ORIGINAL ARTICLE

Higher education's influence on social networks and entrepreneurship in Brazil

Michelle Reddy[1] · Júlio C. Nardelli[2] · Yuri L. Pereira[3] · Leonardo B. Oliveira[3] · Thiago H. Silva[2] · Marisa Vasconcelos[4] · Mark Horowitz[5]

Abstract

Developing and middle-income countries increasingly emphasize higher education and entrepreneurship in their long-term development strategy. Thus, our work focuses on the influence of higher education institutions (HEIs) on startup ecosystems in Brazil, an emerging economy. As traditional data to perform this type of study, such as surveys, are challenging to get, we propose an alternative approach. Given the growing capability of social media databases such as Crunchbase and LinkedIn to provide startup and individual-level data, we draw on computational methods to mine data for social network analysis. Our approach enables different types of analysis. First, we describe regional variability in entrepreneurial network characteristics. Second, we examine the influence of elite HEIs in economic hubs on entrepreneur networks. Third, we investigate the influence of the academic trajectories of startup founders, including their courses of study and HEIs of origin, on the fundraising capacity of startups. We find that HEI quality and the maturity of the ecosystem influence startup success. We also observe that elite HEIs have a powerful influence on local entrepreneur ecosystems. Surprisingly, while the most nationally prestigious HEIs in the South and Southeast have the longest geographical reach, their network influence remains local. This means that investments in entrepreneurship, in the Brazilian context, tend to remain concentrated in wealthier cities, and may actually reinforce or increase regional inequalities. We also find that the startup ecosystem in the wealthier South and Southeast is more diverse in terms of sectors, which is more advantageous to economic development. Our approach can be helpful, especially in countries with limited studies of the interaction between startups and institutional factors supporting them. In terms of policy recommendations, we would recommend more investment at the regional level in terms of cultivating entrepreneurship, given the limited spillover from wealthier regions.

Keywords Higher education · Entrepreneurship · Startup founders · Social networks · Emerging economies · Brazil

✉ Thiago H. Silva
thiagoh@utfpr.edu.br

Michelle Reddy
mireddy@berkeley.edu

Júlio C. Nardelli
julio.2018@alunos.utfpr.edu.br

Yuri L. Pereira
yuri.pereira@dcc.ufmg.br

Leonardo B. Oliveira
leob@dcc.ufmg.br

Marisa Vasconcelos
marisaav@br.ibm.com

Mark Horowitz
horowitz@stanford.edu

1   Goldman School of Public Policy, University of California, Berkeley, CA, USA

2   Informatics, Universidade Tecnológica Federal do Paraná, Curitiba, Brazil

3   Computer Science, Federal University of Minas Gerais, Belo Horizonte, Brazil

4   IBM Research, São Paulo, Brazil

5   Computer Science, Stanford University, Stanford, USA

1 Introduction

Entrepreneurship and higher education are increasingly viewed as drivers for long-term sustainable economic development, not only in high-income, knowledge-based economies but also in emerging economies as well as developing countries. Overall, there is a growing policy interest in the relationship between higher education institutions and universities to foster entrepreneurship. The "triple helix of university–industry–government interactions" with entrepreneurship at the center has become a global economic model (Lee and Mirza 2021). Interest in entrepreneurship has not only economic goals—entrepreneurship has also been poised to be the solution to social problems. Disillusion with politics and government failure to address poverty and wider politics has catalyzed increased interest in social entrepreneurship (for Economic Co-operation and Development 2008). Therefore, entrepreneurship has been viewed as not only the solution to underdeveloped markets but also the solution to poverty. Entrepreneurship creates a form of capital generating new knowledge, innovation spillovers, and diverse business activity (David et al. 2004).

Entrepreneurship reflects a growing dominant ideology in global development that views individuals, rather than state planning, as drivers of development (Hwang 2006). While twentieth-century global development policies largely focused on the development of human capital and in particular, the role of education, in terms of fostering development, at present, entrepreneurship is increasingly becoming a defining feature of global development discourse in the twenty-first century. Institutional theorists argue that world cultural norms of empowered individualism reflect that entrepreneurship has become perceived as something that all individuals can engage in. However, research on entrepreneurship has tended to neglect the wider socioeconomic and institutional context for entrepreneurship to take place (Brandl and Bullinger 2009). For example, regional inequalities within a country are increasing viewed as a challenge to attaining development goals. Even within high-income countries, regional inequality is a major challenge to inclusive growth.

While policymakers have turned to universities as the engines of entrepreneurial development as well as regional development (Schaeffer et al 2018), the role of the university in generating entrepreneurial development has become debatable more recently. For instance, Wadie and Padayachee (2017) argue that universities are only weakly influential in the regional entrepreneurial ecosystem. The study of entrepreneurial activity in developing countries, the institutional context, and the resulting influence on development is primarily constrained by limited data availability (Dvouletý et al 2018). In response, this study uncovers the types of entrepreneurial activities across regions in Brazil, an emerging economy with high variations in regional development, and the relationship of elite higher education institutions (HEIs) to the regional entrepreneur ecosystem and beyond.

Is entrepreneurship easily distributed throughout a country, or might it exacerbate already-existing regional inequalities by naturally finding its strongest base in regions with institutional and socioeconomic advantages? The influence of entrepreneurship on development may vary across time and across regions (Bjørnskov and Foss 2016; Hartog et al 2010; Vivarelli 2013). While numerous studies note the importance of organizations and institutions to foster and disseminate entrepreneurship knowledge (see, e.g., Mitra and Formica 1997), there is limited study of startup ecosystems, that is, the interaction between startups and institutional factors supporting them (Tripathi et al. 2019). Having the appropriate data to make a proper study is challenging in terms of studying startup ecosystems, as conventional data collection methods like questionnaires and interviews do not scale well, i.e., to study large areas, such as countries or the entire world, or have fine-grained data is typically costly and slower. Thus, computational methods exploring public data on the Web become especially important because they usually offer valuable data cheaper and faster. These new methods, such as Web crawlers, text mining, and complex network analysis enable the study of the relationship between entrepreneurship and higher education in places where these types of study tend to be scarce, such as in developing countries. This study explores online social networks and open data to tackle this challenge in the Brazilian context.

If entrepreneurship is heralded as the pathway toward sustainable development, to what extent do higher education institutions (HEIs) influence entrepreneur networks? Will elite HEIs perpetuate existing inequalities, particularly across regions, by having more influence on startup ecosystems? Using social network analysis and mining public Web data of Brazilian entrepreneurs, we hypothesize that entrepreneur networks in regionally disadvantaged areas, such as the Brazilian Northeast, are closely linked with networks from top universities in the wealthier Southeast. We also examine whether the nature of networks will vary by region, given their varying levels of development, as some regions have more access to capital and others to natural resources. In addition, we test our assumption that at the regional level, elite HEIs will influence startups through the social networks formed through HEIs.

Notably, we discuss how elite HEIs, according to national educational quality rankings, influence regional entrepreneurial ecosystems. Overall, we investigate the nature of these networks within Brazil and examine how universities contribute to Brazil's regional entrepreneur networks. Specifically, we aim to address the following questions:

i) To what extent do HEIs influence entrepreneurial networks within and outside their region?
ii) How do entrepreneur networks vary by region in Brazil?
iii) Are entrepreneur networks mostly embedded in elite HEI networks?

As networks provide entrepreneurs with information, capital, and services, we examine entrepreneur networks in Latin America's largest startup ecosystem, Brazil, in this study (Lechner and Dowling 2003; Renzulli and Aldrich 2005). We chose Brazil because there is limited, if any, empirical analysis of the conditions fostering high-tech regions in middle-income countries. In particular, Brazil is a suitable case because while specific regions are middle-income, others, such as the Northeast, have GDPs similar to low-income countries. Given these stark contrasts, we look at HEIs and their influence on regional entrepreneur networks. Our Brazil analysis explores HEI influence on high-tech ecosystems in both a middle- and low-income context. We draw on Neumeyer et al (2019)'s definition of an entrepreneurial ecosystem as a "social network of individuals with reciprocal ties."

We propose using computational methods to mine public data from an online database regarding entrepreneurs in Brazil and triangulate with information publicly available in a social media network, as well as with official open data from Brazil's Ministry of Education. While (Dvouletý et al 2018) examine the established business ownership rate, we look exclusively at startups. First, we download data regarding our target Brazilian startup ecosystems from Crunchbase (Crunchbase 2019) database. Second, we collect relevant data from LinkedIn to enrich our initial data on startup ecosystems. Finally, we add information about the General Index of Courses, an official indicator of quality concerning HEIs in Brazil. Note that the use of computational methods here is significant since Brazil is a continental country, and conventional data collection methods like questionnaires and interviews do not scale well. First, we characterize Brazil's entrepreneur network at the national level. Then, we create a framework for investigating the influence of HEIs, particularly elite HEIs, degree programs, and educational quality, on entrepreneur networks. Overall, our study contributes to education, entrepreneurship, and development research. In addition, our approach to data collection and modeling can be useful in locations with limited studies of the interaction between startups and institutional factors supporting them.

2 Related work

2.1 Brazil

This study focuses on Brazil, given Brazil's position as an emerging economy with high levels of regional inequality. The leading market segments for Brazilian startups are Education (EdTechs), Health and Wellness (Healthtech), Finance (FinTechs), Agribusiness (AgTechs), E-commerce, Development of Software, Human Resources (RHTech), Communication and Marketing, and Civil Construction and Management (ABStartups 2021). Startups are mostly concentrated in the Southeast (51.1%), South (26.5%), Northeast (13.4%), Midwest (5.4%), and North (3.6%) (ABStartups 2021). Overall, the South and Southeast regions together represent 69.9% of the national GDP, and approximately 70% of goods and services produced in one year are concentrated in only five of the 27 federal units in the country (IBGE 2019).

The OECD (2020) notes that small- and medium-sized enterprises are critical for economic growth and social inclusion in Brazil, yet neglects to examine regional variation in terms of entrepreneurial ecosystems. Given a national public policy priority toward entrepreneurship in achieving economic growth, for example, as reflected in the microempreendedor individual (MEI) 2008 policy, Brazil has also invested in entrepreneurship education. According to the OECD (2020), Brazil has a strong entrepreneurial education infrastructure, for instance, government policies aimed at fostering entrepreneurship education include the National Programme for Entrepreneurship Education (Programa Nacional de Educacao Empreendedora), which reached four million students at 6,000 institutions at all levels of education (OECD 2020).

