Italian Economic Journal
https://doi.org/10.1007/s40797-026-00407-z

RESEARCH PAPER

Knowledge Spillovers Versus Competitive Dynamics: Evidence from Startups and Scaleups in Milan's Tech Ecosystem

Alessandro Marra[1,2,3] · Andrea D'Isidoro[1] · Wided Matoussi[4]

Received: 27 May 2025 / Accepted: 27 August 2026

Abstract
Moving beyond the traditional focus on geographical proximity, this study investigates the existence and nature of spillover effects within a non-geographical space. The core novelty of the paper is operational and methodological: we exploit text data scraped from corporate websites to construct a novel semantic proximity matrix within a spatial econometric framework, capturing firms' industrial specialisations, technological adoptions, and cognitive competencies. Focusing on the tech ecosystem of Milan, the study examines how the balance between positive and negative externalities shifts across different phases of firm growth. From a conceptual perspective, the paper integrates growth with spillover dynamics, explaining the mechanisms through which firm maturity transforms knowledge-sharing opportunities into competitive frictions. We distinguish between two archetypes: younger, smaller startups operating within an exploratory innovation paradigm, and more mature scaleups driven by market exploitation and asset accumulation. The empirical results reconcile these perspectives by showing that the net effect of proximity is strictly contingent on the growth stage of neighbouring firms. Proximity to startups is associated with significant positive knowledge spillovers, consistent with their role as open, learning-oriented nodes. By contrast, proximity to scaleups triggers negative competitive pressures and "businessstealing" effects, as their aggressive market expansion imposes costs on technologically and industrially close neighbours. These findings offer fresh insights for researchers and policymakers, demonstrating that in modern tech ecosystems, the impact of proximity depends entirely on who is proximate.

✉ Alessandro Marra
alessandro.marra@unich.it

1 University G. d'Annunzio of Chieti-Pescara, Pescara, Italy

2 GRIF Research Centre, Luiss Guido Carli University, Rome, Italy

3 BLISS Research Centre, Ca Foscari University, Venice, Italy

4 University of Tunis, Tunis, Tunisia

Published online: 06 September 2026

Keywords Knowledge spillovers · Competitive dynamics · Startups · Tech ecosystems

JEL Classification R19 · O39

1 Introduction

Proximity is a double-edged sword for firm performance. While it is widely recognized as a key driver of collaboration, knowledge diffusion, and localized learning, it simultaneously intensifies product market rivalry, congestion, and "business-stealing" effects (Bloom et al. 2013; Davids and Frenken 2018). Although the coexistence of these countervailing mechanisms is well-established, conventional empirical frameworks face two major limitations. First, they often overlook how the balance between positive and negative spillovers shifts across different phases of firm growth. Second, they frequently rely on purely geographical distance, failing to capture the multi-layered industrial, technological, and cognitive spaces where modern firms actually interact.

This paper enters the picture to address these gaps by analysing the net impact of multi-dimensional proximity within the tech ecosystem of Milan. We explicitly account for firm heterogeneity by distinguishing between two distinct archetypes: younger, smaller, innovation-focused startups and more mature, larger, growth-oriented scaleups. We examine whether these different growth stages generate systematically opposite externalities. Specifically, we maintain and test whether startups act primarily as sources of positive, knowledge-based spillovers, whereas scaleups, driven by market exploitation and rapid expansion, tend to trigger negative competitive pressures that crowd out proximate rivals.

By doing so, the paper offers a twofold contribution. The operational and methodological contribution represents the core novelty of our work. Moving beyond traditional geographical or rigid categorical sector matrices (Debarsy and LeSage 2022; Marra and D'Isidoro 2025), we exploit text data scraped from corporate websites to construct a novel semantic proximity matrix within a spatial econometric framework. This approach allows us to map a non-geographical space that simultaneously captures firms' industrial specialisations, technological adoptions, and cognitive competencies. The conceptual contribution, more modestly, allows integrating and clarifying existing strands of literature on proximity, knowledge diffusion, and rivalry. We provide a tighter alignment between firms' growth phases and spillover dynamics, explaining the mechanisms through which firm maturity transforms knowledge-sharing opportunities into competitive frictions (Audretsch et al. 2021).

Our empirical results reconcile these perspectives by showing that the net effect of proximity is strictly contingent on the developmental stage of neighbouring firms. We find that proximity to startups yields significant positive externalities, confirming that early-stage ventures operate as open, learning-oriented nodes that enrich the local knowledge pool. Conversely, proximity to scaleups generates distinct negative competitive spillovers, demonstrating that as firms transition into aggressive market expansion, their growth imposes "business-stealing" costs on technologically and industrially close neighbours. These findings show that the impact of proximity depends entirely on who is proximate.

The paper is structured as follows. Section 2 reviews the literature and develops our research hypotheses. Section 3 describes the dataset and the context of Milan's tech ecosystem. Section 4 outlines our spatial econometric methodology and the construction of the semantic matrix. Section 5 presents and discusses the empirical results alongside robustness checks. Finally, Sect. 6 concludes by outlining policy implications, limitations, and future research paths.

2 Literature Review and Hypotheses Development

This Section integrates and clarifies existing strands of literature on proximity, knowledge spillovers, and competitive dynamics. By combining these perspectives, we aim to clarify how the indirect effects of proximity operate across different phases of firm growth, framing our empirical contribution which utilizes web-based semantic proximity within a spatial econometric framework.

The literature linking proximity to firm performance establishes that closeness facilitates interactions, collective learning, and access to complementary resources (Tubiana et al. 2022; Martin-Rios et al. 2022; Hallin and Lind 2012). While traditional spatial econometric models heavily rely on purely geographical proximity due to its exogeneity (Hazir and Autant-Bernard 2014; Ter Wal 2013), geographic distance alone is insufficient to capture the complexity of modern knowledge economies. Moving beyond purely physical boundaries requires incorporating non-geographical dimensions (such as industrial, technological, cognitive, and semantic proximities) to accurately map the channels of knowledge exchange and competitive pressure (Autant-Bernard 2012; Corrado and Fingleton 2012; Parent and LeSage 2008). These non-geographical dimensions capture different facets of alignment between firms. Cognitive and technological proximity reflect the degree of similarity and complementarity between firms' knowledge bases and technological capabilities, shaping their ability to absorb, integrate, and internalise external knowledge and advancements (Aldieri 2013; Kogler et al. 2013; Kaneva et al. 2023). Industrial proximity captures market communality and shared operational focus (Aarstad et al. 2016). Semantic proximity derived from corporate websites offers a dynamic, multi-layered reflection of a firm's actual market positioning, technological focus, and cognitive orientation (Marra et al. 2024; Marra and Baldassari 2022). Regardless of the specific dimension, the relationship between proximity and performance is inherently nonlinear (Boschma 2005; Nooteboom et al. 2007). An intermediate degree of similarity maximizes mutual understanding and competitive stimulation (Boschma et al. 2009; Timmermans and Boschma 2014); conversely, extreme cognitive or industrial distance prevents meaningful exchange, while excessive similarity induces severe redundancy and operational friction (Marra and D'Isidoro 2025).

As such, proximity activates two distinct, countervailing mechanisms that shape firm performance (Bloom et al. 2013; Carreira and Lopes 2018). On the positive side, proximity generates knowledge spillovers (Audretsch and Belitski 2020). Firms operating in similar fields or technological spaces benefit from a shared sectoral knowledge pool, localized synergies, and collaborative problem-solving (Quatraro 2010; Frenken et al. 2007; Neffke and Henning 2008). This collective learning environment accelerates innovation capacity, productivity, and growth (Cortinovis et al. 2020; Content et al. 2022; Jespersen et al. 2018). On the negative side, proximity intensifies competitive pressures and business-stealing effects. When firms operate in close technological or product market proximity, the expansion of one firm can lead to market share erosion, creative destruction, or duplicative patent races for its neighbours (Bloom et al. 2013; Goya et al. 2016; Hall et al. 2009). Under high market overlap, proximity can induce negative externalities: technological leaders may shut down local operations to prevent unwanted knowledge leakage, and intense competition for regional talent can drive up operational costs (Livanis and Lamin 2016; Hoberg and Phillips 2016).

