Formation of New Startups: Does Business Diversity in Existing Startups Matter?

Chigusa Okamoto *†‡

Abstract

Does a diverse mix of startup businesses within a specific area promote the creation of new startups, and if so, which types of diversity are most influential? To answer this question, this study focuses on growth-oriented startups in central Tokyo and examines how business diversity of incumbent startups influences the number of newly established startups within 500-meter and 1,000-meter mesh units, using a negative binomial regression model. Building on existing literature, this study decomposes diversity into related and unrelated varieties. Relatedness is defined based on startup business fields specifically developed for this study. Accordingly, related variety reflects the degree of diversity within startups' business fields, whereas unrelated variety captures the diversity across different business fields. The findings indicate that both related and unrelated varieties contribute to startup creation in both mesh units, supporting previous research that suggests both types of variety promote entrepreneurship. However, as the size of the area increases, the importance of unrelated variety relative to related variety declines. This suggests that as geographical proximity decreases, cognitive proximity, as represented by business relatedness, becomes more important, implying that geographical proximity may serve as a substitute for cognitive proximity.

*Faculty of Economics, Chuo University; *e-mail:* ochigusa002@g.chuo-u.ac.jp

†This research is part of a collaborative project at Center for Real Estate Innovation (CREI) at The University of Tokyo, conducted in partnership with For Startups, Inc. It is also carried out in collaboration with the Center for Spatial Information Science (CSIS) at the University of Tokyo (No.1234), utilizing the "Detailed Address Matching (Geocoding) Service" and data from the "Grid Square Statistics of 2016 Economic Census for Business Activity by Prefecture," provided by *Sinfonica* of Japan. This work was supported by JSPS KAKENHI Grants 22H00836 and 20K22103.

‡The classification of business attribute tags using community detection is based on the approach and results outlined in CREI Report No. 16.

1 Introduction

The influence of industrial variety, often referred to as diversity, on economic growth has been studies extensively since the pioneering studies of Jacobs (1969) and Glaeser et al. (1992). These studies introduced the concept that knowledge transfer and spillovers often occur across industries, driving innovation and growth. Building on these foundational ideas, Frenken et al. (2007) and subsequent researchers provided a comprehensive view by decomposing variety into unrelated and related types. The key argument is that knowledge spillovers are more likely to occur within related industries, as they share complementary skills and technologies, compared to unrelated industries.

After empirical research extensively examined how related and unrelated varieties affect regional growth (Content and Frenken, 2016), the emphasis shifted to how these varieties affect entrepreneurship, including general and innovative startups, to discover the mechanisms against regional growth. Although related variety has mainly been associated with knowledge spillovers in the context of regional growth, unrelated variety has also been linked to knowledge spillovers as a factor driving radical innovation in the context of entrepreneurship. Therefore, from a theoretical standpoint, both related and unrelated varieties are considered to enhance entrepreneurship. Although the results have not always been consistent across studies, this relationship has been confirmed by Bishop (2012) and Colombelli (2016).

While research on entrepreneurship is growing, prior studies have generally not limited the types of firms considered—such as those based on size or age—when calculating these two varieties. Consequently, the focus has predominantly been on knowledge spillovers across all types of firms, rather than specifically on those within and among incumbent startups. Considering the increasing volume of research in another line that explores the volume and mechanisms of knowledge spillovers occurring specifically within startup ecosystems (Sako et al., 2022; Roche et al., 2022; Atkin et al., 2022; Choi et al., 2024), a key question arises: how does the relatedness among startup businesses within each area influence the emergence of new startups? To the best of my knowledge, no research has specifically examined the impact of regional variety in terms of incumbent startups on new ventures, particularly by decomposing related and unrelated varieties. This paper aims to fill this gap, representing its main contribution to the literature.

To assess the role of related and unrelated varieties in the context of startups, this paper use business attribute tags from a unique database of growth-oriented startups in Japan. This database records the services (i.e., businesses) offered by each startup, with several tags assigned to each service to describe its features. These tags are utilized to quantify the scope of each startup's business activities and to measure diversity in startups, reflecting both diversity within individual startups and diversity among startups. By analyzing the co-occurrence networks of these tags, I categorize them into distinct business fields. Tags within the same business fields are considered "related," whereas those across different fields are considered "unrelated." Related and unrelated varieties for each geographical area are then quantified using tile index decomposition, following Frenken et al. (2007). I analyze the impact of these varieties on the number of newly established startups in each area using a negative binomial regression model. Since startups in Japan are substantially concentrated in Tokyo's 23 wards (i.e., central Tokyo), I define areas, or observation units, as 500-meter or 1,000-meter meshes, allowing for the use of variations across these meshes to assess the influence of diversity within central Tokyo.

The results indicate that both related and unrelated types positively influence startup creation in the 500-meter and 1,000-meter mesh sizes. This suggests that areas displaying diversity, both among startup business fields and within each field, provide a favorable environment for startup creation. At the 500-meter mesh level, the influence of related and unrelated varieties on startups is approximately balanced. However, at the broader 1,000-meter mesh, the impact of the unrelated variety diminishes, while the influence of the related variety grows stronger. This reduced relative importance of the unrelated variety in the broader mesh may be attributed to the need for closer face-to-face interactions to facilitate knowledge spillovers among unrelated startups that generally lack cognitive proximity. As geographic areas expand, the frequency of face-to-face communications declines, thereby weakening the positive effect of unrelated variety on startup creation.

The remainder of this paper is organized as follows. Section 2 provides a literature review, summarizes key findings, and identifies gaps in existing research regarding the role of diversity in entrepreneurship. Section 3 elucidates the data sources used in this study and

Section 4 discusses the specification model and variables. Section 4 details the methodology for constructing startup business fields and quantifying the diversity index using these fields. Section 5 presents the empirical analysis results, while Section 6 concludes the paper by summarizing the main findings and suggesting directions for future research.

2 Literature

The seminal work by Frenken et al. (2007) makes a significant contribution to the study of regional economic growth by introducing the concept of "relatedness" between industries. They propose using the entropy index's ability to decompose variety into "related" and "unrelated" types and investigate which one contributes more to regional economic development. They argue that knowledge spillovers are more likely to occur within related industries, while knowledge exchanges between unrelated sectors are limited. They argue that major innovations often arise from the recombination of knowledge across different industries, based on Jacobs's (1969) ideas, which more likely occur when firms are geographically close and share similar institutional settings. This concept ties into "cognitive proximity," which emphasizes that individuals or organizations need a shared knowledge base to effectively communicate, understand, and assimilate new information (Boschma, 2005). Alongside the benefits associated with the spatial proximity of related industries, Frenken et al. (2007) also highlight the risks that come with it. They point out that it can increase vulnerability owing to correlated demand shocks across those industries. They suggest that diversifying across unrelated sectors is preferable to minimize risks, following a portfolio strategy approach.

According to Content and Frenken (2016), many studies investigating the impact on city growth, particularly in terms of employment or productivity, have reported positive effects of related variety. In contrast, the findings for unrelated variety tend to be mostly insignificant or mixed. Several studies have focused on the heterogeneous effects of these varieties. For example, Cortinovis and van Oort (2015) found a positive effect of related variety in regions characterized by a high technological regime, and Hartog et al. (2012) determined that related variety is significantly positive only in high-tech sectors. Content and Frenken (2016) conclude that the positive effects of related variety on city growth may be limited to knowledge-intensive sectors. This suggests that the related variety may also play a crucial role in startups, particularly innovative ones, given their knowledge-intensive nature.

To understand how related and unrelated varieties drive urban growth, recent studies have focused on their effects on innovation outputs or entrepreneurship. Aligning with findings in city growth literature, positive impacts of related variety have been observed for both general and innovative startups (Bishop, 2012; Guo et al., 2016; Colombelli, 2016). However, research comparing effects of related variety across different types of startups has yielded mixed results. Content et al. (2019) found that related variety positively influences opportunity-driven entrepreneurship, which is thought to include innovative startups, but not necessity-driven entrepreneurs, who typically start businesses due to limited employment options. Conversely, Antonietti and Gambarotto (2020) found positive effects of related variety for other types of startups, but not innovative businesses. This result aligns with Boschma (2005), suggesting that excessive cognitive proximity can hinder learning and innovation due to the need for dissimilar and complementary bodies of knowledge in knowledge building, as well as the risks of cognitive lock-in and involuntary spillovers.