Entrepreneurship has become a taken-for-granted norm in global development policy, however, does entrepreneurship lead to development, or does entrepreneurship continue to reflect existing regional inequalities? While (Dvouletý et al 2018) did not find any impact of entrepreneurship on Human Development Index (HDI), they acknowledge the need to study the role of institutions when analyzing entrepreneurship in developing countries. As a result, our study examines higher education institution quality as an input rather than as an outcome. Regions with higher education institutions perceived as higher quality may be more likely to have a more robust startup ecosystem, and the influence of these universities might even extend into other regions. The success of startups may very well be preconditioned by existing levels of human capital, such as the availability of quality higher education in a region. We examine how higher education institutions influence startup creation and regional startup

ecosystems. Brazil is characterized by spatial inequalities, most notably among regions, and evident in the stark contrast between the Brazilian North and Northeast and the economic hub of the South and Southeast, despite improvements in recent years (Simone Affonso da Silva 2017). Educational inequality, in particular, is well documented, particularly in Brazil (Sanderson 2017) and across emerging economies (Balestra et al 2018), and globally in terms of higher education access (McCowan 2007; Nations Educational, Scientific and Cultural Organization (UNESCO) 2016; Msigwa and Faustina 2016). Overall, there are limited empirical data to support the claim that entrepreneurship stimulates development—see, for example, Hafer (2013), Thurik and Wennekers (2004), and Wennekers et al. (2010, 2005), and the role of universities in fostering regional ecosystems in developing countries has been largely understudied.

2.2 The entrepreneurial university

The entrepreneurial university is a global phenomenon due to the internal development of the university (Etzkowitz et al 2000) and as the transition to a knowledge-based economy became a goal for sustainable economic development (Labra et al 2016). Entrepreneurial activities enhance national and regional economic growth as well as university finances (Etzkowitz et al 2000). Knowledge-intensive entrepreneurship is argued to drive innovation and economic growth. However, the distribution of entrepreneurial activity is often concentrated geographically rather than spreading evenly throughout a country (Saxenian 2018; Audretsch et al 2006). Just as Brazil made strides in terms of startup growth in the past decade, it has exponentially increased access to higher education. Yet, Brazil remains a country with a high degree of economic inequality, particularly between regions. Other resource-rich countries—such as Qatar (Julia et al 2018), Malaysia, and Saudi Arabia (Kumar and van Welsum 2013)—increasingly invest in higher education to move toward a knowledge-based economy and away from natural resource dependency.

The influence of universities on science and technology-based industries is well documented (see, e.g., Rosenberg and Nelson 1994). However, the spatial allocation of entrepreneurship, particularly in developing economies, is less understood (Fischer et al 2019). University–industry linkages include the movement of university graduates into commercial firms and faculty entrepreneurship, faculty involvement on advisory boards, industry gifts supporting university research, and student training, among others (Porter et al 2005). Proximity to research-intensive universities is viewed as a source of expertise for tech entrepreneurs (Etzkowitz and Leydesdorff 1998). There

is a tendency for the research and development efforts of organizations to spill over into the innovation efforts of other organizations (Jaffe 1986), which can occur across industries but is particularly acute within regions, and amplified when key participants are research organizations (Dasgupta and David 1994; Owen-Smith and Powell 2004). However, the role of universities in shaping entrepreneurial activity varies by national level of development (González-Pernía et al 2015), and this may very well be the case in terms of variation in regional development. In particular, HEIs, and their relationship with industry, might be more advantageous in certain regions than in others (Porter et al 2005), mainly if more elite universities are clustered in economically wealthy regions. Overall, the geography of entrepreneurial activity is argued to be quite different in developing countries as opposed to developed countries (Valliere and Peterson 2009; Crescenzi and Rodríguez-Pose 2012; González-Pernía et al 2015).

The technological revolution enabled new entrepreneurial initiatives worldwide, creating an enabling environment for business without the startup costs of the larger firms that dominated the economic landscape of the mid-twentieth century in developed countries (Commission of the European Communities 2013). While technology is vital in the rise of entrepreneurship worldwide, as Banerji and Reimer (2019) note, importance of social networks in entrepreneurship is intuitive. In particular, potential funding agencies predict startup success by examining the social networks of founders (Banerji and Reimer 2019), and networks provide information and opportunities (Burt 2000; Larson 1991), and legitimacy (Klyver et al 2008).

Numerous studies have also observed the role of social ties in entrepreneur networks. In particular, Zimmer and Aldrich (1987) note the importance of social networks in all three aspects of entrepreneurial success: launching a startup, turnover, and sustainability. These findings hold across several cultural contexts, for instance, in China (Bates 1997), as well as for ethnic minorities in the USA (Light 1984). Therefore, our study results are potentially helpful for other cultural contexts, particularly middle-income and developing countries.

2.3 Theoretical framework

Entrepreneurial success is commonly explained through the lens of human capital theory (Unger et al 2011). Human capital consists of knowledge, skills, and health that enable individuals to be economically productive (Bank 2021). For instance, human capital assists with procuring entrepreneurial opportunities, acquiring

financial backing to launch ventures, and accumulating new knowledge (Alvarez and Barney 2007; Marvel et al 2016). However, most existing work on entrepreneurs and their networks focuses exclusively on high-income countries, where human capital is in higher supply. At the same time, many resource-rich and emerging economies have high levels of inequality, which could potentially be exacerbated by the network dominance of startups in capital cities or more well-off regions. As Thornton (2022) suggests, entrepreneurs need three types of capital: human capital, social capital, and financial capital, and it takes social capital to be able to recruit the best human capital and secure financial capital[1].

Social relationships are central to entrepreneurship, and entrepreneurs tend to find organizations where they live or in sectors they have worked in, leading industries to be clustered in certain places (Sorenson 2018). Universities are part of the ecosystem involved in the formation of social capital between networks of entrepreneurs. Sociologist James Coleman defines social capital as the intangible resources that are embedded in relationships or social institutions, for instance, family members, fellow students, etc. (Coleman 1988). For this study, we define social capital as "both the networks of relationships and the resources found and available in these networks" (Nahapiet and Ghoshal 1998; Batjargal 2003) in Hernández-Carrión et al (2020). Therefore, unpacking the role of social capital through social network analysis might better illuminate variation in regional entrepreneur ecosystems in contexts of high inequality.

Social capital theory and network analysis have been largely under-utilized in entrepreneur ecosystem research (Neumeyer et al 2019). Social network analysis enables us to examine the dissemination of social capital from elite universities to local startups through examining nodes and the strength of their ties. While social capital and networks are argued to benefit entrepreneurs, critics maintain that social capital and strong-tie networks can create barriers for disadvantaged groups. Indeed, in our study we find strong bonding capital (within-group ties) between elite universities and the local startup ecosystem in high-income regions and weak bridging capital (outside-group ties) between regions. The combination of strong bonding capital and weak bridging capital has been argued to lead to disconnected social clusters in entrepreneurial ecosystems (Light and Dana 2013), and our study, conducted on a more macro level, finds a combination of strong bonding capital and weak bridging capital between regions in the overall national entrepreneurial ecosystem within a country, Brazil. The favorable effects of social capital are argued to vary with different types of ventures (Morris et al 2018) as well as race and ethnicity (Light and Dana 2013) , and given the variation within regional ecosystems within Brazil in terms of types of ventures (sectors of intervention) and regional variations in race and ethnic composition, we find that that the favorable effects of social capital vary from region to region. What may seem like an appropriate sector of intervention in one region might be different from another based on differences in cultural legitimacy.

Departing from previous studies, we propose an innovative methodological approach by mining Crunchbase data, LinkedIn data, and the index of higher education institution (HEI) quality. Our approach can be applied in different areas of the globe faster than traditional methods that explore, for instance, surveys or interviews. Thus, enabling the understanding of factors supporting startups in places with limited or no previous studies on the topic, such as in several developing countries. The analysis performed in Brazil's context demonstrates the utility and potential of our approach.

3 Data and methods

3.1 Overview

Building a network of connections between HEIs and startups is challenging due to limited data on students' activities during or after graduation. Although some HEIs have career centers, which follow the evolution of the alumni network, these data are decentralized and confidential due to regulations regarding data protection. In addition, Brazil is a large country with more than 2,450 HEIs. Consequently, creating an alumni network through interviews, online questionnaires, or contacting the career centers of these institutions could fail because of the geographic distance, the lack of adequate contacts, and the cost of maintaining a structure capable of collecting this information.

Therefore, a computational methodological approach using public data available on the Web, such as Crunchbase and LinkedIn, combined with the georeferencing of HEIs and startups, is a promising methodology as it scales easily throughout the country and can provide a large amount of information about the trajectory of the alumni network efficiently and with low operating costs. For example, it is possible to obtain a person's complete academic background and entire professional trajectory in a single LinkedIn profile.

We explore three datasets in this study, namely:

Crunchbase Crunchbase is a global database updated daily that contains information about companies, funders, and staff (Eugene and Yuan 2012; Dalle et al 2017). As a partially crowd-sourced database, Crunchbase is increasingly used for academic and commercial purposes (Dalle et al 2017). We acquired a commercial license enabling unlimited access in addition to advanced search functions

---

[1] https://sites.fuqua.duke.edu/dukeven/selected-topics/value-social-capital/.

in Crunchbase. We procured all available data of Brazilian companies up to August 26, 2018. For 3375 companies throughout Brazil, we include company name; LinkedIn profile URL; founding date; company type (or category); the total of investments received; and headquarters location. As Crunchbase uniquely links other data sources such as Twitter (Tata et al 2016, 2017) and LinkedIn (Nuscheler 2016; Dalle et al 2017), we linked Crunchbase with LinkedIn to examine the characteristics of startup founders, their universities, and social networks.

LinkedIn LinkedIn, as a popular social network of professional contacts, provided the educational information of the company founders. We collected the profiles of employees that held titles such as CEO, owner, and founder with the LinkedIn profile URL obtained through Crunchbase. In the end, this yielded 1,177 profiles, and the main data collected were: degree type/level (e.g., Bachelor, Master or Ph.D.); degree area (e.g., Sociology or Computer Science); graduation year; and the name of the alma mater. Multiple degrees for the same profile were common. We collected all information on the LinkedIn profiles.