Consequently, the net impact of proximity depends heavily on whether positive knowledge-sharing dynamics outweigh, or are neutralized by, these negative competitive frictions (Bernal et al. 2022; Audretsch and Belitski 2022; Smrkolj and Wagener 2017). The trade-off between knowledge spillovers and business-stealing is not uniform across the business registry (Cainelli and Ganau 2018; Wang 2015); rather, it is strictly moderated by firm heterogeneity, particularly regarding firm size and growth stage (Civera et al. 2025; Marra et al. 2024). The existing literature recognizes the distinct roles that different organizational archetypes (such as startups, spin-offs, scale-ups, and hidden champions) play within regional ecosystems (Audretsch and Keilbach 2007; Wennberg et al. 2011; Lehmann et al. 2025; Schenkenhofer et al. 2025). Empirical evidence suggests that smaller, younger firms exhibit a fundamentally different sensitivity to horizontal spillovers compared to larger, established, or R&D-intensive incumbents (Grillitsch and Nilsson 2019; Guerrero et al. 2023).

To understand why these differences emerge without falling into tautological assumptions, it is necessary to examine the specific organizational mechanisms, learning behaviours, and strategic resource constraints that characterize a firm's growth phases (Audretsch et al. 2021; Davids and Frenken 2018). On one side, early-stage, younger firms (startups) operate primarily within an exploratory innovation paradigm (Gilbert et al. 2008; Audretsch et al. 2015). Because they generally lack extensive internal R&D infrastructure, established intellectual property portfolios, and formal market power, their survival hinges on external learning networks, informal knowledge sourcing, and high employee mobility (Zhou et al. 2019b). Startups exhibit open organizational boundaries, making them highly permeable nodes within an ecosystem. Furthermore, because their strategic priority is product development, proof-of-concept, and technological validation rather than immediate, aggressive market-share consolidation, their interactions with proximate peers are predominantly collaborative rather than zero-sum (Marra et al. 2024). They contribute to and draw from local knowledge pools, acting as catalysts for regional growth and technology diffusion (Colombelli and Quatraro 2019). Because they do not yet pose an immediate threat of volume-based market displacement to neighbouring firms, the positive externalities of knowledge exchange remain uncompromised by business-stealing anxieties. Thus, we hypothesize:

> Hp#1: Younger, innovation-focused firms (startups) are more likely to generate positive knowledge spillovers.

On the other side, as firms mature and enter a rapid growth phase (scaleups), their strategic orientation shifts fundamentally from exploration to market exploitation, asset accumulation, and revenue scaling (Bock and Hackober 2020). Scaleups have developed robust internal R&D capabilities, formal organizational structures, and valuable proprietary technologies that they must actively protect from imitation (Livanis and Lamin 2016). Consequently, their organizational boundaries become less permeable, and their willingness to engage in informal, open-ended knowledge sharing decreases. More importantly, the defining objective of a scaleup is rapid market expansion. This aggressive pursuit of growth inevitably brings them into direct conflict with proximate competitors for finite regional resources such as specialised venture talent, local supply chain capacity, and customer attention (Bloom et al. 2013). At this developmental stage, high proximity to similar firms triggers intense product market rivalry. The expansion of a scaleup is likely to generate a net negative "business-stealing" externality for its neighbours, as its competitive gains often come at the direct expense of nearby rivals' market shares (Triguero and Fernández 2018; Arvanitis et al. 2020; Alcácer and Chung 2007). Therefore, we hypothesize:

> Hp#2: More mature, growth-oriented firms (scaleups) are more sensitive to competition and may generate negative competitive spillovers.

The transition from exploratory innovation to market exploitation can be further understood through the lens of technological regimes and the fundamental distinction between Schumpeter Mark I and Mark II patterns of innovation. According to Breschi et al. (2000), a Schumpeter Mark I regime is defined by "creative destruction", where technological ease of entry allows new entrepreneurs and young firms to play a central role in innovative activities. This aligns closely with our conceptualization of startups, which operate within an exploratory paradigm, relying on open organizational boundaries and external learning networks to validate new technologies. In contrast, a Schumpeter Mark II regime is characterized by "creative accumulation". In this regime, innovation is dominated by large, established firms that build upon an accumulated stock of knowledge, competencies in R&D, and significant financial resources, which collectively act as barriers to entry for new players. This archetype mirrors the strategic shift observed in scaleups: as these firms mature, their orientation moves toward revenue scaling. At this stage, the focus shifts from the "widening" of the innovative base to a "deepening" of market exploitation, where firms are more inclined to engage in aggressive expansion that may impose business-stealing costs on proximate rivals.

3 Data

Data were obtained from Dealroom (2024). Dealroom is a commercial database specializing in high-technology companies, with a particular focus on the European ecosystem. It integrates machine-learning techniques with user-contributed information, which is subsequently reviewed and validated by a dedicated team of analysts. The database has gained increasing attention in academic research in recent years, particularly for studies examining entrepreneurial ecosystems, high-growth firms, and innovation dynamics (van Meeteren et al. 2022; Consoli et al. 2025; Marra and D'Isidoro 2025).

Our initial dataset comprised approximately 3600 technology-oriented firms located in the city of Milan, Italy's business and financial capital, which generates a regional GDP exceeding €390 billion. Milan's economic structure is characterised by strong specialisations in life sciences, fintech, and fashion and design (OECD 2006). With more than 5000 multinational headquarters and several homegrown unicorns, the city represents a major European hub for advanced services and a particularly dynamic area, having expanded by roughly fifteenfold in terms of combined enterprise value over the past decade and accounting for nearly half of the total value of the Italian tech ecosystem (Antonietti et al. 2013; Arbia et al. 2012).

To ensure data reliability, we retained only companies exceeding a minimum data completeness threshold, as measured by the completeness component of the Dealroom Signal, a composite score that incorporates profile completeness, founder quality, employee growth, and the likelihood of an upcoming funding round. This filtering reduced the sample to approximately 1200 firms. It is worth noting that the collection and updating of economic and financial information by Dealroom benefit from collaborations with local institutional partners. In the case of Milan, the database provider established a partnership with Milano & Partners, the City of Milan, the Chamber of Commerce of Milano Monza Brianza Lodi, and Promos Italia, allowing to obtain reliable proxies for startup performance based on the knowledge of the local ecosystem, an otherwise challenging phenomenon to capture due to its highly dynamic and fragmented nature.

Firm-level financial and structural data were collected in March 2024. Textual data used to construct the semantic proximity matrix were collected through web scraping of official corporate websites conducted in April and May 2024. The empirical analysis adopts a short-term longitudinal perspective, focusing on the period 2022–2023. Specifically, the dependent variable (sales growth rate) is measured as the logarithmic difference between firms' revenues in 2023 and 2022. Similarly, key explanatory variables, including the average number of employees and employee growth rate, are computed over the same period. The initial scaleup status is defined based on firms' revenues at the beginning of the observation window (i.e., 2022), thereby reducing potential endogeneity concerns and capturing firms' baseline characteristics.

From the 1,200 companies, we further selected 505 firms for which consistent information on geographic location, sales, and employment was available. We assessed the representativeness of the final sample across several dimensions (spatial location in Table 1, sales in Table 2, and sectoral distribution in Table 3) confirming that it adequately reflects the broader population of Milan-based tech companies.

The City of Milan is divided into nine administrative zones (formerly Districts), consisting of a central circular zone (the Historic Center) and eight surrounding zones. Figure 1 illustrates these zones and reports, for each of them, the percentage of firms in our sample.

In line with Dealroom's classification, we distinguish between younger, smaller technology-oriented firms and more mature, larger companies. In addition, we identify whether firms display startup characteristics by observing whether they operate at the seed or early growth stage. Dealroom assigns firms to growth stages through a

Table 1 Distribution of firms' headquarters in the City of Milan by zone (Sample vs Total)

Zones	Sample (%)	Total (%)
Zone 1	30.8	33.1
Zone 2	4.0	3.1
Zone 3	9.8	8.2
Zone 4	10.5	9.5
Zone 5	7.8	7.7
Zone 6	6.2	6.5
Zone 7	7.4	6.0
Zone 8	13.6	15.0
Zone 9	9.9	10.9

Table 2 Distribution of firms by sales quartile, in logarithmic scale (Sample vs Total)

Quartile	Sample	Total
Min	0.90	0
25%	5.21	5.06
50%	5.85	5.60
75%	6.42	6.26
Max	10.08	10.21
Mean	5.92	5.73

Table 3 Distribution of firms by industry (Sample vs Total)

Industries	Sample (%)	Total (%)
Fintech	13.96	11.77
Marketing	9.89	9.84
Software enterprise	9.36	9.37
Media	7.77	7.37
Health	7.24	8.57
Food	6.18	7.13
Energy	4.77	4.32
Real estate	4.42	4.24
Home living	4.06	3.36
Transportation	3.53	3.68
Fashion	3.36	4.24
Education	3.36	3.36
Travel	2.65	3.04
Wellness	2.30	2.24
Sports	1.94	1.92
Others	15.19	13.77

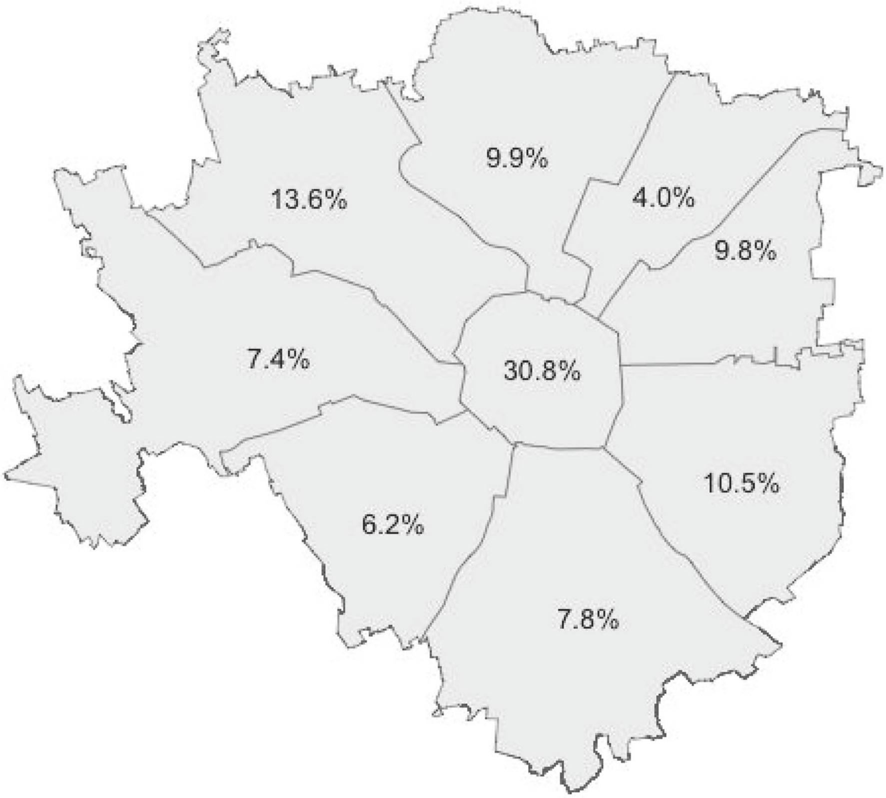

Fig. 1 Distribution of firms' headquarters in the City of Milan by zone

hierarchical classification framework that leverages on employment size, total funding, and firm age. Startups are identified as firms in the seed or early stages: seed-stage firms typically employ fewer than 10 workers, have raised less than €2 million in total funding, or are younger than approximately 1.5–2 years. Early-stage firms generally employ between 11 and 50 workers, have raised €2–10 million in funding, or are between 2 and 5 years old. Firms are classified as late-stage when they employ more than 50 workers, have raised over €10 million in funding, or are older than 5 years.

Startups account for roughly 84% of the sample. Even if companies in the dataset are technology-driven, technological intensity varies across sectors (Table 3).

As detailed in Sect. 4.2 (Model Specification), each firm is assigned a scale-up status measured using the logarithm of revenues at the beginning of the observation period, the average number of employees between 2022 and 2023, and the employee growth rate over the same period, reflecting the prevalence of very young firms in the dataset. In addition, we include a dummy variable distinguishing firms at the seed or early-growth stage from those at a later or more mature stage.

4 Methodology

4.1 Proximity Matrices

Our operational contribution consists in advocating the use of methodologies that exploit text data from corporate websites to construct measures of semantic proximity. This approach allows us to better capture firm interactions and to shed light on the mechanisms through which spillovers operate.

An increasing number of articles are leveraging the potential intrinsic in the descriptions of industrial activities and technological advancements, available on company websites and in other textual data sources (Nathan and Rosso 2015; Kinne and Axenbeck 2020; Papagiannidis et al. 2017). 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, and Zhou et al. (2019a) detect 'typical' research patterns to help technological recombination. With a similar purpose, Ranaei et al. (2020) use textual data from scientific publications. Qi et al. (2022) investigate partner selection for collaborative innovation by mining the content of patent documents, while Marra et al. (2020) use data provided by Crunchbase to explain merger and acquisition strategies in the clean tech sector. Nathan and Rosso (2015) illustrate how text data can deepen our understanding of digital industries. Hoberg and Phillips (2016) examine how firms differentiate themselves from competitors using novel text-based measures of product similarity derived from 10-K product descriptions. 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, Kinne and Axenbeck (2020) employ data from over two million company websites to discover emerging innovation ecosystems. Marra and Baldassari (2022) classify firms and identify technological trajectories across industries using text data from company websites. Losurdo et al. (2019) introduce an innovative measure of industrial specialisation in digital sectors, relying on adopted technologies rather than standard activity codes. Kinne and Lenz (2021) demonstrate the ability of text analysis and big data to uncover collaboration networks and innovative outcomes. Using textual and relational content from corporate websites, Abbasiharofteh et al. (2024) show that cognitive proximity is positively associated with a firm's level of innovation.

We start by collecting textual content from company websites, ensuring that the data is up-to-date and self-reported, accurately reflecting each firm's positioning. Keywords are then extracted to identify the firm's core industrial specialisations, technologies, knowledge and competencies. To standardise across firms, we employ Latent Dirichlet Allocation (LDA), a widely used probabilistic topic modelling method. LDA uncovers latent themes, providing a structured representation of industrial, technological, and cognitive domains and enabling the assignment of broader topics for higher-level generalisation while retaining relevant detail. Mapping firms onto a common set of topics increases comparability and supports the identification of semantic proximity and spillovers.

Keywords are processed using standard natural language processing (NLP) procedures, including text cleaning, tokenisation, lemmatisation, and stop-word removal. Texts are first cleaned by removing symbols, numbers, separators, and case distinctions using functions from the quanteda R package (Benoit et al. 2018). Tokenisation is then performed using the tokens() function, which constructs a tokens object and applies user-specified removal and splitting rules. At the end of this process, each firm is represented as a vector of tokens. Pairwise textual similarity between firms is then computed using cosine similarity, implemented through the textstat_simil family of functions in *quanteda*, which generate similarity matrices that can be easily converted into standard matrix or data-frame formats for subsequent analysis. Cosine similarity is suitable here as it measures orientation rather than magnitude, making it robust to differences in text length and allowing refined comparisons of semantic content across heterogeneous datasets.

Following Boschma (2005) and Nooteboom et al. (2007), we transform the linear cosine similarity scores into an inverted-U function to construct our semantic proximity matrix. This captures the non-linear relationship between proximity and knowledge exchange: moderate proximity maximises the potential for mutual understanding, collaboration, and competitive stimulation, whereas too little or too much similarity can hinder learning, limit idea diversity, and trigger business-stealing effects (Kok et al. 2020; Marra et al. 2020, 2024). In this way, the matrix captures how intermediate levels of proximity are most conducive to innovation and performance gains.

Accordingly, we build a spatial weights' matrix ($\boldsymbol{W}$), which is the semantic proximity matrix retrieved from the cosine similarity calculated above. The $\boldsymbol{W}$ matrix is assumed as exogenous and constructed according to a hyperbola: the values on the main diagonal are assumed as zero; moreover, if two firms have very low or very high cosine similarity, w_{ij} is again zero; then, since higher values of the cosine similarity may have a more pronounced impact on the extent of proximity or spillover effects, w_{ij} is expected to increase up to one (the maximum), if the cosine similarity grows towards the median.

For summary, some key aspects of the data collection and processing steps are summarised in Table 4.

4.2 Model Specification

Spillover effects among firms can be modelled using a spatial econometric framework. As a first step, we define a basic ordinary least square model (OLS):

$$y_i = \alpha + x_i'\beta + \varepsilon_i \tag{1}$$

where, y_i is the sales growth rate (Δ_rev) for the $i = 1 \ldots N$ firms between 2022 and 2023 calculated as the log difference between sales in 2023 and 2022 (Coad 2009; Dosi 2023); α is the intercept; β is the parameters' vector related to each of the covariates; ε_i is the error term. The $r = 1 \ldots P$ variables included in the matrix $\boldsymbol{X}$ are: a proxy for the scaleup status ($Scale\_t0$), measured by the log of revenues at the beginning of the observation period (Civera et al. 2025; Schenkenhofer et al. 2025);

Table 4 Steps for constructing the semantic proximity matrix from collected text data

Step	Method/tool	Notes
Data collection	Web scraping	From company websites, based on current and self-reported information
Keyword extraction	NLP-based	Focus on industrial specialisations, adopted technologies, knowledge and competencies
Topic modelling	Latent Dirichlet Allocation (LDA)	Needed to increase comparability
Text preprocessing	NLP	Lemmatisation, stop-word removal, tokenisation
Firm vector construction	Bag-of-words	Each firm has been represented as vector of tokens
Similarity computation	Cosine similarity	Useful because it measures orientation (rather than magnitude), robust to vector length differences
Proximity transformation	Inverted-U function	Needed to map similarity scores to non-linear semantic proximity, reflecting optimal levels for knowledge exchange. Kernel gaussian transformation around the median

the average number of employees between 2022 and 2023 (Avg_emp), reflecting the firm's stock of intangible or intellectual resources (Bloom et al. 2013; Balsmeier et al. 2014); a dummy variable ($Early$) distinguishing startups at the seed or early-growth stage (Audretsch et al. 2015; Rydehell et al. 2019; Guerrero et al. 2023); and the employee growth rate over the same period (Δ_emp), capturing the pace at which these resources expand, particularly relevant given the prevalence of very young firms in the dataset (Lindelöf and Löfsten 2004; Zhao et al. 2014). The use of employment growth as a regressor for sales growth is motivated by Coad (2010), who conceptualizes firm growth as a multidimensional and co-evolving process in which different growth indicators provide distinct but complementary information. From this perspective, employment growth may represent a key "stimulus" or strategic decision variable that precedes and contributes to subsequent firm expansion. In the context of the technology-oriented firms examined in our dataset, rapid workforce expansion represents a fundamental condition for scaling market activities, supporting organizational development, and commercializing exploratory innovations. Therefore, employment growth captures an important dimension of firms' scaling trajectories and provides relevant information for explaining variations in contemporaneous revenue performance.

Additionally, only in the semantic specification of the spatial model, we include the geographical location of companies, which serves to account for the physical space (Rao-Nicholson et al. 2019; Baldwin et al. 2008), using a synthetic variable that quantify the geographical distance, using latitude and longitude, of each firm from the barycentre of all the firm's locations, which is close to the centre of the city of Milan ($Bary$). Our choice is consistent with existing evidence on the spatial reach of spillovers: Elhorst et al. (2024) suggest that indirect effects are positive but decay with distance, while Baldwin et al. (2008) and Baum-Snow et al. (2024) find that the productivity effects of knowledge spillovers extend only few kilometres beyond individual plants.

The variables do not exhibit collinearity, as indicated by correlation coefficients that all fall within the interval [− 0.22; 0.44]. Notably, the number of employees and sales are only weakly correlated. This pattern is largely attributable to the high share of startups in the sample: in technology-intensive industries, it is common to observe small firms generating relatively high sales or, conversely, larger firms reporting limited or no sales at early stages. As a result, the observed correlation is weaker than would be expected in a sample with a higher proportion of mature firms.

The hypothesis of no spatial autocorrelation is tested using Moran's statistic (King 1981). Values significantly higher than expected lead to the rejection of the no correlation hypothesis in favour of the presence of positive spatial autocorrelation, justifying the introduction of spatial econometrics models.

Several procedures can be followed to select the best spatial specification. We rely on Elhorst (2014), who affirms that the best strategy to test for spatial interaction effects is to adopt a top-down approach, verifying step by step if some parameters should be included or not. In line with this approach, LeSage (2014) suggests that one of the best specifications to start with is the SDM, since it may overcome the bias related to omitted variables (Pace and LeSage 2010).

The SDM is defined as:

$$y_i = \alpha + x_i'\beta + \rho\sum_{j=1}^{N} w_{ij}y_j + \sum_{r=1}^{P}\sum_{j=1}^{N} w_{ij}x_{jr}\theta_r + \varepsilon_i \; with \; j \neq i \tag{2}$$

where ρ is the spatial autoregressive parameter, while the vector of parameters θ, for $r = 1 \ldots P$, measures the influence of the covariates on the dependent variables of neighbours.

To interpret the results, direct and indirect impacts should be computed by calculating the matrix of the partial derivatives of the expected value of y with respect to the rth explanatory variable (LeSage and Pace 2009).

Following the top-down approach, if the spatial parameter ρ results as not significant from the estimation of the SDM, the natural consequence is to choose the SLX specification. In favour of SLX adoption, Elhorst (2017) advises using models that include only the exogenous interaction effects (WX) in case of a high-density W. Moreover, Corrado and Fingleton (2012) suggest using this type of effects, showing that the endogenous spatial lag (WY) can be misleading, as it may only masks the effects of omitted spatially dependent variables included in WX. Additionally, Halleck Vega and Elhorst (2015) assert that the SLX model is one of the simplest in terms of computational burden and interpretative difficulties.

The SLX model takes the form:

$$y_i = \alpha + x_i'\beta + \sum_{r=1}^{P}\sum_{j=1}^{N} w_{ij}x_{jr}\theta_r + \varepsilon_i \; with \; j \neq i \tag{3}$$

which is a reduced form of a SDM without the inclusion of the endogenous effects, that is the spatial interactions between dependent variables.

With regards to the interpretation, in SLX direct impacts are represented by β coefficients, while θ_r parameters are the indirect effects (Elhorst 2014). For example, the estimated coefficient θ for the lag of the variable $Early$ gives immediately the extent of the indirect impact, interpreted as the influence of being a seed or early-growth firm on the performance of neighbours.

To discriminate between SDM and SLX, a recommended approach is to evaluate whether its nested models provide a better fit for the situation by using Likelihood Ratio tests (LR-tests; Elhorst 2014). The LR-test is structured as follows:

$$-2(LogL_{res} - LogL_{unres}) \tag{4}$$

where $LogL_{res}$ represents the log-likelihood of the restricted, and $LogL_{unres}$ represents the log-likelihood of the most general model (i.e., the SDM). The null hypothesis states that the SDM does not perform better than the simpler, more restricted model, making the latter preferable. This test statistic follows a Chi-squared distribution, with degrees of freedom equal to the number of restrictions applied.

Another specification which can be considered in model selection, according to LeSage and Pace (2009) and LeSage (2014), is the Spatial Durbin Error Model (SDEM). This model nests the SLX and adds the spatial autocorrelation between disturbances:

$$y_i = \alpha + x_i'\beta + \sum_{r=1}^{P}\sum_{j=1}^{N} w_{ij}x_{jr}\theta_r + u_i \; with \; u_i = \lambda\sum\nolimits_{j=1}^{N} w_{ij}u_{jr} + \varepsilon_i \; and \; j \neq i \tag{5}$$

where λ is the parameter which considers the spatial autocorrelation between the disturbances. This model could be compared to the SLX using LR-tests because they are nested, since SLX is a SDEM with the constriction λ = 0.

5 Results

5.1 Testing framework

Before using the proximity matrix in the statistical model of firms' performance, we tested for spatial autocorrelation in the residuals of the OLS model with the Moran's I test, for both semantic and geographical proximity (King 1981). Even though the

Table 5 Results of Moran's I test

W matrix	Moran I's statistic	Expectation	p-value
Semantic	0.043	– 0.002	0.000
Geographical	0.009	– 0.002	0.034

Table 6 Estimated parameters for OLS and SLX models

	OLS	SLX
Intercept	1.09***	1.10***
Scale_t0	– 0.015900***	– 0.016229***
Avg_emp	0.000005***	0.000005***
Grow_emp	0.005720***	0.005314**
Early	– 0.002020	– 0.002000
Bary	0.000002**	0.000002**
Lag_Scale_t0		– 0.023872*
Lag_Grow_emp		0.000031*
Lag_Avg_emp		0.022720**
Lag_Early		0.091369**
Lag_Bary		0.000004

p-values *< 0.1; **< 0.05; ***< 0.01

Moran's I statistic is not so far from 0, the p-value indicates that there is still a significant deviation from the expectation of no spatial autocorrelation (Table 5).

Table 6 shows the estimates for both OLS model and SLX model, estimated via Maximum Likelihood (ML), indicating that the coefficients are statistically significant and have the expected signs. This circumstance is confirmed across the two specifications, since the direct impacts of the SLX model show same polarity and similar magnitudes with the OLS coefficients.

Regarding the indirect impacts, results are in line with the literature, by which startups at the seed or early-growth stage ($Early$) are more prone to exchange knowledge (Audretsch et al. 2015; Rydehell et al. 2019; Guerrero et al. 2023). In contrast, neighbours with a higher value for the variable $Scale\_t0$ are expected to generate negative effects, implying some form of competition (Guerrero et al. 2023; Arvanitis et al. 2020; Marra et al. 2024). Broadly speaking, neighbours with higher values of the variables capturing the size of the firms' knowledge base or the pace at which it expands (namely, Avg_emp and Δ_emp) are expected to generate stronger knowledge spillovers (Bloom et al. 2013; Balsmeier et al. 2014). The significance of parameters θ, and the consequent estimated impacts, confirms that the semantic proximity matrix is relevant.

To support the selection of the SLX model, Table 7 summarises the LR-tests performed, with all models estimated via ML. The results indicate that this model is the only one that outperforms the SDM, and the SDEM, making it the most suitable for

Table 7 LR-tests of SDM versus nested models

Model	LR-tests vs SDM	LR-test vs SDEM
OLS	28.769***	28.649***
SLM	16.562***	
SLX	0.096	0.835

***p-value < 0.01

interpreting the current data. This provides evidence that global effects are challenging to justify, particularly when contrasted with local effects (Gibbons and Overman 2012; Lacombe and LeSage 2013).

5.2 Robustness and Endogeneity

The validity of estimated statistical models and their results is typically assessed through robustness checks. These checks aim to evaluate the reliability of key coefficients by altering the model specification. In cases like the one presented in this paper, where the estimated model coefficients align in both polarity and magnitude with those of the OLS and are theoretically plausible, these coefficients may be assumed to be reliable (Lu and White 2014). However, it is always advisable to verify the robustness of the core coefficients by testing their insensitivity to alternative estimations and by introducing redundant variables.

Table 8 presents a comparison of the direct and indirect impacts among three models: the chosen SLX model (Model 1), the corresponding SDM (Model 2), and an SLX model that includes two additional independent variables, specifically Age and $B2B$ (Model 3).

Age is a standard determinant in the literature on firm performance (Coad et al. 2017), as it captures the transition from the experimental routines of very young ventures to the more established operational structures of mature firms. By controlling for age alongside initial revenues ($Scale\_t0$), we ensure that the observed performance effects are specifically linked to its growth stage and scaleup status. The $B2B$ dummy accounts for the fundamental differences in how firms interact with their ecosystem; firms operating in a business-to-business context are inherently more embedded in professional networks, fostering a greater degree of reciprocal exchange of knowledge, ideas, and supply-chain opportunities (Marra and D'Isidoro 2025).

All the estimations, obtained via ML, exhibit the same polarity and similar magnitudes, with very few variables occasionally proving insignificant. This consistency reinforces the robustness of our results.

Additionally, to address endogeneity which could emerge in our specification due to reverse causality of employees related variables, or common unexpected factors that influence start-ups, we employ a 2SLS estimation using the recommended instruments, WX, W^2X, and W^3X where, in the case of the SLX, WX is an internal instrument and W^2 and W^3 are, respectively, the second and the third spatial lag (Kelejian and Prucha 1998; Baltagi et al. 2014). However, although these instruments are recommended,

Table 8 Direct and indirect impacts of model 1 (the adopted model, SLX), model 2 (SDM), and model 3 (SLX with 2 added variables)

	Model 1	Model 2	Model 3
Direct impacts			
Scale_t0	– 0.016229***	– 0.015680***	– 0.017340***
Avg_emp	0.000005***	0.000005***	0.000003*
Grow_emp	0.005314**	0.005411***	0.005136***
Early	– 0.002000	– 0.003758	– 0.003371
Bary	0.000002**	0.000002***	0.000002*
Age			0.000169
B2B			0.000084
Indirect impacts			
Scale_t0	– 0.023872*	– 0.025806**	– 0.021590**
Avg_emp	0.000031*	0.000008	0.000024
Grow_emp	0.022720**	0.008906**	0.024220***
Early	0.091369**	– 0.006186	0.083620**
Bary	0.000004	0.000004	0.000001
Age			0.000393
B2B			– 0.015280

p-values *< 0.1; **< 0.05; ***< 0.01

we perform test of over-identification and weak instruments to ensure their goodness. The results are shown in Table 9.

Weak instruments tests show that the null hypothesis of weakness is rejected in almost all the cases, while the over-identification test cannot reject the null hypothesis of uncorrelation between instruments and error term, proving the validity of the chosen instruments.

Table 9 Test of over-identification and weak instruments

Type of test		Test value
Weak instruments tests	Scale_t0	3.696***
	Avg_emp	16.956***
	Grow_emp	3.741***
	Early	0.557
	Bary	2.153**
Over-identification test	Sargan test	1.42

p-values *< 0.1; **< 0.05; ***< 0.01

5.3 Discussion of Results

We focus on two different impacts: the direct impacts on the dependent variable when covariates change, and the indirect effects (or spillovers), namely the impacts of a covariate on the dependent variable of neighbouring units. In the SLX we define each coefficient β as the direct impact of the associated variable and the parameters θ as the indirect impacts, while in the SDM we must estimate the impacts.

Regarding the SLX model, all the coefficients concerning the direct effects are statistically significant, except for the growth stage. As expected, the negative coefficient of $Scale\_t0$ highlights that the effect of the initial scaleup status is negative, as smaller firms tend to grow more and vice versa. The statistical significance of the $Bary$ variable indicates that the location of firms within the city of Milan is important, confirming the role of the physical space for firm interdependence.

Except for the geographical variable, indirect effects are statistically significant. Such effects are so relevant that they emerge even if limited to first order interaction effects, that is within small 'local' networks of neighbours, without considering the 'global' scope resulting from the propagation of feedback effects (Gibbons and Overman 2012; Lacombe and LeSage 2013). More specifically, the SLX model considers only local spillovers from one unit to the neighbouring ones, neglecting global spillovers involved in SDM and SLM models, which include feedback effects because of impacts passing through neighbouring units (LeSage and Pace 2011; Halleck Vega and Elhorst 2015). Such a result also captures more intuitively the dynamics typical of technology ecosystems (Boyer et al. 2021). Through channels such as collaboration, imitation, and competition, these interactions strengthen or undermine learning, competitive stimuli, and growth, thereby affecting the performance trajectories of participating firms (Fotopoulos and Qian 2025).

As anticipated, these effects are not unconditionally beneficial (Smrkolj and Wagener 2017; Alcácer and Chung 2007): when competitive pressures intensify, such as when firms compete for scarce resources, talent, or market opportunities, these same interactions may trigger counterproductive dynamics (Bloom et al. 2013; Arvanitis et al. 2020). In such cases, initially positive spillovers can be offset or even reversed, as firms may engage in 'defensive' behaviours, limit knowledge sharing, or experience business-stealing effects that reduce the overall gains for the ecosystem (Livanis and Lamin 2016; Goya et al. 2016; Hall et al. 2009).

We find that the neighbours of scaleups do not benefit from positive spillover effects: the negative indirect effects can be attributed to the cannibalisation of each other's business (Zhou et al. 2019b; Marra et al. 2024; Guerrero et al. 2023). This result can also be interpreted in line with the knowledge equilibrium argument, by which knowledge externalities diminish the performance of high-growth firms to benefit low-growth firms, until performance disparities disappear (Grillitsch and Nilsson 2019). Conversely, we acknowledge the positive spillovers generated by startups that exhibit a strong propensity to innovate and originate spillovers that benefit neighbours (Audretsch et al. 2015; Audretsch and Keilbach 2007; Gilbert et al. 2008). The distinction between startups and scaleups is partly mitigated by the positive effects associated with viewing the firm as a 'knowledge base', which are more pronounced in firms with a larger number of employees: each employee would contribute individual knowledge to the firm's aggregate knowledge base, and a larger workforce would increase opportunities for interaction and exchange with employees from other companies, thereby enhancing the overall potential for spillovers (Martin-Rios et al. 2022; Tubiana et al. 2022). Given that most of the observed technology companies are startups, we cannot overlook the firm's employee growth rate, which expands the firm's knowledge base and appears to generate positive indirect effects.

Accordingly, the results support both hypotheses, which cannot be rejected given the statistical significance of the estimated indirect effects across the different models and specifications. While data limitations prevent a more direct empirical investigation of the underlying mechanisms, our analysis contributes to the empirical discussion by interpreting these findings considering well-established theoretical and empirical literature.

6 Conclusions

In this paper, our contribution to the literature is twofold. First, the operational and methodological contribution represents the core novelty of our work. Moving beyond traditional geographical or rigid categorical matrices, we exploit text data scraped from corporate websites to construct a novel semantic proximity matrix within a spatial econometric framework. This approach allows us to map a non-geographical space that simultaneously captures firms' industrial specialisations, technological adoptions, and cognitive competencies. Second, our conceptual contribution integrates existing strands of literature by providing a tighter alignment between firms' growth phases and spillover dynamics (Audretsch et al. 2015, 2021; Bloom et al. 2013). This framework allows us to explain the mechanisms through which firm maturity transforms knowledge-sharing opportunities into competitive frictions (Audretsch and Belitski 2022; Guerrero et al. 2023). Our results reconcile these perspectives by showing that the net effect of proximity is strictly contingent on the developmental stage of neighbouring firms. On the one hand, we find that startups, which operate within an exploratory innovation paradigm, act as open, learning-oriented nodes that generate significant positive knowledge spillovers. Since their priority is technological validation rather than immediate market-share consolidation, their interactions are predominantly collaborative. On the other hand, the negative indirect effects associated with scaleup status point to the prevalence of competitive spillovers and "business-stealing" effects. As firms mature and shift toward market exploitation and asset accumulation, their organizational boundaries become less permeable, and their aggressive expansion imposes costs on technologically and industrially close neighbours. These findings demonstrate that the impact of proximity depends entirely on who is proximate.

Since our empirical evidence is limited to the case of the city of Milan, the results do not provide a universally generalizable basis for drawing clear policy implications, but they nonetheless offer informative insights that may be relevant for similar urban and innovation-driven contexts. Policymakers aiming to foster innovation and economic growth could benefit from tailoring interventions to the composition of firms within a given ecosystem (Audretsch et al. 2015; Civera et al. 2025). Policies promoting collaboration and knowledge-sharing among startups (also through intermediary organizations such as funding schemes for joint research projects, co-working spaces, and innovation hubs) can amplify positive spillover effects, especially when leveraging industrial, technological, or cognitive similarities. Conversely, in areas characterised by a high density of scaleups, policy interventions should aim at mitigating excessive competitive pressures while still harnessing the dynamism of these firms. This could involve supporting upskilling programmes, promoting complementary business opportunities, and encouraging partnerships between neighbouring firms, including startups, to balance competition with cooperation. More broadly, urban planners and regional development agencies could use these insights to design spatial strategies that optimise the co-location of startups and scaleups.

This paper is subject to several limitations. While the database on startups and technology-based firms offers important advantages, the availability of detailed economic and financial variables is inherently limited, reflecting the well-known data constraints affecting young and rapidly evolving firms. These limitations influence the size of the observed sample, the selection of variables, and, consequently, the overall model specification.

Despite these constraints, our findings point to plausible mechanisms that may help explain why proximity to scaleups is more likely to generate negative indirect effects for neighbouring firms. Competitive forces play a role, but they do not operate in isolation. Scaleups typically operate at a more advanced stage of R&D and market development and benefit from superior access to financial, organisational, and reputational resources. Underlying mechanisms may include market saturation, the reallocation of skilled labour and specialised suppliers toward the scaleup, and the concentration of investor attention on a leading firm at the expense of its neighbours. In addition, scaleups often rely on more standardised products, technologies, and business models, which may limit opportunities for learning and differentiation among proximate firms, thereby amplifying rivalry. By contrast, startups are more likely to engage in exploratory activities and informal interactions, making proximity more conducive to positive spillovers. Future research combining richer firm-level data would allow these mechanisms to be investigated more precisely and would further strengthen the interpretation of our results.

Moving forward in our research programme, the next step is to disentangle and jointly model multiple forms of non-geographical and geographical proximity to account for their potentially distinct contributions to the generation of spillovers (Parent and LeSage 2008; Debarsy and LeSage 2022). We aim to assess whether knowledge-driven effects are more closely associated with cognitive or technological proximity, while competitive-driven effects are more likely to arise from industrial proximity. To this end, we plan to adopt a modelling strategy in which interpretability represents a key criterion: the convex combination approach would allow multiple (more than two) proximity structures to be integrated into a single connectivity matrix, with each component weighted by a parameter capturing its relative contribution (Debarsy and LeSage 2022). This framework would generalise beyond a standard two-matrix setting (e.g., geographical and semantic proximity) to an arbitrary number L of matrices, enabling the inclusion of additional separate dimensions such as industrial, technological, or cognitive proximity (Debarsy and LeSage 2022; Pace and LeSage 2010; Hazır et al. 2018). This approach would avoid many of the estimation and interpretation challenges associated with higher-order spatial models, while yielding parameters that directly reflect the relative importance of each form of proximity in shaping spillovers.

Funding Open access funding provided by Università degli Studi G. D'Annunzio Chieti Pescara within the CRUI-CARE Agreement. This work is part of a research program entitled "Firms within networks of words: Using text data to measure proximity between firms and firm performance", of which I am the Principal Investigator, funded by the Italian Ministry of University and Research as a Project of Relevant National Interest (PRIN 2022). Code: 20224EATBN. CUP: D53D23006120006.

Data Availability Authors have associated data in a data repository. Data will be made available on reasonable request.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Aarstad J, Kvitastein OA, Jakobsen SE (2016) Related and unrelated variety as regional drivers of enterprise productivity and innovation: a multilevel study. Res Policy 45(4):844–856

Abbasiharofteh M, Kinne J, Krüger M (2024) Leveraging the digital layer: the strength of weak and strong ties in bridging geographic and cognitive distances. J Econ Geogr 24(2):241–262. https://doi.org/10.1093/jeg/lbad037

Alcácer J, Chung W (2007) Location strategies and knowledge spillovers. Manag Sci 53:760–776. https://doi.org/10.1287/mnsc.1060.0637

Aldieri L (2013) Knowledge technological proximity: evidence from US and European patents. Econ Innov New Technol 22(8):807–819

Antonietti R, Cainelli G, Lupi C (2013) Vertical disintegration and spatial co-localization: the case of Kibs in the metropolitan region of Milan. Econ Lett 118(2):360–363

Arbia GM, Espa G, Giuliani D, Mazzitelli A (2012) Clusters of firms on an inhomogeneous space: the high-tech industries in Milan. Econ Model 29(1):3–11

Arvanitis S, Seliger F, Woerter M (2020) Knowledge spillovers, competition and innovation success†. Oxf Bull Econ Stat 82:1017–1041. https://doi.org/10.1111/obes.12365

Audretsch D, Belitski M (2020) The role of R&D and knowledge spillovers in innovation and productivity. Eur Econ Rev 123:103391. https://doi.org/10.1016/j.euroecorev.2020.103391

Audretsch D, Belitski M (2022) The knowledge spillover of innovation. Ind Corp Change. https://doi.org/10.1093/icc/dtac035

Audretsch DB, Keilbach M (2007) The theory of knowledge spillover entrepreneurship. J Manag Stud 44(7):1242–1254. https://doi.org/10.1111/j.1467-6486.2007.00722.xR

Audretsch DB, Belitski M, Desai S (2015) Entrepreneurship and economic development in cities. Ann Reg Sci 55(1):33–60. https://doi.org/10.1007/s00168-015-0685-x
Audretsch D, Belitski M, Caiazza R (2021) Start-ups, innovation and knowledge spillovers. J Technol Transf 46:1995–2016. https://doi.org/10.1007/s10961-021-09846-5
Autant-Bernard C (2012) Spatial econometrics of innovation: recent contributions and research perspectives. Spat Econ Anal 7(4):403–419. https://doi.org/10.1080/17421772.2012.722665
Baldwin J, Beckstead D, Brown M, Rigby D (2008) Agglomeration and the geography of localization economies in Canada. Reg Stud 42:117–132. https://doi.org/10.1080/00343400701543199
Balsmeier B, Buchwald A, Stiebale J (2014) Outside directors on the board and innovative firm performance. Res Policy 43(10):1800–1815. https://doi.org/10.1016/j.respol.2014.06.003
Baltagi BH, Fingleton B, Pirotte A (2014) Spatial lag models with nested random effects: an instrumental variable procedure with an application to English house prices. J Urban Econ 80:76–86
Baum-Snow N, Gendron-Carrier N, Pavan R (2024) Local productivity spillovers. Am Econ Rev 114(4):1030–1069
Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, Matsuo A (2018) Quanteda: an R package for the quantitative analysis of textual data. J Open Source Softw 3(30):774–774
Bernal P, Carree M, Lokshin B (2022) Knowledge spillovers, R&D partnerships and innovation performance. Technovation. https://doi.org/10.1016/j.technovation.2022.102456
Bloom N, Schankerman M, Van Reenen J (2013) Identifying technology spillovers and product market rivalry. Econometrica 81:1347–1393
Bock C, Hackober C (2020) Unicorns—what drives multibillion-dollar valuations? Bus Res 13(3):949–984
Boschma R (2005) Proximity and innovation: a critical assessment. Reg Stud 39(1):61–74. 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. J Econ Geogr 9(2):169–190
Boyer J, Ozor J, Rondé P (2021) Local innovation ecosystem: structure and impact on adaptive capacity of firms. Ind Innov 28(5):620–650. https://doi.org/10.1080/13662716.2021.1891407
Breschi S, Malerba F, Orsenigo L (2000) Technological regimes and Schumpeterian patterns of innovation. Econom J 110(463):388–410
Cainelli G, Ganau R (2018) Distance-based agglomeration externalities and neighbouring firms' characteristics. Reg Stud 52:922–933. https://doi.org/10.1080/00343404.2017.1360482
Carreira C, Lopes L (2018) Regional knowledge spillovers: a firm-based analysis of non-linear effects. Reg Stud 52(7):948–958. https://doi.org/10.1080/00343404.2017.1360484
Civera A, Schenkenhofer J, Vismara S (2025) A taxonomy of the knowledge spillover theory of entrepreneurship: firm types and knowledge sources. J Technol Transf. https://doi.org/10.1007/s10961-025-10227-5
Coad A (2009) The growth of firms: a survey of theories and empirical evidence. The growth of firms. Edward Elgar Publishing
Coad A (2010) Exploring the processes of firm growth: evidence from a vector auto-regression. Ind Corp Change 19(6):1677–1703
Coad A, Holm J, Krafft J, Quatraro F (2017) Firm age and performance. J Evol Econ 28(1):1–11
Colombelli A, Quatraro F (2019) Green start-ups and local knowledge spillovers from clean and dirty technologies. Small Bus Econ 52(4):773–792. https://doi.org/10.1007/s11187-017-9934-y
Consoli D, Lelli F, Montresor S, Perruchas F, Rentocchini F (2025) Moneytalks. the role of (spatial and digital) proximity in the VC financing of green start-ups (No. 2521). Utrecht University, Department of Human Geography and Spatial Planning, Group Economic Geography
Content J, Cortinovis N, Frenken K, Jordaan J (2022) The roles of KIBS and R&D in the industrial diversification of regions. Ann Region Sci 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? J Reg Sci 52(2):210–239
Cortinovis N, Crescenzi R, Van Oort F (2020) Multinational enterprises, industrial relatedness and employment in European regions. J Econ Geogr 20(5):1165–1205
Davids M, Frenken K (2018) Proximity, knowledge base and the innovation process: towards an integrated framework. Reg Stud 52(1):23–34. https://doi.org/10.1080/00343404.2017.1287349
Dealroom (2024) Data on the Milan's tech ecosystem. https://innovationmap.yesmilano.it/dashboard. Accessed 23 Mar 2024

Debarsy N, LeSage JP (2022) Bayesian model averaging for spatial autoregressive models based on convex combinations of different types of connectivity matrices. J Bus Econ Stat 40(2):547–558. https://doi.org/10.1080/07350015.2020.1840993
Dosi G (2023) The foundations of complex evolving economies: part one: innovation, organization, and industrial dynamics. Oxford University Press
Elhorst JP (2014) Spatial econometrics: from cross-sectional data to spatial panels, vol 479. Springer, Heidelberg, p 480
Elhorst JP (2017) Spatial panel data analysis. Encycl GIS 2:2050–2058
Elhorst JP, Tziolas I, Tan C, Milionis P (2024) The distance decay effect and spatial reach of spillovers. J Geogr Syst 26(2):265–289
Fotopoulos G, Qian H (2025) Entrepreneurial innovation ecosystems in the United States: configurational analysis of the high-tech economy. Small Bus Econ. https://doi.org/10.1007/s11187-025-01109-z
Frenken K, Van Oort FG, Verburg T (2007) Related variety, unrelated variety and regional economic growth. Reg Stud 41(5):685–697
Gibbons S, Overman HG (2012) Mostly pointless spatial econometrics? J Reg Sci 52(2):172–191
Gilbert BA, McDougall PP, Audretsch DB (2008) Clusters, knowledge spillovers and new venture performance: an empirical examination. J Bus Ventur 23(4):405–422. https://doi.org/10.1016/j.jbusvent.2007.04.003R
Goya E, Vayá E, Suriñach J (2016) Innovation spillovers and firm performance: micro evidence from Spain (2004–2009). J Prod Anal 45(1):1–22
Grillitsch M, Nilsson M (2019) Knowledge externalities and firm heterogeneity: effects on high and low growth firms. Pap Reg Sci 98(1):93–114. https://doi.org/10.1111/pirs.12342
Guerrero AJ, Heijs J, Huergo E (2023) The effect of technological relatedness on firm sales evolution through external knowledge sourcing. J Technol Transf 48(2):476–514. https://doi.org/10.1007/s10961-022-09931-3
Hall BH, Mairesse J, Mohnen P (2009) Measuring the returns to R&D. NBER Working Papers 15622, National Bureau of Economic Research, Inc.
Halleck Vega S, Elhorst JP (2015) The SLX model. J Reg Sci 55(3):339–363
Hallin C, Lind C (2012) Revisiting the external impact of MNCs: an empirical study of the mechanisms behind knowledge spillovers from MNC subsidiaries. Int Bus Rev 21:167–179. https://doi.org/10.1016/J.IBUSREV.2010.12.003
Hazır CS, LeSage J, Autant-Bernard C (2018) The role of R&D collaboration networks on regional knowledge creation: evidence from information and communication technologies. Pap Reg Sci 97(3):549–568
Hazir CS, Autant-Bernard C (2014) Determinants of cross-regional R&D collaboration: some empirical evidence from Europe in biotechnology. Ann Reg Sci 53(2):369–393
Hoberg G, Phillips G (2016) Text-based network industries and endogenous product differentiation. J Polit Econ 124(5):1423–1465
Jespersen K, Rigamonti D, Jensen MB, Bysted R (2018) Analysis of SMEs partner proximity preferences for process innovation. Small Bus Econ 51(4):879–904. https://doi.org/10.1007/s11187-017-9969-0
Kaneva M, Untura G, Zabolotsky A (2023) The impact of geographical, technological, and cognitive proximities on knowledge creation in the Russian regions. J Knowl Econ. https://doi.org/10.1007/s13132-023-01552-x
Kelejian HH, Prucha IR (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. J Real Estate Financ Econ 17:99–121
King ML (1981) A small sample property of the Cliff-Ord test for spatial correlation. J R Stat Soc Ser B Stat Methodol 43(2):263–264
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):e0249071
Kogler DF, Rigby DL, Tucker I (2013) Mapping knowledge space and technological relatedness in US cities. Eur Plan Stud 21:1374–1391
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. Res Policy 49:104011. https://doi.org/10.1016/j.respol.2020.104011

Lacombe DJ, LeSage JP (2013) Using Bayesian posterior model probabilities to identify omitted variables in spatial regression models. Pap Reg Sci. https://doi.org/10.1111/pirs.12070
Lehmann EE, Schenkenhofer J, Vismara S (2025) Hidden champions and knowledge spillovers: innovation-enhancing agglomeration effects and niche technology specificity. Small Bus Econ 66:827–847
LeSage JP (2014) What regional scientists need to know about spatial econometrics. Rev Reg Stud 44:13–32
LeSage JP, Pace RK (2009) Introduction to spatial econometrics. Taylor & Francis
LeSage JP, Pace RK (2011) Pitfalls in higher order model extensions of basic spatial regression methodology. Rev Region Stud 41(1):13–26
Lindelöf P, Löfsten H (2004) Proximity as a resource base for competitive advantage: university-industry links for technology transfer. J Technol Transf 29(3–4):311–326. https://doi.org/10.1023/b:jott.0000034125.29979.ae
Livanis G, Lamin A (2016) Knowledge, proximity and R&D exodus. Res Policy 45:8–26. https://doi.org/10.1016/j.respol.2015.09.002
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. Pap Reg Sci 98(2):737–753. https://doi.org/10.1111/pirs.12376
Lu X, White H (2014) Robustness checks and robustness tests in applied economics. J Econom 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):e0270041. https://doi.org/10.1371/journal.pone.0270041
Marra A, D'Isidoro A (2025) Combining geographical and semantic proximity to measure spillovers: the case of Sweden. Spat Econ Anal. https://doi.org/10.1080/17421772.2025.2586550
Marra A, Carlei V, Baldassari C (2020) Exploring networks of proximity for partner selection, firms' collaboration and knowledge exchange. The case of clean-tech industry. Bus Strateg Environ 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 Bus Rev. https://doi.org/10.1007/s40821-024-00270-x
Martin-Rios C, Erhardt NL, Manev IM (2022) Interfirm collaboration for knowledge resources interaction among small innovative firms. J Bus Res 153:206–215. https://doi.org/10.1016/j.jbusres.2022.08.024
Nathan M, Rosso A (2015) Mapping digital businesses with big data: some early findings from the UK. Res Policy 44:1714–1733
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
Nooteboom B, Van Haverbeke W, Duysters G, Gilsing V, van den Oord A (2007) Optimal cognitive distance and absorptive capacity. Res Policy 36(7):1016–1034
OECD (2006) OECD territorial reviews: Milan, Italy. OECD, Paris
Pace RK, LeSage JP (2010) Omitted variable biases of OLS and spatial lag models. Progress in spatial analysis: methods and applications. Springer, pp 17–28
Papagiannidis S, See-To EWK, Assimakopoulos DG, Yang Y (2017) Identifying industrial clusters with a novel big-data methodology: are SIC codes (not) fit for purpose in the Internet age? Comput Oper Res 98:355–366
Parent O, LeSage JP (2008) Using the variance structure of the conditional autoregressive spatial specification to model knowledge spillovers. J Appl Econom 23(2):235–256
Peng W, Yu X, Ji Y (2023) Obtaining advantages from knowledge base: mapping the potential to develop new technologies. Technol Anal Strateg Manag. https://doi.org/10.1080/09537325.2023.2242519
Qi Y, Zhang X, Hu Z, Xiang B, Zhang R, Fang S (2022) Choosing the right collaboration partner for innovation: a framework based on topic analysis and link prediction. Scientometrics 127(9):5519–5550
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. Res 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

Rao-Nicholson R, Khan Z, Akhtar P (2019) Nature of technology and location effects on firm performance in the US medical device industry. Econ Innov New Technol 28(5):498–517. https://doi.org/10.1080/10438599.2018.1536357

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. Pap Reg Sci 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. J Technol Transf 44(5):1547–1576. https://doi.org/10.1007/s10961-018-9668-2

Schenkenhofer J, Block J, Vismara S (2025) University knowledge spillovers and innovation of hidden champions: evidence from Italy. R&D Manag 55(2):598–613

Smrkolj G, Wagener F (2017) Research among copycats: R&D, spillovers, and feedback strategies. Entrepreneurship. https://doi.org/10.2139/ssrn.2485078

Ter Wal AL (2013) Cluster emergence and network evolution: a longitudinal analysis of the inventor network in Sophia-Antipolis. Reg Stud 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. J Econ Geogr 14(2):289–311

Triguero A, Fernández S (2018) Determining the effects of open innovation: the role of knowledge and geographical spillovers. Reg Stud 52(5):632–644. https://doi.org/10.1080/00343404.2017.1395004

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

van Meeteren M, Trincado-Munoz F, Rubin TH, Vorley T (2022) Rethinking the digital transformation in knowledge-intensive services: a technology space analysis. Technol Forecast Soc Change 179:121631. https://doi.org/10.1016/j.techfore.2022.121631

Wang C (2015) Geography of knowledge sourcing, search breadth and depth patterns, and innovative performance: a firm heterogeneity perspective. Environ Plann A 47:744–761. https://doi.org/10.1068/a140024p

Wennberg K, Wiklund J, Wright M (2011) The effectiveness of university knowledge spillovers: performance differences between university spinoffs and corporate spinoffs. Res Policy 40(8):1128–1143. https://doi.org/10.1016/j.respol.2011.05.014R

Zhao M, Park S, Zhou N (2014) MNC strategy and social adaptation in emerging markets. J Int Bus Stud 45:842–861. https://doi.org/10.1057/jibs.2014.8

Zhou X, Huang L, Zhang Y, Yu M (2019a) 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 (2019b) Learning from yourself or learning from neighbours: knowledge spillovers, institutional context and firm upgrading. Reg Stud 53(10):1397–1409. https://doi.org/10.1080/00343404.2019.1566705