In terms of unrelated variety, the seminal work by Frenken et al. (2007), which focuses on city growth, initially associated knowledge spillovers only with related variety. However, recent research on innovation or entrepreneurship has extended this connection to unrelated variety, arguing that combining unrelated knowledge fosters radical innovations and technological breakthroughs. In fact, many studies have reported the positive effects of unrelated variety on both general and innovative startups (Bishop, 2012; Colombelli, 2016; Antonietti and Gambarotto, 2020), highlighting the role of diverse knowledge sources in creating new entrepreneurial opportunities. Other studies have suggested that unrelated variety contributes to radical innovation; for example, Castaldi et al. (2015) and Miguelez and Moreno (2018) determined that unrelated variety positively affects patent quality but not patent quantity. Similarly, Antonietti and Gambarotto (2020) observed a positive impact on the creation of innovative but not other types of startups. However, the impact of unrelated variety is not always consistent, as Guo et al. (2016) reported mixed results, and Content et al. (2019) found negative effects. These mixed findings suggest that further research is required to fully understand the effects of unrelated variety on entrepreneurship and innovation.

These studies on related and unrelated varieties, using entropy index decomposition, primarily rely on hierarchical classification systems such as industrial classifications because relatedness is determined by whether entities belong to the same category. Although some studies have focused on knowledge-sector industries or used patent classifications to capture the regional knowledge base (Bishop, 2012; Colombelli, 2016), related and unrelated varieties are still calculated without specifically narrowing down which firms within each classification are included or considering the institutional origin of the patents used. This approach suggests that the analysis captures knowledge spillovers not only among startups, but also across firms of varying ages and sizes. Since startups constitute only a small proportion of all firms, it is reasonable to conclude that these studies emphasize spillovers between startups and other firms, or even primarily among non-startup firms. To better understand how relatedness influences innovative activities and entrepreneurship through knowledge spillovers, narrowing the scope of knowledge diffusion to specific channels is essential.

Besides the literature on related and unrelated varieties, a growing body of research has focused on the narrow scope of how knowledge flows among startups or their members. This line of inquiry focuses on interactions among startup founders, employees, and networks, elucidating mechanisms that are more relevant to entrepreneurial ventures. For example, Sako et al. (2022) examined knowledge transfer among co-founders, founders' networks, and early employees to explore whether knowledge similarity within these groups fosters start-up growth. Their findings suggest that knowledge similarity plays a significant role in younger firms and nascent ecosystems, whereas knowledge diversity is more important for older firms and mature ecosystems. Although not exclusively focused on start-ups, Atkin et al. (2022) examined face-to-face interactions among patenting firms in Silicon Valley, revealing a strong positive relationship between these interactions and knowledge flow, with serendipitous encounters playing a crucial role. In addition to these studies, as discussed in detail in Section 3.2, Roche et al. (2022) have also highlighted the locality of knowledge spillovers among startups, and Choi et al. (2024) have shown the localized impact of face-to-face interactions on entrepreneurship. Collectively, these studies provide evidence of knowledge spillovers among startups and their members. These findings contrast with prior research on related and unrelated varieties, which did not specify the types of firms among which knowledge spillovers occur.

Beyond knowledge spillovers, relatedness among startups can influence entrepreneurship through other mechanisms. In tech clusters, where labor mobility is particularly high (Kerr and Robert-Nicoud, 2020), the concentration of related startups may benefit entrepreneurs by facilitating the search for co-founders or skilled team members, fostering the creation of new ventures. In fact, Kuusk (2021) highlighted that the role of local labor flows is underestimated in understanding the relationship between related variety and regional employment growth. These flows not only support growth by enabling knowledge transfer but also by improving the matching between employers and employees. Although my study does not specifically quantify the role of local labor flows in new startup formation, these dynamics may be indirectly captured in the observed effects.

3 Data

3.1 Startups information

This study uses a comprehensive database of startups in Japan named "STARTUP DB" provided by *for Startups, Inc.*. The database version used is from April 22, 2022, and includes startups that existed on or before that date. Previous research on entrepreneurship has often used firm-age to define new firms, typically considering a range of six to eight years (Cefis and Marsili, 2011). In line with this approach, this study defines companies listed in this database within six years of their founding as startups. The age of each startup is calculated based on the founding date recorded in the STARTUP DB.

As described on its website, the database focuses on "companies especially venture startups in growth industries."[1] Information in this database is also featured on Crunchbase through data integration.[2] Crunchbase is recognized as a database of innovative startups and companies and has been widely used in economics and management research (Dalle

[1] `https://startup-db.com/about` (Accessed on May 28, 2024)

[2] `https://www.forstartups.com/news/partnership-crunchbase` (Accessed on November 2, 2024)

et al., 2017).[3] It has emerged as a crucial information source for investors, providing detailed data on startup companies to assist potential investors in their evaluations (Dalle et al., 2017; Edwards and Todtenhaupt, 2020). Consequently, startups listed on Crunchbase and STARTUP DB are considered a select group among all new ventures, specifically representing growth-oriented businesses. Indeed, Crunchbase's coverage is relatively limited, with only 0.2% of all new European firms registered on the platform (Leendertse et al., 2022).

To track the location history of startups, the study uses the National Tax Agency's Corporate Number System Web API,[4] which was introduced in Japan in October 2015. Under this system, each organization is assigned a designated number, which is registered along with the organization's name and the address of its head office or principal place of business. This address is used as the office address of startups and tracks the change history of these addresses via the web API.[5][6]

To quantify diversity within and between startups, I use information about the services offered by each startup, rather than the industry classifications commonly used in previous research on related and unrelated varieties. This is because startups may engage in innovative business ventures that do not fit within existing industry categories or may span multiple categories. In fact, STARTUP DB does not contain information on industry classifications to which each startup belongs. Instead, the database provides information on the services

[3] Dalle et al. (2017) comprehensively outlined the scope, coverage, and various research uses of Crunchbase.

[4] The version of STARTUP DB used in this study contains only the address as of a single point in time.

[5] Since only a single address is registered for each organization, this system has a limitation in that it cannot capture all locations where business activities occur. However, this is not a significant problem in this study, because startups are relatively new and have few offices.

[6] Address information is geocoded using the following two-step approach: (1) The University of Tokyo CSV Address Matching Service, specifically the "Detailed Address Matching (Geocoding) Service," is employed. If the reliability of conversion, rated at the highest level of 5 (indicating minimal errors due to place name ambiguity), and the converted address level is "7: Block/lot number" or "8: Building number/branch number," the geocoding result is accepted. (2) Google Geocoding API was used for cases that did not meet these criteria. If the postal code from the geocoding result matches the registered postal code of the corporation, the result is adopted. Using this method, geocoding was performed for all the addresses of companies listed in STARTUP DB from October 1, 2015 (the date of the introduction of the Corporate Number System). Consequently, latitude and longitude coordinates were successfully identified for 99.5% of the addresses, of which 91.4% were determined using method (1).

offered by individual startups, with each service associated with some tags that describe its attributes. Service information is available for 76.9% of startups located in Tokyo's 23 wards as of 2016, with each startup offering an average of 2.99 services. On average, each service is associated with 3.31 tags.[7] The startup business fields are constructed as mutually exclusive groups of tags, where each tag belongs to only one field. These are then used to quantify diversity, as detailed in the next section.

Meanwhile, it is necessary to make several clarifications regarding the data; owing to data limitations, it was not possible to track changes in the service and tag information for each startup. Consequently, this analysis assumes that these details have remained unchanged throughout the year. Additionally, since the primary independent variable (i.e., total, related, and unrelated varieties) is calculated solely based on startups with tag information, these varieties represent the diversity within startups that have tags. To maintain consistency in the sample, only startups with tags were used when calculating the number of newly created and incumbent startups.

3.2 Geographic unit

This study focuses on Tokyo's 23 wards, and uses 500-meter and 1,000-meter meshes as observational units.[8] While previous research on related and unrelated varieties has employed regional units across the country,[9] I concentrated on finer spatial units within a specific region for two reasons.

First, knowledge spillovers among startups and the impact of face-to-face interactions on entrepreneurship reportedly occur at highly localized levels. For example, Roche et al. (2022) investigated knowledge spillovers in co-working hubs, finding that proximity—particularly within 20 meters—significantly enhances knowledge exchange through social interactions.

[7]Tags representing transaction types, such as B2B and B2C, were excluded from this calculation.

[8]The mesh within the 23 wards of Tokyo is identified, and location information is obtained using the *jpmesh* package in R. From these, we used only the meshes that fell within the municipal boundaries obtained from the Municipality Map Maker (`http://www.tkirimura.com/mmm/`).

[9]An exception is Xiong et al. (2023), who used census tracts in Salt Lake County as observation units but applied the Herfindahl-Hirschman Index, rather than the entropy index, to construct related and unrelated varieties.

Similarly, Choi et al. (2024) demonstrated that introducing Starbucks cafés into U.S. neighborhoods previously without coffee shops increased local entrepreneurship, likely by creating enhanced networking opportunities. However, this effect decreased rapidly with distance, reducing to one-fourth at 1–2 kilometers and one-tenth at 2–10 kilometers. Our study aims to capture these localized spillovers by utilizing mesh units that enable adjustments to spatial scale, thereby providing insights into how the effects vary with mesh size.

Second, startups in Japan are heavily agglomerated in Tokyo. As of 2022, 73% of all startups were concentrated in Tokyo Prefecture among the 47 prefectures.[10] At the municipal level (including cities, towns, villages, and the special wards of Tokyo), the top 10 municipalities accounted for 66% of all startups nationwide, out of a total of 1,724 municipalities. This high degree of spatial concentration complicates efforts to measure the effects of related and unrelated varieties using larger regional units, as in previous studies. Therefore, adopting 500-meter and 1,000-meter meshes as observational units allows us to capture the local variation necessary to estimate these effects. Given that 98% of startups in Tokyo are concentrated within the 23 wards, the 500-meter and 1,000-meter meshes were employed in these wards as observational units.

4 Analysis methods

4.1 Specification model

This study examines the impact of startup diversity on the creation of new startups by applying a cross-sectional negative binomial regression model, which is commonly used in studies analyzing the influence of related and unrelated varieties on entrepreneurship (Antonietti and Gambarotto, 2020; Xiong et al., 2023). This model was employed because the dependent variable is a count variable, and some observations have zero values. Since entrepreneurs make their location decisions considering the existing economic environment as exogenous, endogeneity is not a significant issue in the analysis of new business formation

[10] Here, we calculated the number of firms listed in STARTUP DB that were less than six years old. Their addresses as of January 1, 2022, were identified using the National Tax Agency's Corporate Number System Web API

(Rosenthal and Strange, 2003).[11]

The negative binomial regression model is specified as follows:

$$n_{m,2017-2019} = \beta Variety_{m,2016} + \gamma EcosystemActors_{m,2016} + f_j + \epsilon_{m,2017-2019}, \tag{1}$$

where the dependent variable $n_{m,2017-2019}$ is the number of newly established startups at mesh m from 2017 to 2019. $Variety_{m,2016}$ is the vector of variables related to the degree of diversity of startups within each mesh m: total, related, or unrelated varieties. $EcosystemActors_{m,2016}$ is the vector of variables that reflect the number of actors in the startup ecosystem in each mesh m: number of incumbent startups, number of Venture capitals (VCs) and corporate venture capitals (CVCs), academic research intensity, and number of large corporations. Meanwhile, f_j represents the fixed effects of municipality j. The sample is restricted to observations where total, unrelated, or related varieties in startups are available—that is, those with at least one incumbent startup present in 2016. A detailed description of each variable is provided below, with the exception of $Variety_{m,2016}$, which is described in the following subsection. Summary statistics for these variables are presented in Table 1, and heat maps of variables relating to startups at the 1,000-meter mesh level are presented in Figure 1.

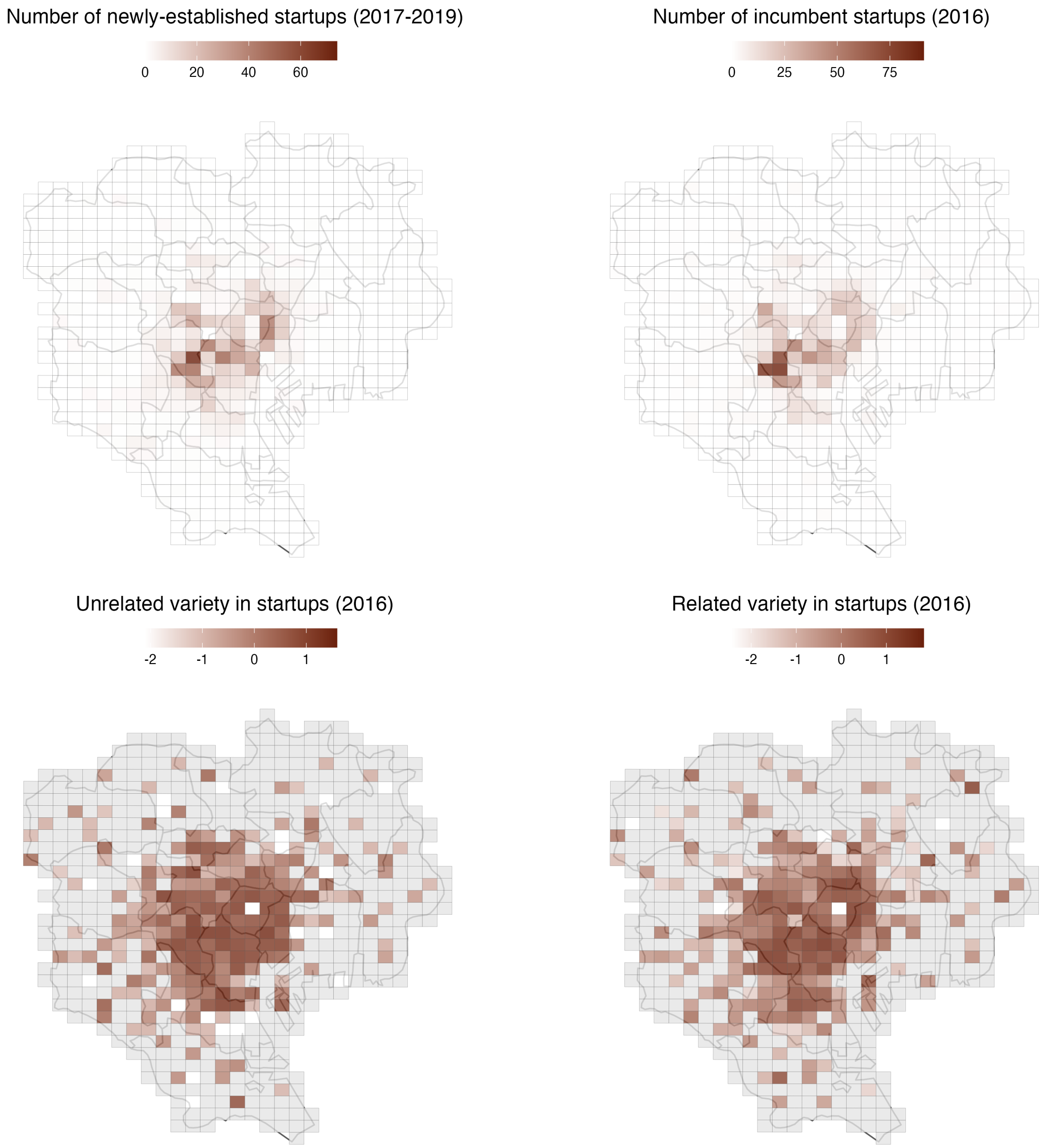
[11] While endogeneity is not a significant concern, panel data analysis could be employed to further address this concern if necessary. However, the available startup location data spans only a four-year period, from 2016 to 2019, due to the timing of the Corporate Number System's introduction and the start of the COVID-19 pandemic. Given this limited period and the potential need for a lag between the independent and dependent variables, the data lacks sufficient temporal depth to conduct panel data analysis.

Table 1: Summary Statistics

Mesh size	500-meter			1,000-meter		
Variable	N	Mean	SD	N	Mean	SD
2017–2019						
Number of newly-established startups	449	2.9	4.8	238	5.9	11
Number of newly-established startups (only non-subsidiary)	449	2.5	4.1	238	5.2	9.2
2016						
Total variety in startups $_*$	449	3.1	1.1	238	3.3	1.3
Total variety in industries $_*$	449	4.6	0.39	238	4.8	0.33
Unrelated variety in startups $_*$	449	1.4	0.72	238	1.5	0.72
Unrelated variety in industries $_*$	449	3.1	0.23	238	3.2	0.21
Related variety in startups $_*$	449	1.7	0.62	238	1.8	0.73
Related variety in industries $_*$	449	1.5	0.27	238	1.6	0.24
Number of incumbent startups	449	4.7	7.4	238	9.1	17
Number of VCs & CVCs	449	0.19	0.59	238	0.37	1.2
Academic research intensity	449	0.2	1.8	238	0.43	2.5
Number of large corporations	449	4.4	6.9	238	9.7	18

Notes: In the main regression analysis, we use only observations where total, unrelated, or related varieties in startups are not missing (i.e., observations with at least one incumbent startup in 2016). Therefore, this table presents the summary statistics for these observations. The total number of 1,000-meter and 500-meter meshes in the 23 wards of Tokyo, identified using the R package *jpmesh* and included within the municipal boundaries obtained from the Municipality Map Maker (`http://www.tkirimura.com/mmm/`), were 685 and 2,580, respectively. In the regression analysis, total, unrelated, and related varieties are scaled to have a mean of 0 and standard deviation of 1; however, here we present the values before scaling, marked with $*$.

Figure 1: Heatmaps of variables related to startups



Notes: This heat map is based on a 1,000-meter mesh size. In the maps showing related and unrelated varieties in startups, meshes colored grey indicate missing data owing to the absence of startups. Lines in the background represent the boundaries of Tokyo's 23 wards. The number of newly-established and incumbent startups is calculated using our targeted sample of startups with business attribute tags in STARTUP DB, focusing specifically on incumbents within six years of their founding. Related and unrelated varieties are scaled to have a mean of zero and standard deviation of one.

The dependent variable, $n_{m,2017-2019}$, represents the number of newly-established startups in each mesh between January 1, 2017, and December 31, 2019. Specifically, the number of startups whose corporate numbers were newly registered during this period were count. To avoid the simultaneity problem, the observation period for the dependent variable began one year after the observation year 2016 for the independent variables. The year 2016 was selected because the corporate number system that is used to construct several independent variables was introduced in late 2015. The observation period for the dependent variable ends in 2019 to avoid potential disruptions in office space demand owing to the outbreak of COVID-19 pandemic in 2020. As illustrated in Figure 1, the number of newly-established startups tends to be higher in central Tokyo's 23 wards, similar to the distribution of incumbent startups in 2016.

In addition to the primary independent variable of interest, $Variety_{m,2016}$, variables related to startup ecosystems (or entrepreneurial ecosystems), denoted as $EcosystemActors_{m,2016}$, were also included as independent variables, as startup ecosystems are recognized as important factors influencing the creation and growth of startups. The entrepreneurial ecosystem is "a set of interdependent actors and factors coordinated in such a way that they enable productive entrepreneurship," as defined by Stam (2015, p.1765). As reviewed by Cavallo et al. (2021), recent studies on entrepreneurial ecosystems have increasingly examined the relationships between new ventures and key ecosystem actors, particularly incubators, VC firms, universities, and large corporations, in fostering the creation and growth of these new ventures. To align with this focus, independent variables were include that quantify the presence of each type of actor within mesh m, incorporating incumbent startups as another ecosystem actor.

The first actors to focus on entrepreneurial ecosystems were incumbent startups, aiming to capture the effect of their agglomeration. The mechanisms that generate agglomeration benefits have been examined since Marshall's (1920) early work, which pointed to input sharing, labor market pooling, and knowledge spillovers, while more recent work by Duranton and Puga (2004) categorized these mechanisms into sharing, learning, and matching, providing micro-foundations for them. Although these concepts refer to general economic activities, they can also be applied to the agglomeration of innovation activities, including entrepreneurship (Carlino and Kerr, 2015; Mathias et al., 2021). To capture the effect of agglomeration of incumbent startups, I include the number of incumbent startups within each mesh as of January 1, 2016, as an independent variable,[12] which can also be considered a proxy for whether there is a suitable local environment for startups. It considers factors such as the presence of incubator facilities, which are not included as separate independent variables because of data limitations, although they are another key actor in entrepreneurial ecosystems.

The second actor in entrepreneurial ecosystems is VC investors, who play a critical role in financing entrepreneurial firms and contribute to the development of entrepreneurial teams through coaching and monitoring (Gompers and Lerner, 2001; Colombo et al., 2019). They tend to invest in geographically-close entrepreneurs, as this proximity allows for frequent face-to-face communications with their portfolio companies, reducing risks such as adverse selection and moral hazard, making coaching more effective (Colombo et al., 2019; Huang et al., 2023; Sorenson and Stuart, 2001). Therefore, it is reasonable to assume that startups seeking funding from VC firms have incentives to locate near them and that their location influences in which locality the startups are established. To capture this effect, we include the number of VC firms and CVCs within each mesh as independent variables. Specifically, we focus on the number of VCs and CVCs who are members of the Japan Venture Capital Association.[13] Similar to the method used for startups, the locations of these VCs and CVCs were identified using the Corporate Number System as of January 1, 2016.

The third actor in entrepreneurial ecosystems includes universities. University research often provides opportunities for entrepreneurship, often resulting in "spin-offs" led by faculty, staff, or graduates (Åstebro et al., 2012). These university spin-off firms are usually located close to their parent universities (Heblich and Slavtchev, 2014), and it has also been shown that university knowledge positively affects entrepreneurship in nearby area (Bonaccorsi et al., 2014; Ghio et al., 2016). To capture this effect, I include academic intensity as an

[12] Since we define firms listed in STARTUP DB within six years of their founding as the startups of our focus, this independent variable represents the number of firms in this database founded between January 1, 2010, and January 1, 2016.

[13] The list of VCs and CVCs that are members of the Japan Venture Capital Association is provided by the Real Estate Companies Association of Japan.

independent variable, measured by total research grants (per 100 million yen) allocated to universities within each mesh. For research grants, I use the amounts allocated to each university as of 2016 under the Grants-in-Aid for Scientific Research program, a national research funding initiative in Japan.[14] The locations of universities are obtained from the School Point Data provided by *Zenrin Marketing Solutions Co., Ltd.*, using the representative longitude and latitude of each university campus recorded in this dataset. If a university has multiple campuses within Tokyo's 23 wards, grant amounts from the university are counted for multiple meshes.

The last actor in our focus on entrepreneurial ecosystems is large corporations. As reviewed by Bhawe and Zahra (2019), multinational enterprises (MNEs) affect local entrepreneurship through knowledge spillovers in multiple ways, including technology, marketing, operations, and managerial skills, and by promoting co-specialization through licensing, selling, or entering alliances. To capture the influence of large corporations within each mesh, the number of large establishments across all industries is included as an independent variable, instead of the number of large corporations due to data availability. Here, large establishments are defined as those with over 300 employees. The data is sourced from the 2016 Economic Census for Business Activity conducted by the Ministry of Internal Affairs and Communications.[15]

Lastly, f_j represents the fixed effects of municipality j. Each municipality (specifically, each of the 23 wards of Tokyo) offers different programs, support, or subsidies to encourage entrepreneurship.[16] In addition, business activities tend to be more concentrated in the central wards within Tokyo's 23 wards, reflecting a more favorable environment for business activities in these areas. Therefore, these municipality fixed effects capture the effects of differences in municipal policies or those resulting from location differences within Tokyo's 23 wards. In cases where a single mesh contains multiple municipalities, the municipality

[14] We use data on both newly-allocated and continuously-allocated grant amounts in 2016, summing them for each university.

[15] The survey date for this study is June 1, 2016.

[16] The following page (written in Japanese) introduces projects that support startups and entrepreneurship in each municipality of Tokyo: `https://www.tokyo-sogyo-net.metro.tokyo.lg.jp/shien_prg/municipal/`

with the largest area is assigned to that mesh.[17]

4.2 Related and unrelated varieties in startup businesses

The main independent variables are indexes representing the diversity of businesses within and among incumbent startups in each mesh. Following previous literature, I primarily use the entropy measure, which can be decomposed into related and unrelated varieties, typically using a classification system with at least two levels, such as an industry classification. For example, the seminal work by Frenken et al. (2007) utilized the 2-digit (broader level) and 5-digit (finer level) Standard Industrial Classification codes, assuming that 5-digit classes across 2-digit classes are unrelated, while those within the same 2-digit class are related. Subsequently, related variety is calculated as the weighted average of the entropy index in 5-digit industries within each 2-digit class, while unrelated variety is the entropy index across 2-digit classes.[18]

Instead of using industry classifications, business attribute tags registered in STARTUP DB are used to compute the entropy index. In particular, business attribute tags are used at a finer level of classification, whereas our originally constructed tag groups (i.e., business fields) are used at a broader level of classification. Therefore, tags within the same tag group are considered related, while those from different tag groups are unrelated (or less related).

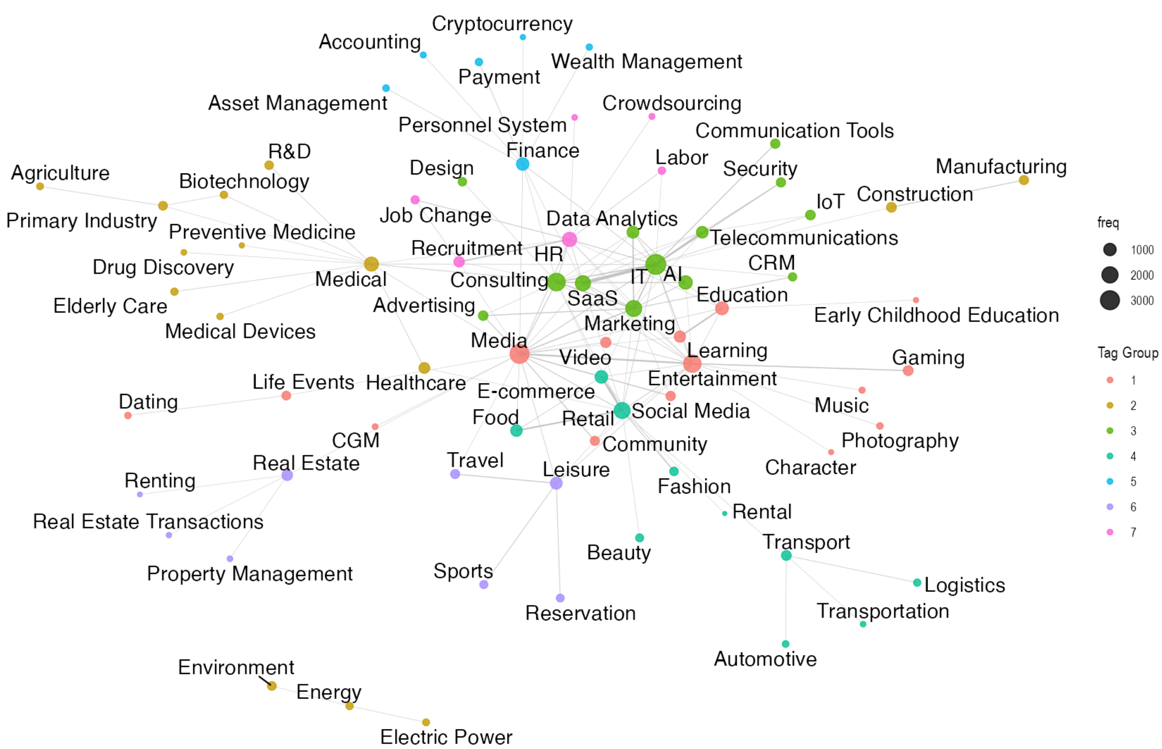
The tag groups are calculated using a community detection method in co-occurrence network analysis. First, a co-occurrence network is constructed by calculating the co-occurrence frequency of each tag pair (i.e., the number of services in which the tag pair is registered together). The co-occurrence network is illustrated in Figure 2. Subsequently, the modularity maximization of Newman and Girvan (2004) was applied for this co-occurrence network to group tags. This method, employed in prior studies on startup community detection (Basole et al., 2019), has the advantage that the number of groups can be determined through analysis, eliminating the need for the analyst to predefine it. As shown in Table 2, seven

[17] Municipal boundaries are obtained from Municipality Map Maker `http://www.tkirimura.com/mmm/`.

[18] In Frenken et al. (2007), the entropy index is calculated using employment in each industry class. However, because it is not possible to identify annual employment data for each startup, the number of startups was used instead.

tag groups are identified, and each group is named after the top two tags with the highest single occurrence in each group: (1) Media and Entertainment, (2) Medical and Healthcare, (3) IT and Consulting, (4) Retail and E-commerce, (5) Finance and Payments, (6) Leisure and Real Estate, and (7) HR and Recruitment. The details of this community detection methodology are described in Appendix A.

Figure 2: Co-occurrence Network of Tags



Note: This network is constructed using startups founded in or after 2010, and the graph displays only the edges where the co-occurrence frequency is 100 or higher. This paper considers that the tag groups represent startup business fields.

Table 2: Overview of Tag Groups

Tag Group (referring to Business fields)	Number of Startups	Top 10 Tags Based on Single Occurrence
1 Media & Entertainment	2987	Media, Entertainment, Education, Learning, Video, Gaming, Social Media, Community, Life Events, Virtual Reality
2 Medical & Healthcare	1809	Medical, Healthcare, Construction, Manufacturing, Environment, Primary Industry, R&D, Robotics, Energy, Biotechnology
3 IT & Consulting	3514	IT, Consulting, Marketing, SaaS, AI, Data Analytics, Telecommunications, IoT, Advertising, Communication Tools
4 Retail & E-commerce	1710	Retail, E-commerce, Food, Transport, Fashion, Beauty, Sharing Economy, Logistics, Automotive, Drone
5 Finance & Payment	743	Finance, Payment, Blockchain, Asset Management, Wealth Management, Government, Accounting, Cryptocurrency, Law, Lending
6 Leisure & Real Estate	870	Leisure, Real Estate, Travel, Sports, Reservation, Property Management, Real Estate Transactions, Renting, Renovation
7 HR & Recruitment	882	HR, Recruitment, Job Change, Labor, Crowdsourcing, Personnel System, Side Jobs

Notes: The number of startups is calculated based on startups listed in STARTUP DB that were established nationwide in or after 2016. This paper considers that the tag groups represent startup business fields.

Before calculating the entropy index, the tags associated with the services (i.e., businesses) provided by each startup are first identified, and then the tags linked to each startup are determined. The entropy index calculated in this study is as follows: Business attribute tags (referred to as tag) t belong to tag group g. P_t denotes the share of startups with tag t in total startups, and P_g denotes the share of startups with tag group g in total startups. Considering that startups frequently have multiple tags t, the count for each startup for each tag is weighted as one divided by the total number of unique tags.[19]

The entropy index before decomposition, called total variety (TV), is given by

$$TV = \sum_{t} P_t \log_2 \left(\frac{1}{P_t} \right) \tag{2}$$

Related variety (RV), which is the weighted average of the entropy index within each tag group, is given by

$$RV = \sum_{g} P_g H_g \tag{3}$$

where:

$$H_g = \sum_{t \in g} \frac{P_t}{P_g} \log_2 \left(\frac{1}{P_t / P_g} \right) \tag{4}$$

The unrelated variety (UV), which is the entropy index at tag group level, is given by

$$UV = \sum_{g} P_g \log_2 \left(\frac{1}{P_g} \right) \tag{5}$$

As first explained by Theil (1972), this total variety (TV) can be decomposed into an unrelated variety (UV) and a related variety (RV). These varieties reflect the degree of diversity in businesses, capturing both diversity within incumbent startups and diversity among startups. Specifically, TV is equal to the sum of RV and UV. These variety indexes are scaled to have a mean of zero and standard deviation of one to allow for comparison across various models.

[19] Overlapping tags, which may occur when a startup offers multiple services, are treated as a single occurrence to capture the overall scope of its businesses.

5 Results

5.1 Baseline results and robustness checks

Table 3 presents the results using total variety in startups, while Table 4 decomposes it into unrelated and related varieties in startups. Each column with the same number corresponds to the results obtained from the same model specifications. Column (1)–(3) displays the results at the 500-meter mesh level, and Columns (4)–-(6) the results at the 1,000-meter mesh level. Columns (1) and (4) in each table illustrate the baseline results and Columns (2)–-(3) and (5)–-(6) show the results of robustness checks. Since we include municipality fixed effects, if all outcomes are zero in a municipality, observations from that municipality are excluded. The number of excluded municipalities is reported in the footnotes of each results table.

First, we focus on the baseline results shown in Column (1) for the 500-meter mesh and in Column (4) for the 1,000-meter mesh in each table. Table 3 shows that total variety has a significant positive effect for both mesh sizes, with a higher coefficient observed for the larger mesh size. This finding suggests that areas with greater diversity among startup businesses tend to experience a higher rate of new business formation over the succeeding three years. Moreover, the effect of a one standard deviation increase in diversity is found to be larger for broader meshes. When this diversity index is divided into related and unrelated varieties, as shown in Table 4, we observe that both types are significantly positive for both mesh sizes. These positive effects align with existing literature on entrepreneurship, which suggests that both related and unrelated varieties enhance entrepreneurship: related variety promotes knowledge spillovers through cognitive proximity, while unrelated variety fosters radical innovation by integrating distinct types of knowledge.

Notably, the relative strength of effects between related and unrelated varieties varies with mesh size. In the 500-meter mesh in Column (1), the coefficients for unrelated and related varieties are almost equal, with the unrelated variety being slightly higher. However, in the larger 1,000-meter mesh size in Column (4), the coefficient for the related variety increases, while that for unrelated variety decreases relative to Column (1). At this broader regional scale, the coefficient for related variety was approximately 2.7 times larger than that for unrelated variety. The significance level of related variety remains at 1%, while that of the unrelated variety decreases to 5%.

Therefore, on a broader regional scale, the influence of unrelated variety diminishes, while related variety has a stronger impact on startup creation. These results align with Roche et al. (2022), who argue that physical proximity is less important for promoting knowledge exchange among similar startups but becomes more crucial for dissimilar startups. This also aligns with Boschma's (2005) view, which suggests that cognitive proximity can substitute geographical proximity. As geographical proximity weakens, higher relatedness needs to be compensated. Therefore, at a broader mesh level, it can be inferred that the relative effect of unrelated variety diminished.

The robustness checks are presented in Columns (2), (3), (5), and (6). The first robustness check in Columns (2) and (5) addresses the ownership status of new startups. In the baseline results in Columns (1) and (4), all new startups are included in the outcome variables regardless of subsidiary status. However, the location of subsidiary startups is likely influenced by the location of their parent firms, suggesting that regional diversity in startup businesses may have a lower influence on the creation of subsidiary startups. Therefore, including newly-established subsidiary startups in the outcome may underestimate the effect of variety on startup creation. Our database categorizes startups as (i) non-listed, (ii) non-listed and non-subsidiary, or (iii) listed. Columns (2) and (5) re-calculate the dependent variable using only (ii) non-listed and non-subsidiary startups. This adjustment does not affect the significance or magnitude of the coefficients for related or unrelated varieties. One point to note is that, as these statuses do not necessarily reflect the founding period, the number may not accurately capture newly-established subsidiary startups.

The second robustness check in Columns (3) and (6) relates to the potential mechanisms of varieties. The dataset in the baseline results includes meshes that contain only one incumbent startup, meaning that the variety indices for these meshes reflect the business diversity within that single startup. Even with only one incumbent startup, knowledge spillovers could still occur among its members and between them and potential entrepreneurs. Additionally, the presence of a single incumbent startup may contribute to labor mobility in the surrounding area, especially if the startup provides multiple services and therefore likely employs a larger workforce. This labor mobility could not only create opportunities to recruit new members with startup-specific skills but also facilitate knowledge spillovers through the transfer of specialized skills and expertise. However, to narrow the focus on potential mechanisms of varieties to interactions among multiple startups, the observation units are limited to meshes containing more than two incumbent startups. This adjustment did not substantially change the coefficient size for either mesh size. The estimated impact of a one standard deviation increase in these varieties, calculated as the exponentiated coefficient, decreases by approximately 3–5%, with the exception of the unrelated variety in the 1,000-meter mesh analysis, which increases by 4.5%. The significance levels remain stable in the 1,000-meter mesh analysis but drop to 5% in the 500-meter mesh analysis. Therefore, while limiting mechanisms to interactions between startups slightly reduces the effects of both unrelated and related varieties, focusing exclusively on interactions between multiple startups, or not, does not alter the conclusions.

The findings on variables related to actors in startup ecosystems are outlined in the rest of this section. Across various mesh sizes and specification models, the number of incumbent startups shows a significantly positive effect at the 1% level, whereas academic research intensity has no statistically significant effect. This finding suggests that future startup creation is influenced by the volume of incumbent startups, whereas the presence of research universities has no such impact in Central Tokyo. Alternatively, the effect of research universities may be captured by municipality-fixed effects.

Other actors within the startup ecosystem—VCs, CVCs, and large corporations—positively influence startup creation, but their effects are significant only within the 500-meter mesh, and significance varies with the specification. In the baseline model in Column (1) of Tables 3 and 4, the number of VCs and CVCs is weakly significant at the 10% level, but this effect becomes insignificant in Column (2), indicating that VCs and CVCs have a limited impact on non-subsidiary startup creation. However, in Column (3), the significance level increases to 5%, indicating that, in areas where there is already a concentration of incumbent startups, the presence of VCs and CVCs positively influences future startup creation. For large corporations, the baseline in Column (1) shows 5% significance, which decreases to 10% when the focus is on non-subsidiary startup creation, likely attributed to the exclusion of spin-offs from large corporations. This effect becomes insignificant in Column (3), indicating that large corporations have a lower influence on future startup creation in areas with a higher number of incumbent startups.

Table 3: Regression Results for Total Variety in Startups

	500-meter mesh			1,000-meter mesh		
	(1)	(2)	(3)	(4)	(5)	(6)
Total variety in startups	0.447***	0.444***	0.367***	0.573***	0.600***	0.579***
	(0.081)	(0.085)	(0.124)	(0.092)	(0.097)	(0.129)
Number of incumbent startups	0.042***	0.042***	0.040***	0.024***	0.023***	0.023***
	(0.011)	(0.012)	(0.011)	(0.004)	(0.005)	(0.005)
Number of VCs & CVCs	0.142*	0.119	0.159**	0.026	0.029	0.042
	(0.077)	(0.081)	(0.074)	(0.045)	(0.047)	(0.043)
Academic research intensity	0.006	0.011	0.012	0.001	0.001	-0.004
	(0.026)	(0.027)	(0.029)	(0.017)	(0.018)	(0.017)
Number of large corporations	0.018**	0.016*	0.011	0.006	0.005	0.004
	(0.008)	(0.009)	(0.008)	(0.004)	(0.004)	(0.004)
Fixed effects of Municipality	Yes	Yes	Yes	Yes	Yes	Yes
Non-subsidiary		Yes			Yes	
Multiple incumbents			Yes			Yes
Num.Obs.	408	408	219	238	234	126
Log.Lik.	-752.977	-724.500	-536.376	-452.461	-441.052	-334.816

Notes: Standard errors are presented within parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$. If *Non-subsidiary* is Yes, the dependent variable is calculated using startups registered as non-listed and non-subsidiaries in the STARTUP DB. If *Multiple incumbents* are Yes, meshes with two or more incumbents are used for the observations. Observations from municipalities with all-zero outcomes were excluded because of municipality-fixed effects. The numbers of excluded municipalities in each column are as follows: (1) 5, (2) 5, (3) 5, (4) 0, (5) 1, and (6) 4.

Table 4: Regression Results for Unrelated and Related Varieties in Startups

	500-meter mesh			1,000-meter mesh		
	(1)	(2)	(3)	(4)	(5)	(6)
Unrelated variety in startups	0.284***	0.282***	0.242**	0.183**	0.208**	0.227*
	(0.072)	(0.075)	(0.116)	(0.088)	(0.092)	(0.117)
Related variety in startups	0.258***	0.257***	0.202**	0.491***	0.495***	0.446***
	(0.075)	(0.078)	(0.094)	(0.094)	(0.098)	(0.123)
Number of incumbent startups	0.042***	0.042***	0.040***	0.023***	0.022***	0.022***
	(0.011)	(0.012)	(0.011)	(0.004)	(0.005)	(0.004)
Number of VCs & CVCs	0.143*	0.119	0.159**	0.025	0.028	0.040
	(0.077)	(0.081)	(0.074)	(0.043)	(0.045)	(0.042)
Academic research intensity	0.006	0.011	0.012	0.001	0.001	-0.004
	(0.026)	(0.027)	(0.029)	(0.017)	(0.017)	(0.017)
Number of large corporations	0.018**	0.015*	0.011	0.004	0.004	0.004
	(0.008)	(0.009)	(0.008)	(0.004)	(0.004)	(0.004)
Fixed effects of Municipality	Yes	Yes	Yes	Yes	Yes	Yes
Non-subsidiary		Yes			Yes	
Multiple incumbents			Yes			Yes
Num.Obs.	408	408	219	238	234	126
Log.Lik.	-752.966	-724.490	-536.376	-450.177	-439.258	-334.087

Notes: Standard errors are presented within parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$. If *Non-subsidiary* is Yes, the dependent variable is calculated using startups registered as non-listed and non-subsidiaries in the STARTUP DB. If *Multiple incumbents* are Yes, meshes with two or more incumbents are used for the observations. Observations from municipalities with all-zero outcomes were excluded because of municipality-fixed effects. The numbers of excluded municipalities in each column are as follows: (1) 5, (2) 5, (3) 5, (4) 0, (5) 1, and (6) 4.

5.2 Comparison of diversity indices

Our analysis suggests that both related and unrelated varieties in startups encourage the creation of new startups, but the effect of the unrelated variety diminishes as the geographic area expands. In this subsection, these findings are compared with the results obtained using other diversity indices, specifically industry variety, which is more commonly referenced in the literature. Table 5 presents the results, with Columns (1) through (4) showing the 500-meter mesh results and Columns (5) through (8) illustrating the 1,000-meter mesh. Specifically, Columns (1) and (3) in Table 5 align with Columns (1) and (4) in Table 3, respectively, whereas Columns (5) and (7) correspond to Columns (1) and (4) in Table 4.

Total, unrelated, and related varieties in terms of industry is calculated following methodologies established in previous literature. The data used here is sourced from the 2016 Economic Census for Business Activity conducted by the Ministry of Internal Affairs and Communications in Japan. This dataset provides the number of establishments within each mesh, categorized by industry "division" and "major group." Major group comprises a detailed category within this division. It can be assumed that major groups within the same division are related, while major groups between different divisions are unrelated. Accordingly, total variety is measured by the entropy index at the major group level, related variety is the weighted average of the entropy index within each division, and unrelated variety is the entropy index at the division level.[20] To be clear, these industry varieties capture the degree of diversity among all firms and are not limited to startups.

Table 5 demonstrates that total variety in industries is negative but insignificant in both Column (2) for the 500-meter mesh and Column (6) for the 1,000-meter mesh. This finding suggests that industry diversity, which is often emphasized in literature, may not significantly influence startup creation in this analysis. When decomposing this total variety in the 500-meter mesh in Column (4), unrelated variety in industries is significantly negative at the 5% significance level, indicating that broader diversity across industry divisions hinders new startup entries. In contrast, related industry variety is positive but statistically insignificant, and this combination of opposing coefficient signs in related and unrelated varieties leads to the overall insignificance of the total variety in Column (2). In the 1,000-meter mesh in Column (8), the signs for unrelated and related varieties in industries reverse compared to the 500-meter mesh in Column (4). Unrelated variety now presents a positive coefficient, while the related variety shows a negative coefficient. However, both coefficients are small and

[20] In the 2016 Economic Census for Business Activity, there are 18 divisions and 95 major groups. However, since the number of establishments within each mesh for major groups under divisions A "Agriculture and Forestry," B "Fisheries," and C "Mining and Quarrying of Stone and Gravel" are not disclosed, we only use divisions D through R, totaling 15.

statistically insignificant, suggesting that even with decomposed variety, industry diversity does not impact new startup entries in this broader area.

Table 5: Regression Results Comparing Variety in Startups versus Industries

	500-meter mesh				1,000-meter mesh			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total variety in startups	0.447***	0.449***			0.573***	0.572***		
	(0.081)	(0.081)			(0.092)	(0.092)		
Total variety in industries		-0.038				-0.026		
		(0.059)				(0.064)		
Unrelated variety in startups			0.284***	0.273***			0.183**	0.183**
			(0.072)	(0.071)			(0.088)	(0.088)
Related variety in startups			0.258***	0.245***			0.491***	0.498***
			(0.075)	(0.074)			(0.094)	(0.095)
Unrelated variety in industries				-0.122**				0.015
				(0.061)				(0.069)
Related variety in industries				0.092				-0.062
				(0.072)				(0.079)
Number of incumbent startups	0.042***	0.042***	0.042***	0.043***	0.024***	0.024***	0.023***	0.023***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.004)	(0.004)	(0.004)	(0.004)
Number of VCs & CVCs	0.142*	0.138*	0.143*	0.149*	0.026	0.025	0.025	0.020
	(0.077)	(0.077)	(0.077)	(0.076)	(0.045)	(0.045)	(0.043)	(0.044)
Academic research intensity	0.006	0.006	0.006	0.010	0.001	0.001	0.001	-0.001
	(0.026)	(0.026)	(0.026)	(0.026)	(0.017)	(0.017)	(0.017)	(0.017)
Number of large corporations	0.018**	0.019**	0.018**	0.017**	0.006	0.006	0.004	0.005
	(0.008)	(0.008)	(0.008)	(0.008)	(0.004)	(0.004)	(0.004)	(0.004)
Fixed effects of Municipality	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num.Obs.	408	408	408	408	238	238	238	238
Log.Lik.	-752.977	-752.755	-752.966	-750.651	-452.461	-452.371	-450.177	-449.822

Notes: Standard errors are presented within parentheses. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$. Columns (1) and (3) corresponds to Columns (1) and (4) of Table 3, respectively, and Columns (5) and (7) correspond to Columns (1) and (4) of Table 4. Observations from municipalities with all-zero outcomes were excluded because of municipality-fixed effects. The number of excluded municipalities for each column is as follows: (1)–(4) 5, and (5)–(8) 0.

6 Conclusion

Are new startups more likely to emerge in areas with a diverse mix of existing startup businesses? This study investigates this question while seeking to clarify which type of diversity exerts a greater influence. Recent studies on the impact of diversity on regional growth, including entrepreneurship, increasingly emphasize distinguishing between types of diversity, specifically related and unrelated varieties(Frenken et al., 2007). While prior research calculated these varieties using data on economic activities that were not startup specific, this study uses a unique dataset of Japanese startups to quantify related and unrelated varieties, specifically in the context of businesses offered by startups. Therefore, this study is the first attempt to capture the effect of diversity in terms of existing startups on new startup formation. Rather than conventional industry classifications, this dataset includes business attribute tags that characterize each startup's services. First, business attribute tags are classified into several tag groups representing business fields. This classification is then used to determine whether tags are related or unrelated, allowing for the calculation of both related and unrelated varieties. Subsequently, I examine the impact of these measures on the number of newly-established startups within each 500-meter or 1,000-meter mesh across Tokyo's 23 wards.

Both related and unrelated varieties of startup businesses were found to encourage startup creation regardless of the size of geographic unit. This indicates that diversity across different startup business fields and within each field contributes to promoting startup creation. Our findings are consistent with existing literature on entrepreneurship, which suggests that both related and unrelated varieties enhance entrepreneurship. Specifically, related variety is considered to promote knowledge spillovers through cognitive proximity, while unrelated variety fosters radical innovation by integrating diverse, unrelated knowledge.

In the 500-meter mesh analysis, the effects of related and unrelated varieties are nearly equal. However, as the geographic units increase to 1,000 meters, the impact of unrelated varieties diminishes, while that of related varieties increases. This reduction in the relative importance of unrelated varieties could be explained by the communication costs between startups. Among unrelated startups, which typically lack shared knowledge bases, more intensive face-to-face communications are required to generate sufficient knowledge spillover. In larger geographic areas, opportunities for direct interaction are less frequent, thereby reducing the positive impact of unrelated varieties. Boschma (2005) and Roche et al. (2022) also noted this substitution effect between geographical and cognitive proximity. This study offers additional empirical support for this perspective.

I also compared the results when including unrelated and related varieties based on industry classifications, as used in most previous literature, but found no significant effects for these conventional varieties, except for the unrelated variety in the 500-meter mesh. These variables are constructed using all establishments belonging to each industry. Therefore, these results suggest that, among various types of diversity, diversity among startups—especially those in similar positions to potential entrepreneurs—is most important for fostering new startup creation.

Although this study examines the impact of diversity on the formation of new startups, it is equally important to consider how diversity affects the growth of these startups. Future research should investigate the points at which startup lifecycle diversity has the greatest impact, providing deeper insights into the role of diversity in fostering entrepreneurship throughout the different stages of development. Additionally, this study examines relatedness within incumbents, but does not account for business relatedness between entrants and incumbents. Understanding the effect of relatedness is crucial, as it may influence the ease with which new startups access resources, knowledge, and networks established by existing firms. Future research should explore the degree to which business relatedness affects startup success and how it interacts with diversity to shape entrepreneurial outcomes.

Acknowledgements

We would like to thank Keisuke Kawata (UTokyo), Masatoshi Kato (Kwansei Gakuin University), and Kenta Ikeuchi (RIETI) for their valuable feedback on this paper, as well as the CREI members—Sachio Muto, Yosuke Nagase, and Daisuke Hasegawa—who frequently provided valuable feedback on this research. We also extend our gratitude to the Real Estate Companies Association of Japan and members of their Roundtable for their insightful discussions. Additionally, we would like to express our gratitude to the following participants for their valuable comments, including those provided on earlier versions of this study: the Entrepreneurship Workshop, the Sustainable Infrastructure Workshop, the CREI International Forum, the Japan Society of Real Estate Financial Engineering, the 37th Annual

Conference of the Applied Regional Science Conference, UTokyo Urban Economics Workshop, Chuo University Institute of Business Research Workshop, the Osaka Metropolitan University Economics Research Workshop, and the Hitotsubashi University Environment and Technology Workshop.

References

Antonietti, Roberto and Francesca Gambarotto (2020) "The role of industry variety in the creation of innovative start-ups in Italy," *Small Business Economics*, 54, 561–573.

Atkin, David, M Keith Chen, and Anton Popov (2022) "The returns to face-to-face interactions: Knowledge spillovers in silicon valley," NBER Working Paper Series No.30147.

Basole, Rahul C, Hyunwoo Park, and Raul O Chao (2019) "Visual Analysis of Venture Similarity in Entrepreneurial Ecosystems," *IEEE Transactions on Engineering Management*, 66 (4), 568–582.

Bhawe, Nachiket and Shaker A Zahra (2019) "Inducing heterogeneity in local entrepreneurial ecosystems: the role of MNEs," *Small Business Economics*, 52, 437–454.

Bishop, Paul (2012) "Knowledge, diversity and entrepreneurship: a spatial analysis of new firm formation in Great Britain," *Entrepreneurship and Regional Development*, 24 (7-8), 641–660.

Bonaccorsi, Andrea, Massimo G Colombo, Massimiliano Guerini, and Cristina Rossi-Lamastra (2014) "The impact of local and external university knowledge on the creation of knowledge-intensive firms: evidence from the Italian case," *Small Business Economics*, 43, 261–287.

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

Brandes, Ulrik, Daniel Delling, Marco Gaertler, Robert Gorke, Martin Hoefer, Zoran Nikoloski, and Dorothea Wagner (2008) "On Modularity Clustering," *IEEE Transactions on Knowledge and Data Engineering*, 20 (2), 172–188.

Carlino, Gerald and William R Kerr (2015) "Chapter 6 - Agglomeration and Innovation," in Duranton, Gilles, J Vernon Henderson, and William C Strange eds. *Handbook of Regional and Urban Economics*, 5, 349–404: Elsevier.

Castaldi, Carolina, Koen Frenken, and Bart Los (2015) “Related Variety, Unrelated Variety and Technological Breakthroughs: An analysis of US State-Level Patenting,” *Regional Studies*, 49 (5), 767–781.

Cavallo, Angelo, Antonio Ghezzi, and Cristina Rossi-Lamastra (2021) “Small-medium enterprises and innovative startups in entrepreneurial ecosystems: exploring an under-remarked relation,” *International Entrepreneurship and Management Journal*, 17, 1843–1866.

Cefis, Elena and Orietta Marsili (2011) “Born to flip. Exit decisions of entrepreneurial firms in high-tech and low-tech industries,” *Journal of Evolutionary Economics*, 21, 473–498.

Choi, Jinkyong, Jorge Guzman, and Mario Small (2024) “Third places and neighborhood entrepreneurship: Evidence from Starbucks cafés,” NBER Working Paper Series No.32604.

Colombelli, Alessandra (2016) “The impact of local knowledge bases on the creation of innovative start-ups in Italy,” *Small Business Economics*, 47, 383–396.

Colombo, Massimo G, Diego D’Adda, and Anita Quas (2019) “The geography of venture capital and entrepreneurial ventures’ demand for external equity,” *Research Policy*, 48 (5), 1150–1170.

Content, Jeroen and Koen Frenken (2016) “Related variety and economic development: a literature review,” *European Planning Studies*, 24 (12), 2097–2112.

Content, Jeroen, Koen Frenken, and Jacob A Jordaan (2019) “Does related variety foster regional entrepreneurship? Evidence from European regions,” *Regional Studies*, 53 (11), 1531–1543.

Cortinovis, Nicola and Frank van Oort (2015) “Variety, economic growth and knowledge intensity of European regions: a spatial panel analysis,” *The Annals of Regional Science*, 55, 7–32.

Dalle, Jean-Michel, Matthijs den Besten, and Carlo Menon (2017) “Using Crunchbase for economic and managerial research,” OECD Science, Technology and Industry Working Papers No.2017/08.

Duranton, Gilles and Diego Puga (2004) "Chapter 48 Micro-foundations of urban agglomeration economies," in Henderson, J Vernon and Jacques-François Thisse eds. *Handbook of Regional and Urban Economics*, 4 of Handbook of regional and urban economics, 2063–2117: Elsevier.

Edwards, Alexander and Maximilian Todtenhaupt (2020) "Capital gains taxation and funding for start-ups," *Journal of Financial Economics*, 138 (2), 549–571.

Frenken, Koen, Frank Van Oort, and Thijs Verburg (2007) "Related Variety, Unrelated Variety and Regional Economic Growth," *Regional Studies*, 41 (5), 685–697.

Ghio, Niccolò, Massimiliano Guerini, and Cristina Rossi-Lamastra (2016) "University knowledge and the creation of innovative start-ups: an analysis of the Italian case," *Small Business Economics*, 47, 293–311.

Glaeser, Edward L, Hedi D Kallal, Jose A Scheinkman, and Andrei Shleifer (1992) "Growth in Cities," *Journal of Political Economy*, 100 (6), 1126–1152.

Gompers, Paul and Josh Lerner (2001) "The venture capital revolution," *Journal of Economic Perspectives*, 15 (2), 145–168.

Guo, Qi, Canfei He, and Deyu Li (2016) "Entrepreneurship in China: The role of localisation and urbanisation economies," *Urban Studies*, 53 (12), 2584–2606.

Hartog, Matté, Ron Boschma, and Markku Sotarauta (2012) "The impact of related variety on regional employment growth in Finland 1993–2006: High-tech versus medium/low-tech," *Industry and Innovation*, 19 (6), 459–476.

Heblich, Stephan and Viktor Slavtchev (2014) "Parent universities and the location of academic startups," *Small Business Economics*, 42, 1–15.

Huang, Ying Sophie, Buhui Qiu, Jiajia Wu, and Juan Yao (2023) "Institutional distance, geographic distance, and Chinese venture capital investment: do networks and trust matter?" *Small Business Economics*, 61, 1795–1844.

Jacobs, Jane (1969) *The economy of cities*, New York, NY: Vintage Books.

Kerr, William R and Frederic Robert-Nicoud (2020) "Tech Clusters," *Journal of Economic Perspectives*, 34 (3), 50–76.

Kuusk, Kadri (2021) "Regional differences in how related variety 'works': the case of labour mobility," *European Planning Studies*, 29 (10), 1951–1973.

Leendertse, Jip, Mirella Schrijvers, and Erik Stam (2022) "Measure Twice, Cut Once: Entrepreneurial Ecosystem Metrics," *Research Policy*, 51 (9), 104336.

Marshall, Alfred (1920) *Principles of economics*, London, England: Macmillan, 7th edition.

Mathias, Blake D, Brian T McCann, and Daniel S Whitman (2021) "A meta-analysis of agglomeration and venture performance: Firm-level evidence," *Strategic Entrepreneurship Journal*, 15 (3), 430–453.

Miguelez, Ernest and Rosina Moreno (2018) "Relatedness, external linkages and regional innovation in Europe," *Regional Studies*, 52 (5), 688–701.

Newman, M E J and M Girvan (2004) "Finding and evaluating community structure in networks," *Physical Review E*, 69 (2), 026113.

Åstebro, Thomas, Navid Bazzazian, and Serguey Braguinsky (2012) "Startups by recent university graduates and their faculty: Implications for university entrepreneurship policy," *Research Policy*, 41 (4), 663–677.

Roche, Maria, Alexander Oettl, and Christian Catalini (2022) "(Co-)Working in Close Proximity: Knowledge Spillovers and Social Interactions," NBER Working Paper Series No.30120.

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

Sako, Mari, Matthias Qian, and Mark Verhagen (2022) "Knowledge similarity among founders and joiners: Impact on venture scaleup in fintech and lawtech," SSRN No.4003043.

Sorenson, Olav and Toby E Stuart (2001) "Syndication Networks and the Spatial Distribution of Venture Capital Investments," *American Journal of Sociology*, 106 (6), 1546–1588.

Stam, Erik (2015) "Entrepreneurial ecosystems and regional policy: A sympathetic critique," *European Planning Studies*, 23 (9), 1759–1769.

Theil, Henri (1972) *Statistical decomposition analysis: With applications in the social and administrative sciences*: North-Holland.

Xiong, Ning, Yehua Dennis Wei, and Yangyi Wu (2023) "Tech firm births and agglomeration economies: (un)related variety, specialization, and spatial externalities," *Cities*, 138, 104349.

Appendix: Community detection method

Using the calculated co-occurrence frequency, I performed community detection in co-occurrence network analysis to group tags (i.e., startups' business fields). Specifically, I applied modularity maximization as the community detection method (Newman and Girvan, 2004), a method also employed in prior studies on community detection in startups (Basole et al., 2019). It identifies the optimal partitioning of a network and the corresponding number of communities by maximizing the modularity index, reflecting the quality of the network division.[21] The modularity index Q is defined as follows:

$$Q = \sum_{c=1}^{n_c} \left[\frac{l_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right] \tag{6}$$

where n_c denotes the number of communities, m the total number of edges in the network, l_c the number of edges within community c, and d_c the degree of nodes within community c. Therefore, l_c/m represents the proportion of edges within community c in the actual network, whereas $(d_c/2m)^2$ provides the expected proportion of edges within community c under the assumption of a random distribution of edges, given the same node degrees. In this context, communities represent tag groups or business fields, nodes correspond to individual tags, and edges reflect the co-occurrence of tags when jointly registered for a single service.

[21] In this paper, I used the `cluster_optimal` function within the *igraph* package in R, following the approach of Brandes et al. (2008), to perform modularity optimization.