IGC The General Index of Courses (IGC INEP 2019b; OECD 2018) is the official quality indicator for HEIs in Brazil. Annually, the National Institute of Educational Studies and Research (INEP (OECD 2018; INEP 2019c)) performs the Census of Higher Education (CENSUP (INEP 2019d)), which is used to calculate the IGC, a metric of HEI quality. We used the IGC to classify HEIs as *elite* or *nonelite* institutions.[2]

3.2 Data pre-processing

For data pre-processing, we first obtained the geolocation of companies and HEI addresses. We used Google Geocode API to yield formatted address and geographic coordinates (e.g., latitude and longitude). We also standardized the name field. As Crunchbase has over 1,400 different categories for companies, we matched Crunchbase categories to categories used by the Brazilian Association of Startups (Abstartups) (ABStartups 2018). Since LinkedIn users report their educational background by open response, we standardized the names of HEIs using the IGC list, from the INEP Web site (INEP 2019a), and searched phonetically, using manual coding when necessary, to match IGC and Linkedin HEI names.

Here, we describe our methodological approach to identifying startup ecosystems. Most studies of entrepreneur networks are rich in interview and survey data (Banerji and Reimer 2019), see, for example, Zimmer and Aldrich (1987), Bates (1997), and Light (1984). Recent access to databases

Table 1 Dataset overview

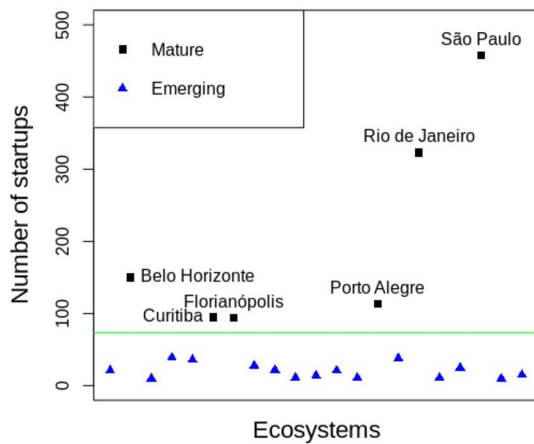|	
Startup creation period	2004 to 2018
Number of startups	1,547
Number of founders	454
Number of HEIs	146
Number of degrees obtained by founders	648
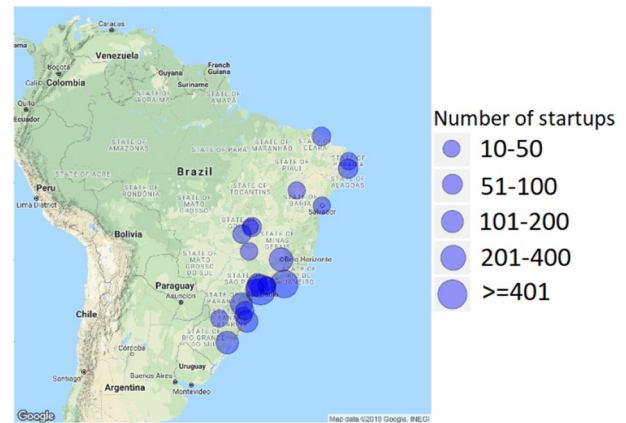
Fig. 1 Map of Brazilian ecosystems. The redder and larger the circle, the greater the number of startups. The center of the circle indicates the location of the ecosystem

such as LinkedIn and Crunchbase facilitates more generalizable results, given the ability to generate a larger sample size (Banerji and Reimer 2019). Thus, we draw on LinkedIn and Crunchbase databases and use social network analysis to test our main research questions.

We use our own data collection process, i.e., raw data extraction, transformation, and loading clean data into a centralized database. We also draw on computational methods known in the literature to solve some particular problems: To extract the main activity of startups, we use LDA topic analysis, a natural language processing method. To examine the structure of the entrepreneurship network, we construct complex network models, which fit well in cases where we want to investigate how the elements of a system interact with each other. To measure the influence of HEIs on the network, we use complex network metrics such as degree centrality, intermediation, and proximity. To measure the reach of links, we use a method of calculating the centrality of nodes based on geographic distance. To assess the students' academic trajectory similarity, we used the cosine similarity calculus to measure the similarity between two vectors and provides a better fit, in this case, than the Euclidean distance.

We examined *startups* that are at most 15 years old. From the 3375 companies we extracted from Crunchbase, we selected only 1957 (57.98%). Next, we grouped the startups

---

[2] Stanford and USP are absent from IGC rank. Yet, due to their academic excellence (Times Higher Education 2019), we regarded them as elite HEIs.

Fig. 2 Brazilian ecosystems by city. *Mature* ecosystems are in the cities where the number of startups is greater than the national mean (green line), in contrast to *emerging* ecosystems

Table 2 Mature startup ecosystems

Ecosystem	Region	Size[a]	Fundraising
São Paulo	Southeast	458	$2.6B
Rio de Janeiro	Southeast	323	$283.6M
Belo Horizonte	Southeast	151	$33.8M
Porto Alegre	South	114	$11.1M
Curitiba	South	95	$67.3M
Florianópolis	South	94	$43.5M

[a]Size stands for the number of startups

Table 3 Top emerging ecosystems

Ecosystem	Region	Size	Fundraising
Brasilia	Center-West	39	$3.7M
Recife	Northeast	38	$8.5M
Campinas	Southeast	36	$13.4M
Fortaleza	Northeast	28	$105.4K
SJ Campos	Southeast	25	$4.1M
Goiânia	Center-West	22	no info.
Barueri	Southeast	21	$10.9M
Joinville	South	21	$41.3M
Uberlândia	Southeast	15	$2.6M
João Pessoa	Northeast	14	$194.8K

by city and considered those cities with at least ten startups as *ecosystems*. Then, we examined only startups associated with our ecosystems. This yielded 21 ecosystems covering 1, 547 startups (45.83%) of our initial set. We then collected founders' data from LinkedIn, yielding 146 HEIs and 648 academic degrees of founders. Table 1 summarizes our
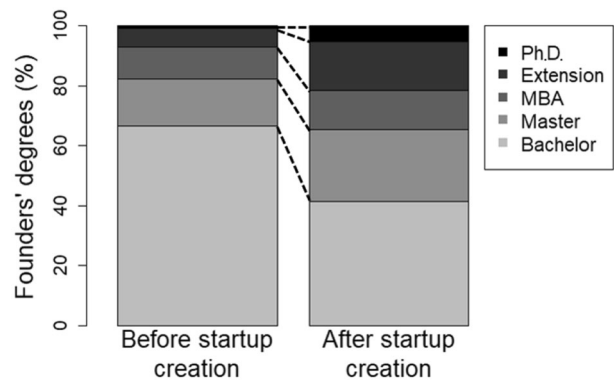
Fig. 3 Founders' degrees

dataset numbers. Figure 1 shows the geographical distribution of the ecosystems present in our dataset.

To address the gap regarding the number of startups among Brazilian systems, we divided the ecosystems into *mature* and *emerging* ecosystems. Figure 2 illustrates the difference between the two groups. Ecosystems with 74 startups or more—above the green line, which represents the average number of the observed distribution—are considered mature ecosystems. According to Crunchbase, the largest ecosystems (Table 2) are in Brazilian state capitals such as São Paulo (SP); Rio de Janeiro (RJ); Belo Horizonte (MG); Porto Alegre (RS); Curitiba (PR); and Florianópolis (SC). Together, they comprise 79.82% of startups and 97.05% of total fundraising. All of the largest ecosystems are located in the South or Southeast, the economic hub of Brazil. The emerging ecosystems, on the other hand, encompass 15 cities (Table 3). Emerging ecosystem locations are more diverse in terms of region and city size, ranging from regional capitals such as Brasilia (DF), Fortaleza (CE), and Goiânia (GO) to smaller cities such as Uberlândia (MG), Joinville (SC), and São José dos Campos (SP).

4 Results

Section 4.1 explores the network relationship between startups and HEIs, and the academic trajectory of startup founders. Section 4.2 analyzes the relationship between startups and HEIs. Finally, we investigate how educational quality influences the success of an ecosystem in Sect. 4.3.

4.1 Network characterization

Here, we analyze founders in terms of HEI, major, and degree nature (type and level). Figure 3 shows the degrees held by company founders before and after company creation. The most common degree is the Bachelor's degree, followed by the MBA and then other master's degrees.
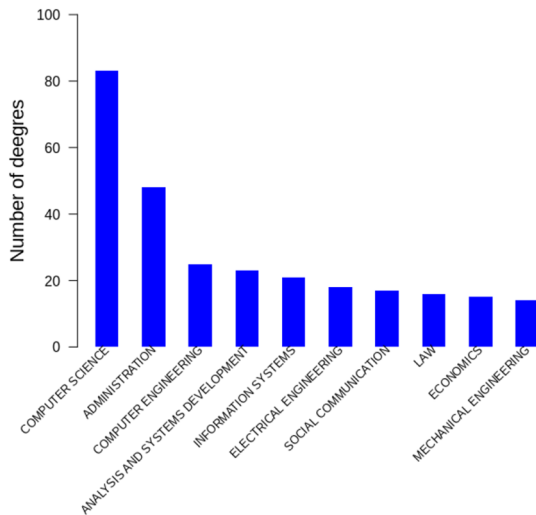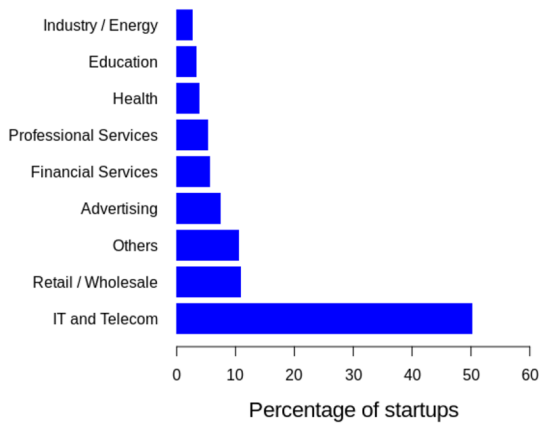
Fig. 4 Popular majors pre-startup creation

Fig. 5 Startup categories

Fig. 6 CDF of cosine similarity over HEI

We find that most founders obtain their Bachelor's degree before startup creation (Fig. 3, left). Most startups were opened during undergraduate studies. After startup creation (Fig. 3, right), the demand for other courses increased 50% (Master), 23% (MBA), 156% (Extension[3]), and 647% (Ph.D.).

This suggests that, after launching a startup, some founders may look for new educational opportunities that may add value to their business. Many startups employ disruptive technologies and non-traditional business strategies, e.g., MVP and lean-startups, and Brazilian undergraduate courses, which are the predominant types before or during startup creation, currently do not cover all these novelties. Therefore, to be well prepared and obtain better performances in the startup market, which is uncertain and

extremely competitive, founders might seek to improve themselves by taking these courses even after their companies are founded.

Figure 4 presents the founders' bachelor degree areas, before startup creation. Most degrees come from STEM (Science, Technology, Engineering, and Mathematics) ($\approx$ 59%) and social sciences ($\approx$39%). Computer Science is the most popular course of study among startup founders, and many other courses are related to Computer Science (e.g., Computer engineering). Nearly half of startups are in IT or Telecom, perhaps drawing on the computer science background of many founders and their networks (Fig. 5).

In a next step, we also examine whether founders of the same startup have similar academic trajectories, indicating a likely social relationship from university. For each founder, we consider an academic trajectory vector where the ith position represents the number of degrees concluded in HEI i. We then measure the academic trajectory similarity between each pair of founders of the same startup using cosine similarity; then, we average those values aggregating by startup. Figure 6 shows the cumulative distribution function (CDF) of this average similarity coefficient.

Note that approximately 55% of the startups have nonzero cosine similarity, which means that founders had at least one common HEI in their academic trajectory.

By further investigating the data, we found that, for the same group startups, 83% of them have contemporary founders (i.e., studied at the same HEI during the same period). Many founders may have met while at university, through acquaintances, or other university affiliations, such as being in the same social network even after university. This supports existing studies which indicate that entrepreneurs tend to have other entrepreneurs within their social circle, particularly given social costs, as starting a company can appear to differ from the established path of working for a company and having a consistent salary, (AmitaiEtzioni 1987; Toby and Waverly 2006; Aleksandra Kacperczyk

---

[3] In Brazil, *extension* courses are certified programs that do not require a Bachelor's degree, like *continuing studies* in the U.S.
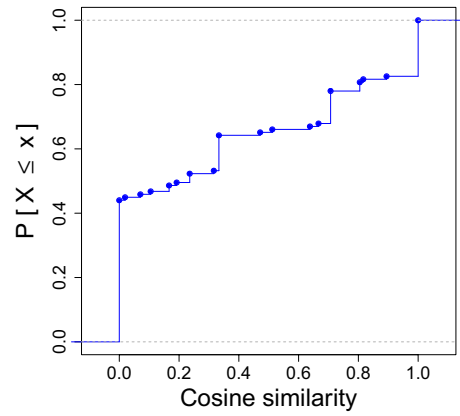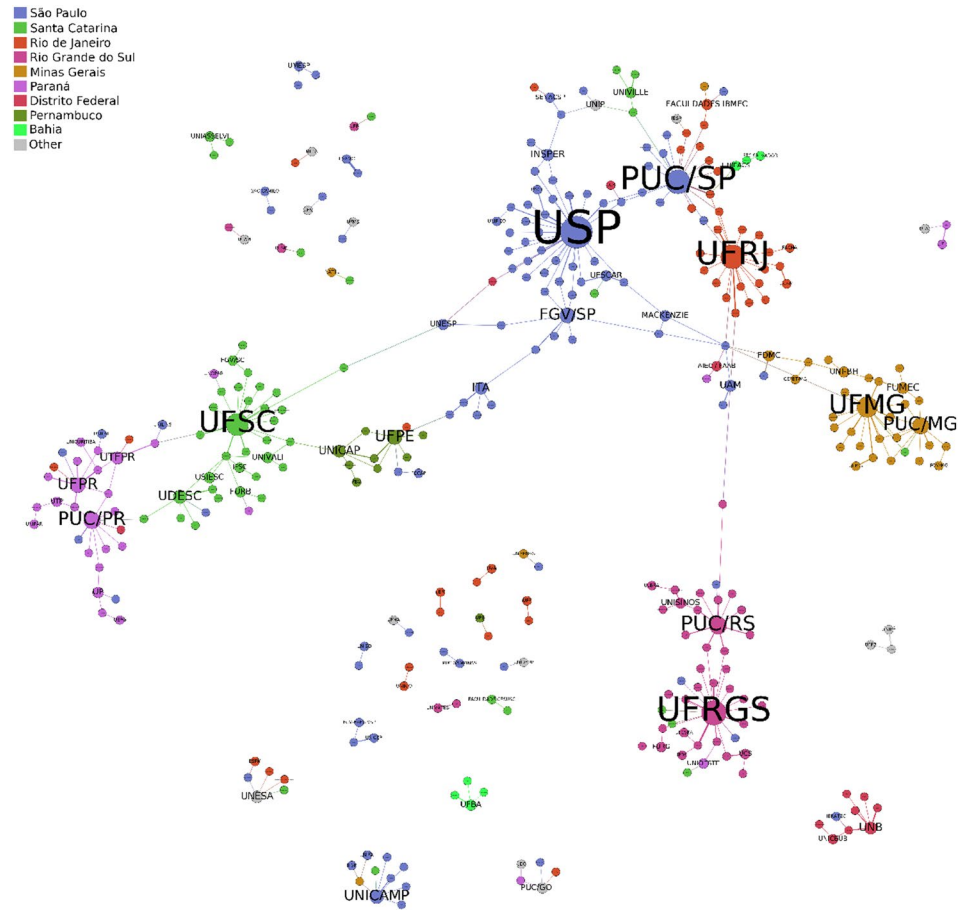
Fig. 7 Undergrad network. Node colors represent the Brazilian state where they are located in

4.2 Relationship between HEIs and startups

In this section, we analyze the network relationship between HEIs and startups. We use social network analysis to compare the described ecosystems in terms of academic trajectories, connectivity, and spatial distribution. Finally, Sect. 4.3 analyzes the success of ecosystems as a function of HEI quality rankings.

2013; Olav Sorenson 2018). Relatedly, Easley and Wang indicate that students having mentors who were entrepreneurs were more likely to become entrepreneurs (Eesley and Wang 2017). Convincing employees with permanent work to join a startup also relies upon personal connections (Ruef et al. 2003; Ruef 2010) and the quality of the startup depends heavily on early employees (Olav Sorenson 2018), therefore benefiting founders with the most extensive social networks (Michael and Olav 2014). Despite the ability to be mobile, entrepreneurs tend to locate their businesses in the places where they have the most connections (their hometown or where they attended university or graduate schools, which, at the same time, enhances their success (Claudio and Olmo 2007; Johan et al. 2017; Michael and Olav 2009; Olav 2018).

4.2.1 Network approach

We use an undirected bipartite graph $G = (U, V, E)$, where nodes $v_i \in V$ are startups, nodes $u_i \in U$ are HEIs, and an edge $e_{i,j} = (v_i, u_j)$ exists from node v_i to u_j if a startup v_i founder is a HEI u_j alum. For our analysis, we consider two networks of this kind: (i) `Undergrad` comprising only Bachelor degrees of founders (Fig. 7) and (ii) `All-Degrees` including any founder degree (Fig. 8). Both networks also include the HEIs that issued the degrees.

4.2.2 HEIs centrality

Table 4 shows the top ten HEIs according to networks' degree, closeness, and betweenness centralities (Newman 2010).

Degree centrality reflects the importance of a node through its number of connections. Notably, in our network, HEIs are only linked to startups. Therefore the degree centrality reflects the direct influence of the HEI in the startup formation. We found that University of Sao Paulo (USP), top-ranked, is the most central node in both `Undergrad` and `All-Degrees` networks (Table 4). In addition, an international HEI figures top-ranked on

Fig. 8 All-Degree network. Node colors represent the Brazilian state where they are located in
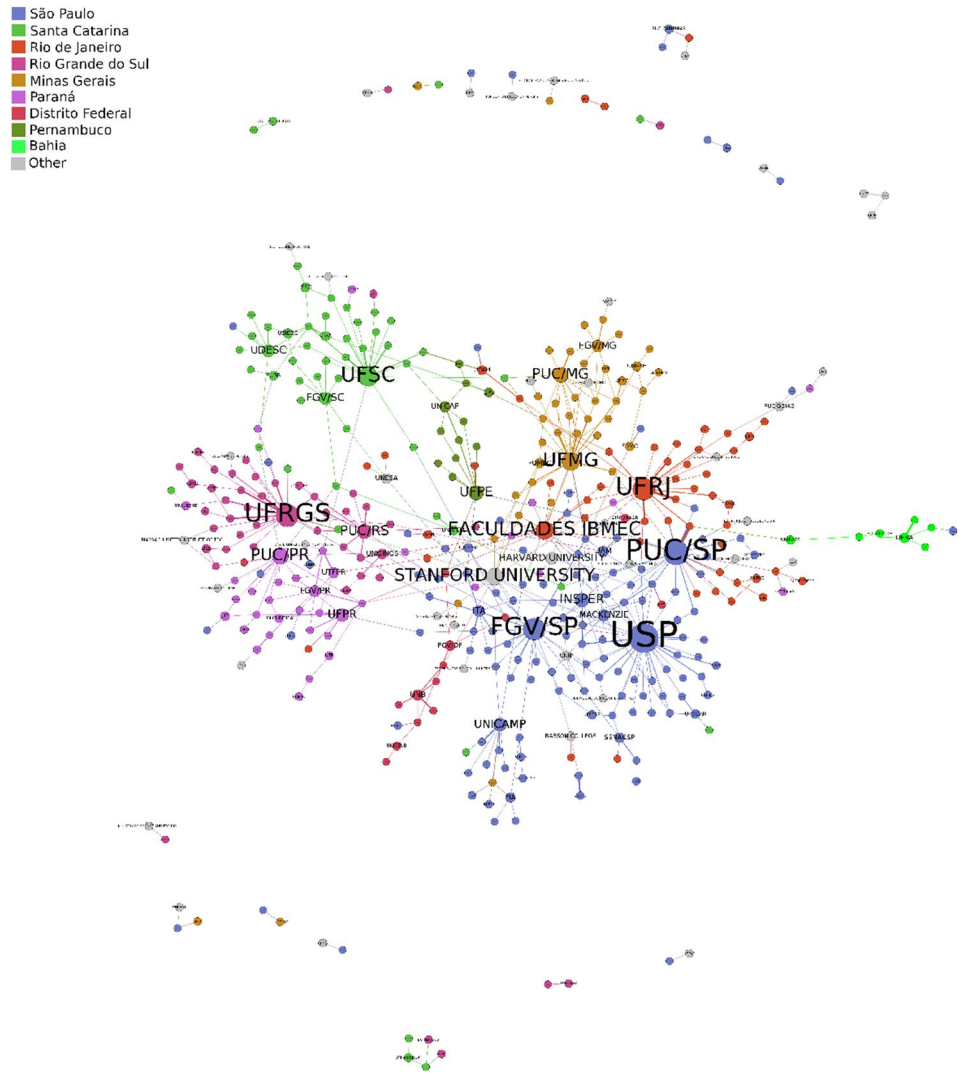
Table 4 Top 10 HEIs by degree, closeness, and betweenness centrality

Ranking	Degree		Closeness		Betweenness	
	Undergrad	All-degrees	Undergrad	All degrees	Undergrad	All degrees
01	USP	USP	FGV/SP	FGV/SP	UFSC	UNIRIO
02	UFRGS	PUC/SP	UNESP	STANFORD	FGV/SP	FGV/SP
03	UFRJ	UFRGS	USP	IBMEC	UNESP	PUC/SP
04	PUC/SP	FGV/SP	UFRJ	UFMG	USP	STANFORD
05	UFSC	UFRJ	UAM	UFRJ	UAM	USP
06	UFMG	UFSC	MACKENZIE	PUC/SP	PUC/RS	UFRJ
07	PUC/RS	UFMG	UFMG	USP	UFRJ	UFRGS
08	PUC/PR	IBMEC	FDMC	INSPER	PUC/SP	UFSC
09	PUC/MG	STANFORD	AIEC/FAAB	MACKENZIE	UFRGS	UFMG
10	FGV/SP	PUC/PR	PUC/SP	UAM	UFMG	IBMEC

`All-Degrees`: Stanford University. Upon closer examination, we find that these founders took extension courses at Stanford.

Broadly, closeness centrality captures the distance to all other nodes in the network. Here, the closeness centrality suggests that more elite HEIs reach (or influence) the network faster. In terms of undergraduate degrees among founders, Universidade Estadual Paulista (UNESP), though not top-ranked according to degree centrality, appears in the second position in terms of closeness centrality, likely because UNESP is present in 24 cities. Additionally, among the top-ranked HEIs in terms of founder undergraduate degrees, there is AIEC/FAAB (AIEC 2019), an HEI that offers online courses nationwide. Finally, FGV/SP, Stanford, and IBMEC are the most central HEIs among `All-Degrees`. This is likely due to their online course delivery and the high ranking of their business programs.

Betweenness centrality tells how often a node is within the shortest path with another in the network. In our study, this metric unveils HEIs that connect distinct social circles and then foster entrepreneurship. Here, Universidade Federal de Santa Catarina (UFSC) is the most central in `Undergrad`, and Federal University of the State of Rio de Janeiro (UNIRIO) in `All-Degrees` (Table 4). Finally, centrality top-ranked HEIs, in general, are most elite (IGC $\geqslant$ 4) HEIs. (There are two exceptions whose IGC = 3, though: AIEC and FDMC.) Also, 95 HEIs of 146 are located in major cities in the South or Southeast of Brazil, the economic hub of the country.

4.2.3 HEIs spatial degree centrality

We draw on definitions by Lima and Musolesi (2012) for our spatial degree analysis below. Each node i, $i \in V$ or $i \in U$ in our affiliation network $G = (U, V, E)$, is assigned a set of j neighbors nodes, i.e., the neighbors of node i is the set of nodes j that are reachable from i through the out-link $e_{ij} \in E$. All of them are represented by points on Earth $P_i = \{p_0^{(i)}, p_1^{(i)}, ..., p_{|j|}^{(i)}\}$, expressed through latitude and longitude.

For the spatial degree analysis, we first define a spatial neighborhood S as a circular region specified by its center and radius. Given a node i, its spatial coordinates (latitude and longitude) represent the center of a spatial neighborhood S_i with a certain radius. The intersection $P_i \cap S_i$ contains all the points, i.e., nodes representing HEIs and startups, falling inside the region S_i that are neighbors of i in G. In this way, we can compute the spatial degree centrality C of node i with spatial neighborhood S as:
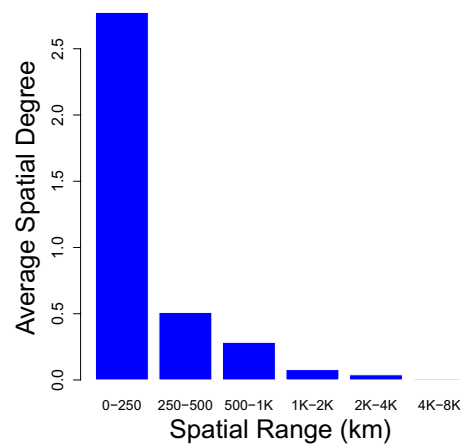
$$C_{i,S} = |P_i \cap S_i| \tag{1}$$

Fig. 9 Spatial degree analysis for different spatial neighborhood S

In this study, we are interested in the average spatial degree centrality $\overline{C}$ for HEIs. Thus, for our network $G = (U, V, E)$ this metric is expressed as:

$$\overline{C}_{U,S} = \frac{1}{|U|} \sum_{u \in U} C_{u,S}, \tag{2}$$

where the set U represents HEIs.

Figure 9 shows the spatial degree considering different non-overlapping spatial ranges, meaning the ranges are circle and expanding annular rings around each HEI. This analysis takes into account a network composed of startups that founders obtained any degree from any HEI in the 15 years before the startup creation. Most of the connections are short distance, up to 250 km, suggesting that the influence of HEIs are mostly local. However, we find that elite HEIs, such as PUC/SP, UNICAMP, and IBMEC, have the longest spatial ranges; therefore, their influence is more likely to extend beyond their local ecosystem and into other regions.

In addition, we calculate the similarity of connections in the network concerning the nodes' state, using the assortativity coefficient (Newman 2003). In general, the coefficient lies between -1 and 1. The network has perfect assortative mixing patterns when the assortativity coefficient is 1. The assortativity coefficient by *state* is 0.72 for the same network studied in the spatial analysis. This means that the majority of connections happen between nodes from the same state, corroborating with what is observed by the spatial degree centrality analysis. Thus, HEIs, in general, have more influence within their region.

The most central nodes are elite HEIs, i.e., institutions with excellence in teaching and research. Teaching is the main means of training the workforce, and research is an instrument of innovation. For example, USP offers more than 183 undergraduate courses and 239 graduate programs, with more than 80,000 students. USP is directly involved in
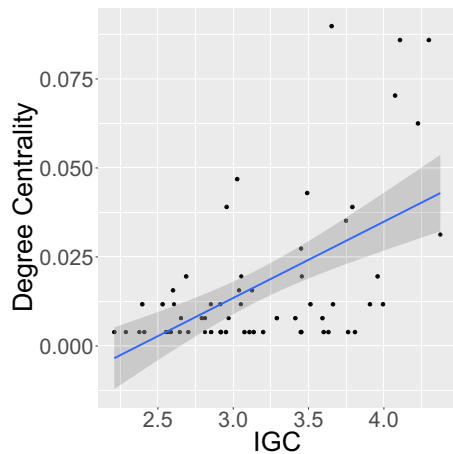
Fig. 10 Scatterplot of node degree centrality and IGC. Pearson's correlation is 0.56 (*p* value < 0.001)
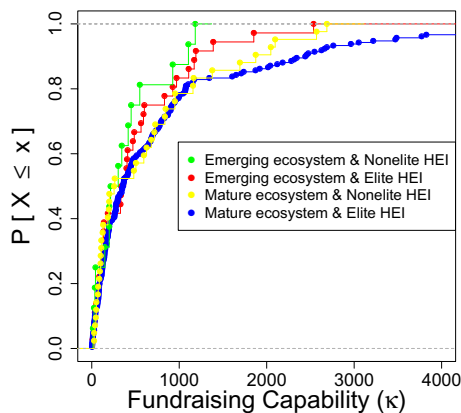
Fig. 11 CDF for fundraising capability (κ) of startups

the local innovation ecosystem with startup incubators and acceleration programs, which are extracurricular activities to stimulate and train young entrepreneurs, hackathon competitions, lectures, and workshops involving public agents, innovation agencies, companies, teachers, and students. Also, the top-ranked universities tend to attract the most highly ranked students on the national exam, who then may be more likely to form more successful startups.

4.3 Elite HEI alumni and enhanced startup fundraising capabilities

We also examine how education quality drives the success of an ecosystem. We assume that HEIs with higher educational quality rankings are perceived as more elite. We calculated the Pearson correlation between the HEIs' IGC and the HEIs' degree centrality and constructed a scatter plot in Fig. 10. The Pearson correlation is moderate and around 0.56 (p-value < 0.001). Figure 10 also plots the

linear regression, with a 95% confidence interval of the best-fit line. These findings suggest that elite HEIs (high-ranked IGC) have more startup connections and overall support our hypothesis that elite HEIs have more influence on startup ecosystems.

We also analyzed the fundraising capability (κ) of startups. Equation 3 describes how κ is calculated for a given startup *i*:

$$\kappa_i = \frac{F_{i,t_o}}{Li \times E_{i,t_o}}, \tag{3}$$

where κ_i is the fundraising capability for startup *i* up to time t_o (now), F_i the total fund raised in the life cycle (from creation up to t_o), L_i, startup age in months, and E_i current number of employees. This equation was also used by Perotti et al. (2015).

Figure 11 shows a cumulative distribution function of κ for startups whose founders are elite HEIs alumni. A founder, or a group of founders from the same company, is a product of an elite HEI if the average IGC of all HEIs she/he attended is over or equal to four. There is a positive correlation between mature ecosystems and fundraising capability. In addition, startups whose founders are elite HEI alumni tend to increase κ. Finally, the combination of a mature ecosystem and elite HEI affiliation correlates with better fundraising capability, again supporting our hypothesis that elite HEIs have more influence on regional startup ecosystems.

4.4 Regional analysis

We investigate regional variation in startup sectors. First, we examine the extent to which Brazilian regions focus on different sectors. Some sectors, for instance, might lead to faster or more sustainable economic growth. In addition, the diversity of firms is necessary for high rates of economic growth (Ingrid and Andre 2007), as Niveditha diversity enhances competition between firms leading to broader and higher quality of goods and services. Second, we examine whether there is more sector diversity for startups in wealthier regions of the country with more elite universities—namely in the South and Southeast in Brazil. For this analysis, we use the words listed in *Categories* field of Crunchbase as input to a latent Dirichlet allocation (LDA) model (Blei et al 2003). LDA is a generative statistical model to infer the topics in a collection of documents automatically. We used LDA to infer the startup niche per regionl.[4]

For each document with categories representing a certain startup, we put all words into lowercase, tokenized them, and removed stopwords, i.e., words which does not add much

---

[4] Brazil is officially divided into five regions: Center-West, North, Northeast, South, and Southeast. More info at Duran (2013).
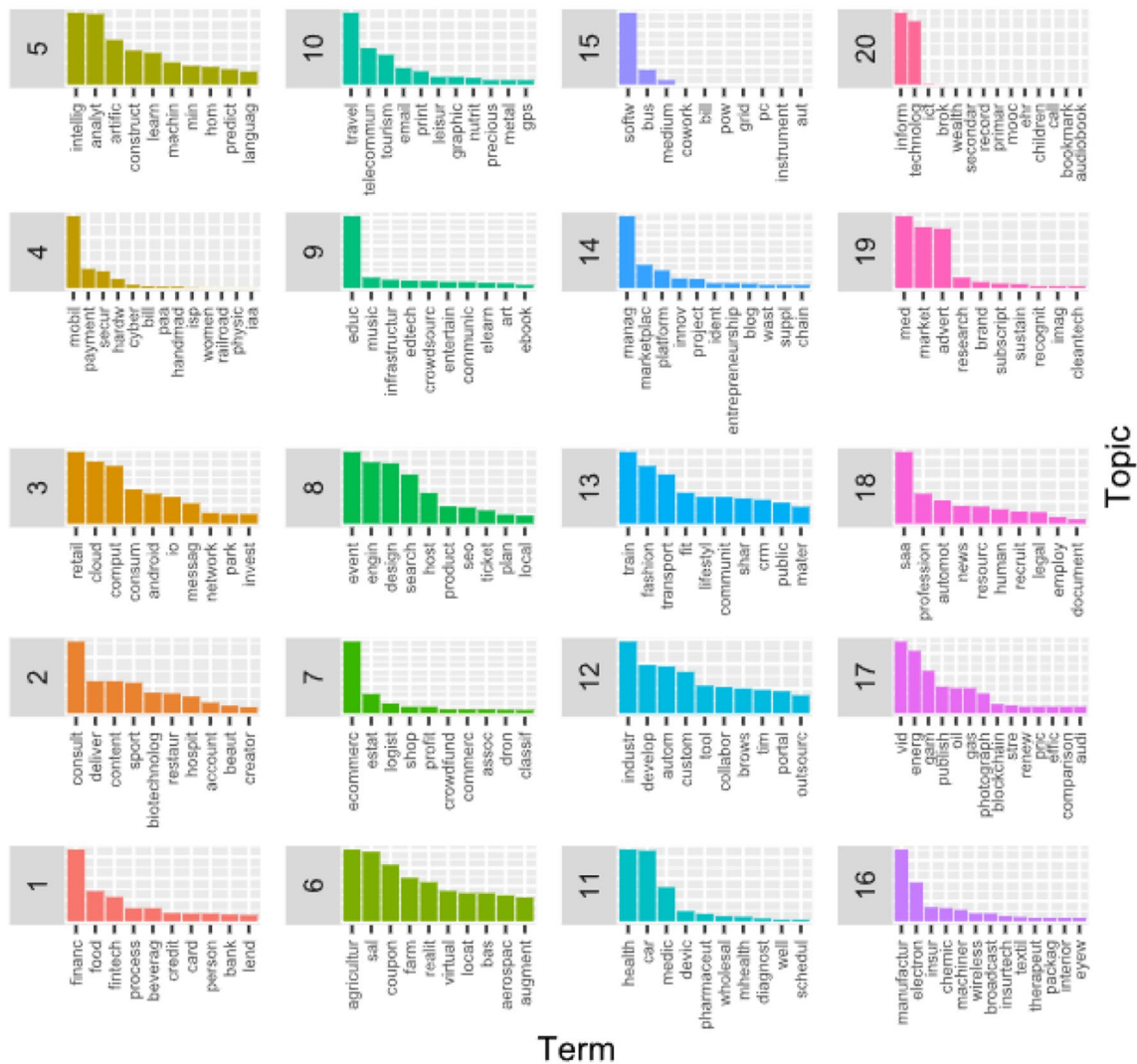
Fig. 12 Top ten words for each LDA topic. Topic labels (word with the highest probability): 1—Finance; 2—Consulting; 3—Retail; 4—Mobile; 5—IA; 6—Agriculture; 7—e-Commerce; 8—Event; 9—Education; 10—Travel; 11—Health; 12—Auto. Ind.; 13—Fashion; 14—Management; 15—Software; 16—Manufacture; 17—Energy; 18—SAAs; 19—Medical; 20—ICT

meaning to a sentence and words that might be considered stopwords in the startup sector (such as technology and service). Then, we ran the LDA algorithm using *topicmodels* package (Hornik and Grün 2011) written in R, and the best number of topics *k* returned by the algorithm using a coherence metric, implemented by *textmineR* package,[5] which approximates semantic coherence or human understandability of a topic. The highest coherence score for our dataset was *k* = 20. The most representative words of each topic are presented in Fig. 12.

We then assigned one topic to each startup, analyzing the probability of each word in the startup corpus belonging to each of the identified topics. We selected the topic with the highest aggregated probability, given all words that describe that startup as its representative topic/area.

Figure 13 shows the distribution of topics per region. As we can see, this distribution is not uniform, i.e., some regions tend to have startups associated with particular topics more than others. For instance, the North region focuses on topics 11, 12, 13, 17, and 20 related to *health*, automobile industry, *energy*, *fashion*, and ICT. Northeast region is characterized by topics 1, 2, and 6 that consists of the subjects *finance*, *consulting*, and *agriculture*, respectively, while West Centre region focuses on topics 8 (*event*), 14 (*management*) and 19 (*medical*). Finally, South and Southeast regions do not show a focus on certain topics (except for *agriculture* in South region). In general, they show a uniform distribution

Fig. 13 Distributions of the percentage of LDA topics among regions of Brazil
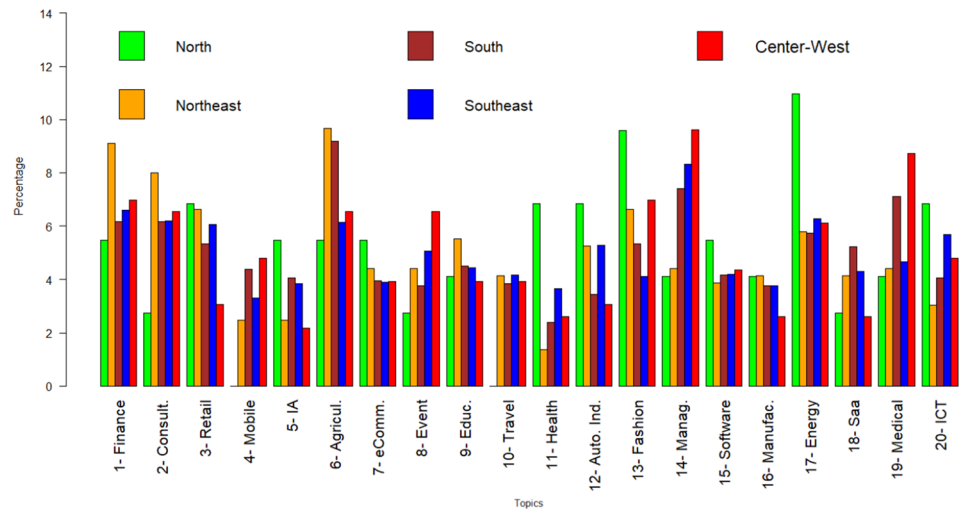
among the 20 topics. These results indicate that there are regions in Brazil that tend to concentrate more startups in certain areas, while in South and Southeast regions there is more higher diversity.

5 Discussion

The inequality trend in Brazil is also observed in the development of startup ecosystems. Although startups concentrate on more developed locations such as economic capitals (Table 2), there are also locations outside them (3). This may be evidence that entrepreneurship is an activity that spreads across the country and that regional differences favor certain sectors and create opportunities. For example, the Midwest is a region with a lot of agricultural activity, which might favor the development of *AgTechs* and possibly stimulate the creation of *startups* in diversified sectors of commerce. The North and Northeast regions have the lowest human development index (HDI) in the country (IPEA et al 2016), which could help to explain the concentration of startups in the area of health, education, consulting for small and medium businesses, and financial services; i.e., possibly to meet emerging market needs due to the lack of health, education and income conditions that characterize human development in these regions. As noted by Tolbert et al (2011); Facchini et al (2021), research on the influence of institutional cultural norms on entrepreneurship remains limited. Intervening in certain types of sectors may seem more legitimate for an organization in one region rather than another. While institutional theory maintains that global norms and culture diffuse ideals into concrete scripts that become taken-for-granted, there are regional variations in the types of startups formed and therefore there is a need for further work examining the spread of entrepreneurship culture and why organizations and networks are different from one regional

ecosystem to another. The global entrepreneurial script may well be defined differently based on the regional ecosystem.

In the Southeast region, which concentrates most of the Brazilian GDP (IBGE 2019), the financial sector is predominant (*FinTechs*), which involves credit and payment services, possibly because the concentration of wealth favors this type of activity. Also, large urban centers such as São Paulo (SP), Rio de Janeiro (RJ), and Belo Horizonte (MG) are places where disruptive education services (*EdTechs*), health and wellness (*HealthCare*)) have increasingly sought after by the population. The South, the country's second economy (IBGE 2019), stands out for its *AgTechs* as the states of Rio Grande do Sul, Santa Catarina, and Paraná are major agricultural producers. The health and wellness services startups and the real estate sector also stand out, possibly to meet the population's needs.

A recent study by ABStartups (2021) revealed that the main market segments for Brazilian startups are Education (EdTechs), Health and Wellness (Healthtech), Finance (FinTechs), Agribusiness (AgTechs), E-commerce, Development of Software, Human Resources (RHTech), Communication and Marketing, and Civil Construction and Management. The study also revealed that startups are primarily concentrated in the Southeast (51.1%), South (26.5%), Northeast (13.4%), Midwest (5.4%), and North (3.6%), which is a result consistent with the one presented in this article. However, it was carried out traditionally using questionnaires.

In this study, we investigate how HEIs contribute to Brazil's regional entrepreneur networks and their nature. Our original hypothesis suggested that entrepreneur networks in regionally disadvantaged areas, such as the North and Northeast of Brazil, will be closely linked with networks from elite HEIs in the wealthier South and Southeast. Though we find that elite HEIs, such as PUC/SP, UNICAMP, and IBMEC have the most extended spatial ranges, their network influence is still primarily local and, for instance, does not

extend to the North and Northeast. Most of the connections between HEIs and startups are local, and the most elite HEI within the region tends to have the most influence within the regional network. We also found a strong presence of Stanford University in our networks, likely reflecting the global influence of this university and Silicon Valley in business education, technology, entrepreneurship, and innovation.

In terms of variability among regional entrepreneur ecosystems, we found that the nature of networks varied by region, given their varied level of development. While IT and Telecom are the most common sectors across regions, there is more variation among startup sectors in the wealthier Southeast than in other areas. In addition, the most economically disadvantaged area in the Northeast has a strong presence of health startups, likely due to historical underdevelopment. This possibly is an avenue for future research regarding the potential of startups aimed at alleviating gaps in social service provision in low-income contexts.

Overall, we find support for our hypothesis that regional elite HEIs (the HEI with the highest educational quality rankings in the region) influence regional startup ecosystems and the fundraising capability of founders. We also found that most startup founders were contemporaries while at university, meaning that they overlapped during their course of study or met through social networks during or after their studies. We found that most startup founders studied computer science, likely reflective of the strong presence of IT and telecom startups in Brazil. In addition, there are opportunities for investment in regional entrepreneur ecosystems, and missed opportunities for stronger network tie between regions in terms of diversifying sectors of interventions.

6 Methodological limitations

Our methodology aims to understand the relationship between startups and universities in areas where this type of study is difficult to perform, particularly in developing countries. However, it is essential to acknowledge some limitations of our data, which could represent potential limitations to our methodology.

Regarding the limitations of our dataset, though Crunchbase is the most comprehensive data source for startups, it does not include all existing startups. The platform has a lower percentage of startups registered for developing and emerging economies. Also, as we rely on data entered by individual users on LinkedIn to furnish information on founders, it is important to say that not all founders in our dataset have LinkedIn profiles. The ones with profiles may be entry mistakes or misinformation. However, we believe that profile mistakes are rare, given that startup founders would risk their credibility by inputting false information.

7 Future work

Overall, a big question is whether entrepreneurship facilitates sustainable development, or does it expand existing regional inequalities. Our study begins to examine this but in future, it would also be interesting to deepen the analysis of the national entrepreneurship network and the impacts on the economy and society in terms of wealth generation and social well-being, such as the increase in quality of life indices. This could influence public policies and government strategies for sustainable economic growth and the reduction of regional inequalities. Future work might more specifically investigate whether national and university policies invest more in entrepreneurship education at elite universities, as well as the efforts by HEIs to influence startup ecosystems. At the regional level, future work could examine differences in university—startup linkages.

At a more micro-level, future work could also uncover what explains the variation in success of founders in underprivileged regions like the Northeast. Once could examine entrepreneurial orientation (EO), which is characterized by 5 dimensions: innovation, proactivity, capacity to take risks, autonomy, and competitive aggressivity. Another promising direction would be to investigate, in greater detail, the profile of the founders, identifying, in their academic and professional trajectory, the most important points that led them to become an entrepreneur and, above all, successful. This is known in the literature as entrepreneurial orientation, and the strategies employed in this study, such as using public data available on the Web and text mining techniques, could contribute to this line of research, as they scale easily to obtain massive data and further processing and analysis. This could help guide a more assertive educational experience in training young talents to create innovative businesses or even entrepreneurs who wish to increase their skills.

8 Conclusion

In this study, we take a unique conceptual approach to examine the influence of HEIs on entrepreneur networks in Brazil, a middle-income country with low-income regions. We employ an innovative methodological approach by mining publicly available data from Crunchbase, LinkedIn, and the official index of higher education institution quality, to construct and examine the social networks of startup founders. We find that most of the founders were contemporaries at the same HEI and that entrepreneurs frequently seek additional training after startup creation. We observe that the most influential nodes in the network are elite HEIs, though they usually remain within a more localized geographical range. The most nationally prestigious HEIs in

the South and Southeast have the longest spatial range into other regions yet remain fairly local, nor do they extend into the economically disadvantaged North and Northeast. One explanation may be that the startup ecosystem in the economic capitals located in the South and Southeast is not yet saturated. Perhaps once there is more saturation in this emerging market context, there will be more spillover into other regions. In addition, we find that HEI quality and the maturity of the ecosystem influence startup success. While our mature startup ecosystems are in the wealthier South and Southeast, we see some regional movement in the top emerging ecosystems.

Our findings, therefore, inform research in emerging, developing, and developed countries aiming to stimulate higher education and entrepreneurship, particularly in a context of regional inequality. We find support for the notion that contemporaries at HEIs, particularly elite HEIs, have powerful influences on entrepreneurial social networks. Our findings contribute to education, entrepreneurship, and development research more globally than studies exclusively focused on high-income countries by examining entrepreneur networks in a middle-income country, Brazil, that also has low-income regions.

## Appendix A: Full university names and abbreviations

- Universidade de São Paulo (USP)
- Universidade Anhembi Morumbi (UAM)
- Unidervsidade Estadual Paulista (UNESP)
- Universidade Presbiteriana Mackenzie (MACKENZIE)
- Faculdade Milton Campos (FDMC)
- Faculdade de Administração de Brasília (FAAB)
- Fundação Getúlio Vargas de São Paulo (FGV/SP)
- Instituto de Ensino e Pesquisa (INSPER)
- Instituto Brasileiro de Mercado de Capitais (IBMEC)
- Associação Internacional de Educação Continuada (AIEC)
- Universidade Federal do Minas Gerais (UFMG)
- Universidade Federal do Rio de Janeiro (UFRJ)
- Universidade Federal do Santa Catarina (UFSC)
- Universidade Federal do Grande do Sul (UFRGS)
- Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
- Pontifícia Universidade Católica de Paraná (PUC/PR)
- Pontifícia Universidade Católica de São Paulo (PUC/SP)
- Pontifícia Universidade Católica de Minas Gerais (PUC/MG)
- Pontifícia Universidade Católica de Rio Grande do Sul (PUC/RS)

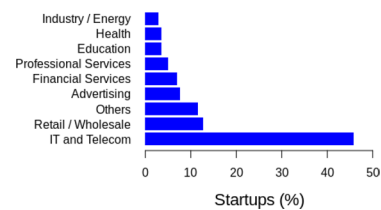## Appendix B: Startup sectors by region

See Figs. 14, 15, 16, 17, and 18.

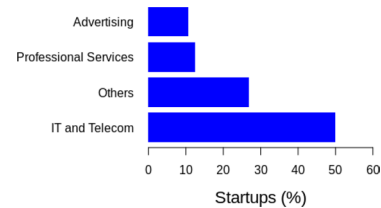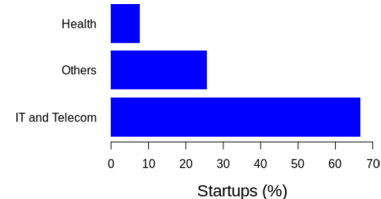## Appendix C: Summary of all ecosystems studied

See Table 5.
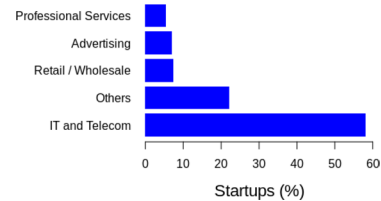
Fig. 14 South. Others: biotechnology, entertainment, industry/energy, education, financial, services, real state, health

Fig. 15 Southeast. Others: consumer products, tourism

Fig. 16 Center-West. Others: biotechnology, agribusiness, HR tech, education, consumer products, financial services, retail/wholesale, health

Fig. 17 Northeast. Others: HR Tech, lawtech, financial services, entertainment, real state, retail/wholesale, industry/energy, professional services, education
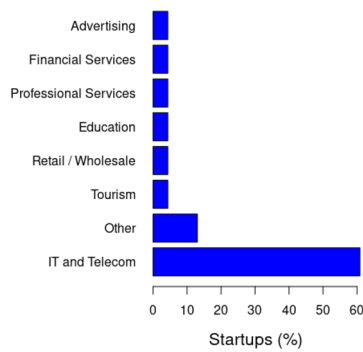
Fig. 18 North. Others: category unidentified

Table 5 Summary of all ecosystems studied

Ecosystem	State	Region	# of startups	Fundraising
São Paulo	SP	Southeast	458	$2.6B
Rio de Janeiro	RJ	Southeast	323	$283.6M
Belo Horizonte	MG	Southeast	151	$33.8M
Porto Alegre	RS	South	114	$11.1M
Curitiba	PR	South	95	$67.3M
Florianópolis	SC	South	94	$43.5M
Brasilia	DF	Center-West	39	$3.7M
Recife	PE	Northeast	38	$8.5M
Campinas	SP	Southeast	36	$13.4M
Fortaleza	CE	Northeast	28	$105.4K
São José dos Campos	SP	Southeast	25	$4.1M
Goiânia	GO	Center-West	22	no info.
Barueri	SP	Southeast	21	$10.9M
Joinville	SC	South	21	$41.3M
Uberlândia	MG	Southeast	15	$2.6M
João Pessoa	PB	Northeast	14	$194.8K
Itá	PB	Northeast	11	$2.8M
Jussara	BA	Northeast	11	$130.1K
Salvador	BA	Northeast	11	$1.8M
Blumenau	SC	South	10	$1.3M
Sorocaba	SP	Southeast	10	$530.3K

Appendix D: Classification and framework details

Classification

The classifications we used in this regard were to outline the startup ecosystems, the startups' market segments, the classification of HEIs regarding teaching and research quality, and the classification of courses.

The first concept is the startup ecosystem, a set of startups located in the same city. Ecosystems were classified as "mature" or "emerging," depending on the number of startups. Mature ecosystems were those whose number of startups was more significant than the national average. It is important to note that mature ecosystems are found in the capitals of the South and Southeast states, Brazil's wealthiest regions.

The startup industry was extracted from CrunchBase's raw data and refined through a topic mining algorithm called LDA, which is part of the natural language processing (NLP) discipline.

The HEIs were classified as "elite or not" based on the IGC, which is the official index of the Ministry of Education to assess the quality of HEIs. They were also ranked in universities, colleges, and university centers. A more detailed description of the composition of the IGC is presented as follows:

CPC—An indicator that assesses the course of study on a scale from 1 to 5. For the calculation, the following are considered: Enade Concept (student performance in the Enade test—nationwide test); Difference Indicator between Observed and Expected Performance (IDD); faculty (information from the Higher Census on the percentage of masters, doctors, and work regime) and students' perception of their training process (information from the Enade Student Questionnaire).

IGC—An indicator that evaluates the educational institution. The following are part of the IGC calculation: the average of the CPCs of the last three years of Enade (2016, 2017, and 2018) related to the evaluated courses of the institution; the average of the evaluation concepts of the master's and doctoral programs awarded by the Coordination for the Improvement of Higher Education Personnel (Capes), in the last available triennial evaluation; and distribution of students among the different levels of education, undergraduate and graduate courses.

The classification of bachelor's, master's, Ph.D.'s, MBA, or extension courses was carried out according to the courses' descriptions in the Linkedin profile.

Framework

The framework consists of acquiring public data available on the Web, and processing and calculating complex network metrics, among other metrics. The following items are the main steps of our framework:

1. Load the raw data from Crunchbase.
2. Loading of raw Linkedin data.
3. Loading of raw IGC data.
4. Extraction of categories of startups employing LDA.
5. Extraction of courses from Linkedin profiles.
6. Consolidation of clean data on a single basis.

7. Enrichment of the base with georeferencing data from startups.
8. Enrichment of the base with IGC from HEIs.
9. Construction of networks.
10. Calculation of centrality metrics.
11. Calculation of startup fundraising.
12. Data analysis and insights discussion.

Our framework was created from the combination of different elements identified in traditional models (Schaeffer et al 2018; Conceição et al 2017; Avnimelech and Feldman 2015), models based on analysis of social networks (Diánez-González and Camelo-Ordaz 2019; Hayter 2015; Soetanto and Van Geenhuizen 2015; Lyu et al 2019; Minguillo and Thelwall 2015; Porter et al 2005) and alumni networks (Rubens et al 2011) to create informal links between universities and startups, and investigation of the influence of HEIs in these network structures.

Appendix E: HEI quality and the maturity of the ecosystem

This is the algorithm used for the task studying the HEI quality and the maturity of ecosystems:

1. For each startup, calculate the fundraising (k).
2. For each ecosystem, tick "Elite HEI" if there is at least one HEI with maximum IGC (IGC = 5).
3. Calculate startup fundraising CDFs and break them down into emerging ecosystems with or without elite HEIs, and mature ecosystems with or without elite HEIs.
4. Plot the comparative CDFs between these four possibilities: mature ecosystem (with or without elite HEI) and emerging ecosystem (with or without elite HEI).

Availability of data and materials The datasets generated and analyzed during the current study will be available in a public repository after review.

Code availability Codes will be made available in a public repository after review.

Declarations

Conflict of interest/Competing interests (check journal-specific guidelines for which heading to use) The authors declare that they have no competing interests.

Ethics approval Not applicable

Consent to participate Not applicable

Consent for publication Not applicable

References

ABStartups (2018) O momento da startup Brasileira e o futuro do ecossistema de inovação. Abstartups and Accenture. http://abstartups.com.br/PDF/radiografia-startups-brasileiras.pdf

ABStartups (2021) Mapeamento de Startups. https://abstartups.com.br/mapeamento-de-comunidades/. Accessed 28 March 2021

AIEC (2019) Faculdade AIEC. https://www.aiec.br/. Accessed 14 April 2019

Alvarez SA, Barney JB (2007) Discovery and creation: alternative theories of entrepreneurial action. Strateg Entrep J 1(1–2):11–26

AmitaiEtzioni (1987) Entrepreneurship, adaptation and legitimation: a macro-behavioral perspective. J Econ Behav Org 8(2):175–189. https://doi.org/10.1016/0167-2681(87)90002-3

Audretsch DB, Keilbach MC, Lehmann EE et al (2006) Entrepreneurship and economic growth. Oxford University Press, Oxford

Avnimelech G, Feldman MP (2015) The stickiness of university spin-offs: a study of formal and informal spin-offs and their location from 124 us academic institutions. Int J Technol Manage 68(1–2):122–149

Balestra C, Llena-Nozal A, Murtin F et al (2018). Inequalities in emerging economies. https://doi.org/10.1787/6c0db7fb-en

Banerji D, Reimer T (2019) Startup founders and their LinkedIn connections: are well-connected entrepreneurs more successful? Comput Hum Behav 90:46–52. https://doi.org/10.1016/j.chb.2018.08.033

Bank W (2021) World development report 2021: Data for better lives. The World Bank

Bates T (1997) Financing small business creation: the case of Chinese and Korean immigrant entrepreneurs. J Bus Ventur 12(2):109–124. https://doi.org/10.1016/S0883-9026(96)00054-7

Batjargal B (2003) Social capital and entrepreneurial performance in Russia: a longitudinal study. Organ Stud 24(4):535–556. https://doi.org/10.1177/0170840603024000400

Bjørnskov C, Foss NJ (2016) Institutions, entrepreneurship, and economic growth: what do we know and what do we still need to know? Acad Manag Perspect 30(3):292–315. https://doi.org/10.5465/amp.2015.0135

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3(1):993–1022

Brandl J, Bullinger B (2009) Reflections on the societal conditions for the pervasiveness of entrepreneurial behavior in western societies. J Manag Inq 18(2):159–173

Burt RS (2000) The network structure of social capital. Res Org Behav 22:345–423. https://doi.org/10.1016/S0191-3085(00)22009-1

Coleman JS (1988) Social capital in the creation of human capital. Am J Sociol 94:S95–S120

Commission of the European Communities (2013) Entrepreneurship in Europe. http://ec.europa.eu/invest-in-research/pdf/download_en/entrepreneurship_europe.pdf

Conceição O, Faria AP, Fontes M (2017) Regional variation of academic spinoffs formation. J Technol Transf 42(3):654–675

Crescenzi R, Rodríguez-Pose A (2012) Infrastructure and regional growth in the European union. Pap Reg Sci 91(3):487–513

Crunchbase (2019) Main site. https://www.crunchbase.com/. Accessed 14 April 2019

Michael SD, Olav S (2009) The embedded entrepreneur. Eur Manag Rev 6(3):172–181. https://doi.org/10.1016/j.respol.2017.07.004

Michael SD, Olav S (2014) The who, why and how of spinoffs. Ind Corp Chang 23(3):661–688. https://doi.org/10.1093/icc/dtt032

Dalle JM, den Besten M, Menon C (2017). Using Crunchbase for economic and managerial research. https://doi.org/10.1787/6c418d60-en

Dasgupta P, David P (1994) Toward a new economics of science. Res Policy 23:487–521. https://doi.org/10.1016/0048-7333(94)01002-1

David B. Audretsch, Max Keilbach (2004) Entrepreneurship capital: determinants and impact. Discussion Papers on Entrepreneurship, Growth and Public Policy (ISSN 1613-8333)

Diánez-González JP, Camelo-Ordaz C (2019) The influence of the structure of social networks on academic spin-offs' entrepreneurial orientation. Ind Mark Manag 80:84–98. https://doi.org/10.1016/j.indmarman.2017.12.009

Duran R (2013) Brazilian regions. The Brazil Business. https://thebrazilbusiness.com/article/brazilian-regions

Dvouletý O, Gordievskaya A, Procházka DA (2018) Investigating the relationship between entrepreneurship and regional development: case of developing countries. J Glob Entrepreneurship Res. https://doi.org/10.1186/s40497-018-0103-9

Economic Co-operation O, Development (2008) OECD factbook 2008. OECD publishing, Paris

Charles E, Yanbo W (2017) Social influence in career choice: evidence from a randomized field experiment on entrepreneurial mentorship. Res Policy 46(3):636–650. https://doi.org/10.1016/j.respol.2017.01.010

Etzkowitz H, Leydesdorff L (1998) A triple helix of university-industry-government relations: introduction. Ind High Educ 12(4):197–201

Etzkowitz H, Webster A, Gebhardt C et al (2000) The future of the university and the university of the future: evolution of ivory tower to entrepreneurial paradigm. Res Policy 29(2):313–330. https://doi.org/10.1016/S0048-7333(99)00069-4

Eugene LY, Yuan SD (2012) Where's the Money? The social behavior of investors in Facebook's small world. In: 2012 IEEE/ACM international conference on advances in social networks analysis and mining, pp 158–162, https://doi.org/10.1109/ASONAM.2012.36

Facchini F, Jaeck L, Bouhaddioui C (2021) Culture and entrepreneurship in the United Arab Emirates. J Knowl Econ 12(3):1245–1269

Fischer BB, Schaeffer PR, Vonortas NS (2019) Evolution of university-industry collaboration in Brazil from a technology upgrading perspective. Technol Forecast Soc Chang 145:330–340

González-Pernía JL, Jung A, Peña I (2015) Innovation-driven entrepreneurship in developing economies. