heterogeneity is not statistically significant, due to the distinct geographical concentration of Sweden's tech firms, we conducted an additional check. The SLX model was estimated first for Stockholm and then for a broader subset including Gothenburg and Malmö. The comparison of these models with the overall SLX model based on geographical proximity yields comparable results (Table 8).

Table 8. Estimation results of SLX geographical models estimated for Sweden (model A), Stockholm, Gothenburg and Malmö (model B), and Stockholm (model C) subsets.

Variables	model A	model B	model C
$intercept$	-0.4874*	-0.4296*	-0.2964**
$Sales\_t0$	-0.0366***	-0.0432***	-0.0583***
$Emplo$	0.000006*	0.000003*	0.000011*
$Startup$	-0.05246***	-0.06670***	-0.0801***
Lag_Sales_t0	0.1076***	0.1069***	0.1099***
Lag_Emplo	-0.00005	-0.00003	-0.00011
$Lag\_Startup$	0.1608**	0.1533**	0.0917*

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

6. Conclusions

Our purpose in this paper was to combine a geographical proximity matrix with a semantic matrix to explore knowledge spillovers between firms. The semantic proximity matrix was built using web-derived data, capturing firms' expertise related to industrial specializations and technologies.

We tested that companies generate knowledge spillovers that positively affect the performance of neighbouring firms. Our findings showed that a firm's economic performance is shaped not only by its intrinsic characteristics, but more notably by the spillover effects that arise from neighbouring units in both geographical and semantic proximity. These effects were most pronounced when both forms of proximity were combined optimally.

The use of web-derived textual data has allowed us to gather information along two key dimensions: what firms do (that is, their industrial specializations) and the technologies they employ, along with the underlying expertise embedded in their workforce. These dimensions are not independent but deeply interconnected. A firm's industrial focus influences the technologies it adopts, and vice versa, with both dimensions shaped by the firm's internal knowledge base and expertise. This intersection of industrial, technological, and cognitive dimensions provides a richer, more integrated perspective, enhancing our understanding of how knowledge flows emerge and how different forms of proximity support innovation.

From a policymaking standpoint, these dimensions are vital when developing strategies to drive economic growth (McCann and Ortega-Argilés, 2016). Understanding how these proximities represent firms' interactions allows policymakers to design more effective policies: potential interventions include promoting R&D in related sectors, encouraging the formation of innovation clusters, and strengthening regional industrial and technological infrastructures (Montresor et al., 2023; Losurdo et al., 2019). As shown throughout the paper, the wide range of applications that leverage textual data to create firm profiles and novel proximity measures demonstrates the strength of these emerging methods. Their value lies not only in the richness and timeliness of the data they elaborate, but also in the flexibility they offer, allowing researchers and policymakers to tailor insights to specific purposes. This adaptability makes such approaches particularly powerful for informing targeted innovation strategies and evidence-based policy design.

However, this study is not without limitations.

Firstly, the dataset used, while valuable, restricted the number of variables that could be applied in the statistical model due to limited data on companies, particularly because many are startups in the early or seed stages. This limitation influenced our variable selection and model configuration. However, it also strengthens the contribution of the analysis, as it sheds light on the innovative startup phenomenon, which is typically difficult to capture due to the scarcity of reliable data in this sector (Giuliani et al., 2024). Thus, despite these constraints, the study provides insights into a segment that is often underexplored.

Secondly, we are reconsidering the adequacy of cosine similarity for estimating non-geographical proximity and are exploring alternative text-based methods. As seen, cosine similarity necessitated some preliminary technical steps to smooth out the strict co-occurrence of keywords: by employing topic modeling, which groups words into broader themes and generated new keywords, and semantic enrichment, which incorporates contextual understanding, we only partially mitigated the limitations of the adopted technique. The result was a less sophisticated measure of semantic proximity. We look for a more refined technique to fully capture the nuances of semantic similarity, which may 'understand' the relationships between concepts, beyond mere keyword matching (Lara-Clares et al., 2021). This would allow to propose a more streamlined methodology, reducing the need for multiple steps and minimizing approximation errors.

Thirdly, the choice to apply our analysis to firms across an entire country rather than a more localized area, such as a region or metropolitan area, may have somewhat weakened our spillover effects. In a more localized context, not only physical distances are significant (Bereitschaft, 2019), but semantic distances as well (Fritz and Manduca, 2021). Sweden has a diversified economy, which reduces the companies' common base of expertise on industrial specializations and adopted technologies necessary to foster knowledge exchange and collaborative interactions (Grillitsch and Nilsson, 2019).

Fourthly, a key limitation of our analysis is the difficulty in disentangling overlapping knowledge spillovers from startups and larger, more established firms. While our model identifies statistically significant spillover patterns, these mechanisms likely coexist and interact. Startups may contribute disruptive, experimental knowledge, while incumbents provide codified knowledge and support incremental innovation. This complexity cautions against attributing performance effects to a single type of actor. For policymakers, this underscores the need for balanced strategies that support both entrepreneurial dynamism and organizational stability. Likewise, firms should recognize that proximity-related knowledge benefits often arise from multiple, interwoven sources. More specifically, this implies that firms should strategically assess not only the presence of startups or incumbents in their vicinity, but also the type of knowledge interactions these actors facilitate. Future research could address this by using more detailed firm classifications, network data, or longitudinal designs to better capture the evolving nature of spillovers across time and regions.

All four of the above-mentioned limitations represent valuable avenues for future research. We plan to explore these directions further in order to contribute to the ongoing debate with additional evidence.

7. References

Aarstad, J., Kvitastein, O. A., & Jakobsen, S. E. (2016). Related and unrelated variety as regional drivers of enterprise productivity and innovation: A multilevel study. Research Policy, 45(4), 844–856.

Abbasiharofteh, M., Krüger, M., Kinne, J., Lenz, D., Resch, B., 2023. The digital layer: alternative data for regional and innovation studies. Spat. Econ. Anal. 1–23. https://doi.org/10.1080/17421772.2023.2193222.

Aldieri, L. (2013), "Knowledge technological proximity: Evidence from US and european patents", Economics of Innovation and New Technology, 22(8), 807-819.

Amoroso, S., Diodato, D., Hall, B. H., & Moncada-Paternò-Castello, P. (2023). Technological relatedness and industrial transformation:: Introduction to the Special Issue. Journal of Technology Transfer, 48(2), 469–475. https://doi.org/10.1007/s10961-022-09941-1

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. Geographical analysis, 20(1), 1-17.

Audretsch, D.B., Feldman, M.P.: R&D spillovers and the geography of innovation and production. Am. Econ.Rev. 3, 630–640 (1996)

Autant-Bernard C. (2012) Spatial Econometrics of Innovation: Recent Contributions and Research Perspectives, Spatial Economic Analysis, 7:4, 403-419, DOI: https://doi.org/10.1080/17421772.2012.722665

Autant-Bernard, C. and LeSage, J.P. (2011), Quantifying Knowledge Spillovers Using Spatial Econometric Models. Journal of Regional Science, 51: 471-496. https://doi.org/10.1111/j.1467-9787.2010.00705.x

Balsmeier, B., Buchwald, A., & Stiebale, J. (2014). Outside directors on the board and innovative firm performance. Research Policy, 43(10), 1800–1815. https://doi.org/10.1016/j.respol.2014.06.003

Baltagi, B. H., Fingleton, B., & Pirotte, A. (2014). Estimating and forecasting with a dynamic spatial panel data model. Oxford Bulletin of Economics and Statistics, 76(1), 112-138.

Bereitschaft, B. (2019). Are walkable places tech incubators? Evidence from Nebraska's 'Silicon Prairie.' *Regional Studies, Regional Science*, *6*(1), 339–356. https://doi.org/10.1080/21681376.2019.1620631

Boschma R. (2005), "Proximity and Innovation: A Critical Assessment". Reg Stud [Internet];39(1):61–74. Available from: https://doi.org/10.1080/0034340052000320887

Boschma, R., Eriksson, R., Lindgren, U. (2009), "How does labour mobility affect the performance of plants? The importance of relatedness and geographical proximity", Journal of Economic Geography, 9(2): 169–190.

Breschi, S., Lissoni, F.: Knowledge spillovers and local innovation systems: a critical survey. Ind. Corp.Change 4, 975–1005 (2001)

Cao, Z., Derudder, B., & Peng, Z. (2019). Interaction between different forms of proximity in inter-organizational scientific collaboration: The case of medical sciences research network in the Yangtze River Delta region. Papers in Regional Science, 98(5), 1903–1924. https://doi.org/10.1111/pirs.12438

Cappelli, R., & Cucculelli, M. (2024). The role of business models in explaining differences in firm performance. In Unpacking Innovation (pp. 92-101). Edward Elgar Publishing.

Chasco, C., Le Gallo, J., & López, F. A. (2018). A scan test for spatial groupwise heteroscedasticity in cross-sectional models with an application on houses prices in Madrid. Regional Science and Urban Economics, 68, 226-238.

Cicerone, G., Losacker, S., & Ortega-Argilés, R. (2024). Regional diversification into green and digital economic activities – The case of UK Local Authorities. ESCoE Discussion Paper No. 2024-07. http://escoe-website.s3.amazonaws.com/wp-content/uploads/2024/07/29090432/ESCoE-DP-2024-07.pdf

Coad, A., Holm, J., Krafft, J., & Quatraro, F. (2017). Firm age and performance. Journal of Evolutionary Economics, 28, 1 - 11. https://doi.org/10.1007/s00191-017-0532-6.

Cohen, W.M., Levinthal, D.A., 1990. Absorptive capacity: a new perspective on learning and innovation. Adm. Sci. Q. 35 (1), 128. https://doi.org/10.2307/2393553

Colombelli, A., & Quatraro, F. (2019). Green start-ups and local knowledge spillovers from clean and dirty technologies. Small Business Economics, 52(4), 773–792. https://doi.org/10.1007/s11187-017-9934-y

Content, J., Cortinovis, N., Frenken, K., & Jordaan, J. (2022). The roles of KIBS and R&D in the industrial diversification of regions. Annals of Regional Science, 68(1), 29–64. https://doi.org/10.1007/s00168-021-01068-9

Corrado, L., & Fingleton, B. (2012). Where is the economics in spatial econometrics?. Journal of Regional Science, 52(2), 210-239.

Cortinovis, N., Crescenzi, R., & van Oort, F. (2020). Multinational enterprises, industrial relatedness and employment in European regions. Journal of Economic Geography, 20(5), 1165–1205. https://doi.org/10.1093/jeg/lbaa010

Cortinovis, N., Van Oort, F. (2015), "Variety, economic growth and knowledge-intensity of European regions: A spatial panel analysis", Regional Studies, 41(5), 685–697.

Davids, M., & Frenken, K. (2018). Proximity, knowledge base and the innovation process: towards an integrated framework. Regional Studies, 52(1), 23–34. https://doi.org/10.1080/00343404.2017.1287349

Dealroom (2024). Data on the Sweden's national tech ecosystem. Date of the last extraction: July 21st, 2024. Available from: https://dealroom.co/reports/sweden-tech-2023-review

Debarsy, N., & LeSage, J. (2018). Flexible dependence modeling using convex combinations of different types of connectivity structures. Regional Science and Urban Economics, 69, 48-68.

Debarsy, N., & Lesage, J. P. (2021). Using convex combinations of spatial weights in spatial autoregressive models. In Handbook of Regional Science (pp. 2267-2282). Berlin, Heidelberg: Springer Berlin Heidelberg.

Debarsy, N., LeSage, J. P. (2022). Bayesian Model Averaging for Spatial Autoregressive Models Based on Convex Combinations of Different Types of Connectivity Matrices. Journal of Business & Economic Statistics, 40(2), 547–558. Available from: https://doi.org/10.1080/07350015.2020.1840993.

Döring, T., Schnellenbach, J. (2006). What do we know about geographical knowledge spillovers and regional growth? A survey of the literature. Reg. Stud. 3, 375–395

Ebert, T., Brenner, T., & Brixy, U. (2019). New firm survival: the interdependence between regional externalities and innovativeness. Small Business Economics, 53(1), 287–309. https://doi.org/10.1007/s11187-018-0026-4

Elhorst, J. P. (2010). Applied spatial econometrics: raising the bar. Spatial economic analysis, 5(1), 9-28.

Elhorst, J. P. (2014). Spatial econometrics: from cross-sectional data to spatial panels (Vol. 479, p. 480). Heidelberg: Springer.

Elhorst, J. P. (2017). Spatial Panel Data Analysis. Encyclopedia of GIS, 2, 2050-2058.

Elhorst, J. P., Lacombe, D. J., & Piras, G. (2012). On model specification and parameter space definitions in higher order spatial econometric models. Regional Science and Urban Economics, 42(1-2), 211-220.

Freitas, E., Britto, G., & Amaral, P. (2024). Related industries, economic complexity, and regional diversification: An application for Brazilian microregions. Papers in Regional Science, 103(1). https://doi.org/10.1016/j.pirs.2024.100011

Frenken, K., Van Oort, F.G., Verburg, T. (2007), "Related variety, unrelated variety and regional economic growth", Regional Studies, 41(5): 685-697.

Frévent, C., Ahmed, M. S., Soula, J., Smida, Z., Cucala, L., Dabo-Niang, S., & Genin, M. (2022). The R Package HDSpatialScan for the Detection of Clusters of Multivariate and Functional Data using Spatial Scan Statistics. R Journal, 14(3).

Fritsch, M., & Kublina, S. (2018). Related variety, unrelated variety and regional growth: the role of absorptive capacity and entrepreneurship. Regional Studies, 52(10), 1360–1371. https://doi.org/10.1080/00343404.2017.1388914

Fritz, B. S. L., & Manduca, R. A. (2021). The economic complexity of US metropolitan areas. Regional Studies, 55(7), 1299–1310. https://doi.org/10.1080/00343404.2021.1884215

Gibbons, S., & Overman, H. G. (2012). Mostly pointless spatial econometrics?, Journal of regional Science, 52(2), 172-191.

Giuliani, D., Toffoli, D., Dickson, M. M., Mazzitelli, A., & Espa, G. (2024). Assessing the role of spatial externalities in the survival of Italian innovative startups. Regional Science Policy and Practice, 16(1). https://doi.org/10.1111/rsp3.12653

Golra, O. A., Rosiello, A., & Harrison, R. T. (2024). Proximity and its impact on the formation of product and process innovation networks among producer firms. Regional Studies, 58(4), 768–786. https://doi.org/10.1080/00343404.2023.2249029

Grillitsch, M., & Nilsson, M. (2019). Knowledge externalities and firm heterogeneity: Effects on high and low growth firms. Papers in Regional Science, 98(1), 93–114. https://doi.org/10.1111/pirs.12342

Guerrero, A. J., Heijs, J., & Huergo, E. (2023). The effect of technological relatedness on firm sales evolution through external knowledge sourcing. Journal of Technology Transfer, 48(2), 476–514. https://doi.org/10.1007/s10961-022-09931-3

Halleck Vega, S., & Elhorst, J. P. (2015). The SLX model. Journal of Regional Science, 55(3), 339-363.

Harris, R., Moffat, J., & Kravtsova, V. (2011). In search of 'W'. Spatial Economic Analysis, 6(3), 249-270.

Hazir, C. S., & Autant-Bernard, C. (2014). Determinants of cross-regional R&D collaboration: some empirical evidence from Europe in biotechnology. The Annals of Regional Science, 53, 369-393.

Hazır, C. S., LeSage, J., & Autant-Bernard, C. (2018). The role of R&D collaboration networks on regional knowledge creation: Evidence from information and communication technologies. Papers in Regional Science, 97(3), 549-568.

Jespersen, K., Rigamonti, D., Jensen, M. B., & Bysted, R. (2018). Analysis of SMEs partner proximity preferences for process innovation. Small Business Economics, 51(4), 879–904. https://doi.org/10.1007/s11187-017-9969-0

Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. The journal of real estate finance and economics, 17, 99-121.

Kinne, J., & Axenbeck, J. (2020). Web mining for innovation ecosystem mapping: a framework and a large-scale pilot study. Scientometrics, 125(3), 2011–2041. https://doi.org/10.1007/s11192-020-03726-9

Kinne, J., Lenz, D. (2021), "Predicting innovative firms using web mining and deep learning", PLoS ONE, 16(4): https://doi.org/10.1371/journal.pone.e0249071

Kok, H., Faems, D., & De Faria, P. (2020). Ties that matter: The impact of alliance partner knowledge recombination novelty on knowledge utilization in R&D alliances. Research Policy, 49, 104011. https://doi.org/10.1016/j.respol.2020.104011

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

Kulldorff, M., Huang, L., & Konty, K. (2009). A scan statistic for continuous data based on the normal probability model. International journal of health geographics, 8, 1-9.

Lacombe, D. J. (2004). Does econometric methodology matter? An analysis of public policy using spatial econometric techniques. Geographical analysis, 36(2), 105-118.

Lacombe, D.J., LeSage J.P. 2013. "Using Bayesian Posterior Model Probabilities to Identify Omitted Variables in Spatial Regression Models," Papers in Regional Science, https://doi.org/10.1111/pirs.12070

Lara-Clares, A., Lastra-Díaz, J. J., & Garcia-Serrano, A. (2021). Protocol for a reproducible experimental survey on biomedical sentence similarity. *PLoS ONE*, *16*(3 March). https://doi.org/10.1371/journal.pone.0248663

Lee, L. F., & Liu, X. (2010). Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. Econometric Theory, 26(1), 187-230.

LeSage J. P., & Pace, R. K. (2009). Introduction to Spatial Econometrics. Taylor & Francis.

LeSage, J. P., & Pace, R. K. (2011). Pitfalls in higher order model extensions of basic spatial regression methodology. Review of Regional Studies, 41(1), 13-26.

LeSage, J.P. 2014. What regional scientists need to know about spatial econometrics. The Review of Regional Studies. 44: 13-32.

Li, D., Heimeriks, G., & Alkemade, F. (2021). Recombinant invention in solar photovoltaic technology: can geographical proximity bridge technological distance? Regional Studies, 55(4), 605–616. https://doi.org/10.1080/00343404.2020.1839639

Lindelöf, P., & Löfsten, H. (2004). Proximity as a resource base for competitive advantage: University-industry links for technology transfer. Journal of Technology Transfer, 29(3–4), 311–326. https://doi.org/10.1023/b:jott.0000034125.29979.ae

Liu, J., & Ma, T. (2019). Innovative performance with interactions between technological proximity and geographic proximity: evidence from China electronics patents. Technology

Analysis and Strategic Management, 31(6), 667–679. https://doi.org/10.1080/09537325.2018.1542672

Lopolito, A., Falcone, P. M., & Sica, E. (2022). The role of proximity in sustainability transitions: A technological niche evolution analysis. Research Policy, 51(3). https://doi.org/10.1016/j.respol.2021.104464

Losurdo, F., Marra, A., Cassetta, E., Monarca, U., Dileo, I., & Carlei, V. (2019). Emerging specializations, competences and firms' proximity in digital industries: The case of London. Papers in Regional Science, 98(2), 737–753. https://doi.org/10.1111/pirs.12376

Lu, R., Song, Q., Xia, T., Lv, D., Reve, T., & Jian, Z. (2021). Unpacking the U-shaped relationship between related variety and firm sales: Evidence from Japan. Papers in Regional Science, 100(5), 1136–1157.

Lu, X., & White, H. (2014). Robustness checks and robustness tests in applied economics. Journal of econometrics, 178, 194-206.

Marra A, Baldassari C (2022) Using text data instead of SIC codes to tag innovative firms and classify industrial activities. PLoS ONE 17(6): https://doi.org/10.1371/journal.pone.0270041

Marra A, Carlei V, Baldassari C. Exploring networks of proximity for partner selection, firms' collaboration and knowledge exchange. The case of clean-tech industry. Bus Strat Env. 2020; 29: 1034–1044. https://doi.org/10.1002/bse.2415

Marra, A., Cucculelli, M., & Cartone, A. (2024). So far, yet so close. Using networks of words to measure proximity and spillovers between firms. Eurasian Business Review. https://doi.org/10.1007/s40821-024-00270-x

Martin-Rios, C., Erhardt, N. L., & Manev, I. M. (2022). Interfirm collaboration for knowledge resources interaction among small innovative firms. Journal of Business Research, 153, 206–215. https://doi.org/10.1016/j.jbusres.2022.08.024

McCann, P. (2014). Schools of thought on economic geography, institutions, and development. In M. Fischer & P. Nijkamp (Eds.), Handbook of Regional Science (pp. 527–538). Berlin, Heidelberg: Springer. https://10.1007/978-3-642-23430-9

McCann, P. and Ortega-Argilés, R. (2016) Regional innovation, R&D and knowledge spillovers: the role played by geographical and non-geographical factors. In: R. Shearmu, C. Carrincazeaux, and D. Doloreux (eds), Handbook on the Geographies of Innovation. Northampton, MA: Elgar, pp. 22–44.

Montresor, S., Orsatti, G., & Quatraro, F. (2023). Technological novelty and key enabling technologies: evidence from European regions. Economics of Innovation and New Technology, 32(6), 851–872. https://doi.org/10.1080/10438599.2022.2038147

Nathan, M., Rosso A., (2015), "Mapping digital businesses with big data: some early findings from the UK", Res. Policy, 44, pp. 1714-1733.

Neffke, F., Henning, M. (2008), "Revealed relatedness: Mapping industry space. Papers in Evolutionary Economic Geography, No. 8.19, Utrecht, the Netherlands: Urban and Regional Research Centre, University of Utrecht.

Nilsson, M. (2019). Proximity and the trust formation process. European Planning Studies, 27(5), 841–861. https://doi.org/10.1080/09654313.2019.1575338

Nooteboom B, Van Haverbeke W, Duysters G, Gilsing V, van den Oord A. Optimal cognitive distance and absorptive capacity. Res Policy [Internet]. 2007;36(7):1016–34. Available from: https://www.sciencedirect.com/science/article/pii/S0048733307000807

Pace, R. K., & LeSage, J. P. (2010). Omitted variable biases of OLS and spatial lag models. Progress in spatial analysis: Methods and applications, 17-28.

Panori, A., Kakderi, C., & Dimitriadis, I. (2022). Combining technological relatedness and sectoral specialization for improving prioritization in Smart Specialisation. Regional Studies, 56(9), 1454–1467. https://doi.org/10.1080/00343404.2021.1988552

Papagiannidis, S., See-To, E. W. K., Assimakopoulos, D. G., & Yang, Y. (2017). Identifying industrial clusters with a novel big-data methodology: Are SIC codes (not) fit for purpose in the Internet age?, Computers & Operations Research, Volume 98, 2018. ISSN, 355–366, 0305–0548

Parent, O., & LeSage, J. P. (2008). Using the variance structure of the conditional autoregressive spatial specification to model knowledge spillovers. Journal of applied Econometrics, 23(2), 235-256.

Peng, W., Yu, X., & Ji, Y. (2023). Obtaining advantages from knowledge base: mapping the potential to develop new technologies. Technology Analysis and Strategic Management. https://doi.org/10.1080/09537325.2023.2242519

Petralia, S. (2020). Mapping general purpose technologies with patent data. Research Policy, 49(7). https://doi.org/10.1016/j.respol.2020.104013

Qin, Y., Qin, X., Chen, H., Li, X., & Lang, W. (2021). Measuring cognitive proximity using semantic analysis: A case study of China's ICT industry. Scientometrics, 126(7), 6059–6084. https://doi.org/10.1007/s11192-021-04021-x

Quatraro, F. (2010), "Knowledge coherence, variety and economic growth: Manufacturing evidence from Italian regions", Research Policy, 39(10), 1289–1302.

Ranaei, S., Suominen, A., Porter, A., & Carley, S. (2020). Evaluating technological emergence using text analytics: two case technologies and three approaches. Scientometrics, 122(1), 215–247. https://doi.org/10.1007/s11192-019-03275-w

Russo, M., Caloffi, A., Colovic, A., Pavone, P., Romeo, S., & Rossi, F. (2022). Mapping regional strengths in a key enabling technology: The distribution of Internet of Things competences across European regions. Papers in Regional Science, 101(4), 875–900. https://doi.org/10.1111/pirs.12679

Rydehell, H., Isaksson, A., & Löfsten, H. (2019). Business networks and localization effects for new Swedish technology-based firms' innovation performance. Journal of Technology Transfer, 44(5), 1547–1576. https://doi.org/10.1007/s10961-018-9668-2

Sharma, A., Adhikary, A., Bikash Borah, S., & Pathak, S. (2024). Supply base concentration and firm innovation performance: A contingency study of supply base breadth, depth, dispersion, and collaboration. Journal of Business Research, 174. https://doi.org/10.1016/j.jbusres.2023.114450

Sheng, Y., & LeSage, J. (2021). A spatial regression methodology for exploring the role of regional connectivity in knowledge production: Evidence from Chinese regions. Papers in Regional Science, 100(4), 847–874. https://doi.org/10.1111/pirs.12601

Shkolnykova, M. (2023). Assessing the importance of proximity dimensions for the diffusion of radical innovations in German biotechnology. European Planning Studies, 31(7), 1510–1531. https://doi.org/10.1080/09654313.2022.2147392

Ter Wal, A. L. (2013). Cluster emergence and network evolution: a longitudinal analysis of the inventor network in Sophia-Antipolis. Regional Studies, 47(5), 651-668.

Timmermans, B., Boschma, R. (2014), “The effect of intra- and inter-regional labour mobility on plant performance in Denmark: The significance of related labour inflows”, Journal of Economic Geography 14 (2): 289–311.

Tubiana, M., Miguelez, E., & Moreno, R. (2022). In knowledge we trust: Learning-by-interacting and the productivity of inventors. Research Policy, 51(1). https://doi.org/10.1016/j.respol.2021.104388

Van Oort, F., de Geus, S., Dogaru, T. (2015), “Related variety and regional economic growth in a cross-section of european urban regions”, European Planning Studies, 23(6), 1110-1127.

Whittle, A. (2020). Operationalizing the knowledge space: theory, methods and insights for Smart Specialisation. Regional Studies, Regional Science, 7(1), 27–34. https://doi.org/10.1080/21681376.2019.1703795

Zhang, J., Yan, Y., & Guan, J. (2019). Recombinant distance, network governance and recombinant innovation. Technological Forecasting and Social Change, 143, 260–272. https://doi.org/10.1016/j.techfore.2019.01.022

Zhou, X., Huang, L., Zhang, Y., & Yu, M. (2019b). A hybrid approach to detecting technological recombination based on text mining and patent network analysis. Scientometrics, 121(2), 699–737. https://doi.org/10.1007/s11192-019-03218-5

Zhou, Y., Zhu, S., & He, C. (2019a). Learning from yourself or learning from neighbours: knowledge spillovers, institutional context and firm upgrading. Regional Studies, 53(10), 1397–1409. https://doi.org/10.1080/00343404.2019.1566705

Appendix

The convex combined adjacency matrix *W* is constructed starting from two similarity matrix B_1 and B_2, which are based respectively on geographical and semantic proximities.

In formulas:

$$W = \varphi_1 B_1 + \varphi_2 B_2 = \varphi_1 * \begin{pmatrix} 0 & b1_{12} & \cdots & b1_{1n} \\ b1_{21} & & \ddots & b1_{2n} \\ \vdots & & & \vdots \\ b1_{n1} & b1_{n2} & \cdots & 0 \end{pmatrix} + \varphi_2 * \begin{pmatrix} 0 & b2_{12} & \cdots & b2_{1n} \\ b2_{21} & & \ddots & b2_{2n} \\ \vdots & & & \vdots \\ b2_{n1} & b2_{n2} & \cdots & 0 \end{pmatrix} =$$

$$\begin{pmatrix} 0 & \varphi_1 * b1_{12} & \cdots & \varphi_1 * b1_{1n} \\ \varphi_1 * b1_{21} & & \ddots & \varphi_1 * b1_{2n} \\ \vdots & & & \vdots \\ \varphi_1 * b1_{n1} & \varphi_1 * b1_{n2} & \cdots & 0 \end{pmatrix} + \begin{pmatrix} 0 & \varphi_2 * b2_{12} & \cdots & \varphi_2 * b2_{1n} \\ \varphi_2 * b2_{21} & & \ddots & \varphi_2 * b2_{2n} \\ \vdots & & & \vdots \\ \varphi_2 * b2_{n1} & \varphi_2 * b2_{n2} & \cdots & 0 \end{pmatrix} =$$

$$\begin{pmatrix} 0 & \varphi_1 * b1_{12} + \varphi_2 * b2_{12} & \cdots & \varphi_1 * b1_{1n} + \varphi_2 * b2_{1n} \\ \varphi_1 * b1_{21} + \varphi_2 * b2_{21} & \ddots & & \varphi_1 * b1_{2n} + \varphi_2 * b2_{2n} \\ \vdots & & & \vdots \\ \varphi_1 * b1_{n1} + \varphi_2 * b2_{n1} & \varphi_1 * b1_{n2} + \varphi_2 * b2_{n2} & \cdots & 0 \end{pmatrix}$$

As a numerical example, giving a geographical nearest neighbor matrix, with $k = 50$ neighbors

$$B_1 = \begin{pmatrix} 0 & 0 & \cdots & 0.02 \\ 0.02 & & \ddots & 0 \\ \vdots & & & \vdots \\ 0.02 & 0.02 & \cdots & 0 \end{pmatrix};$$

and a semantic proximity matrix

$$B_2 = \begin{pmatrix} 0 & 0.1 & \cdots & 0.42 \\ 0.15 & & \ddots & 0 \\ \vdots & & & \vdots \\ 0.60 & 0.02 & \cdots & 0 \end{pmatrix};$$

We estimate the convex combined parameters as:

$\varphi_1 = 0.74$ and $\varphi_2 = 0.26$.

Accordingly, we obtain:

$$W = \varphi_1 B_1 + \varphi_2 B_2 = 0.74 * \begin{pmatrix} 0 & 0 & \cdots & 0.02 \\ 0.02 & & \ddots & 0 \\ \vdots & & & \vdots \\ 0.02 & 0.02 & \cdots & 0 \end{pmatrix} + 0.26 * \begin{pmatrix} 0 & 0.1 & \cdots & 0.42 \\ 0.15 & & \ddots & 0 \\ \vdots & & & \vdots \\ 0.60 & 0.02 & \cdots & 0 \end{pmatrix} =$$

$$\begin{pmatrix} 0 & 0 & \cdots & 0.015 \\ 0.015 & & \ddots & 0 \\ \vdots & & & \vdots \\ 0.015 & 0.015 & \cdots & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0.026 & \cdots & 0.109 \\ 0.039 & & \ddots & 0 \\ \vdots & & & \vdots \\ 0.156 & 0.005 & \cdots & 0 \end{pmatrix} =$$

$$\begin{pmatrix} 0 & 0.026 & \cdots & 0.124 \\ 0.054 & & \ddots & 0 \\ \vdots & & & \vdots \\ 0.171 & 0.02 & \cdots & 0 \end{pmatrix}.$$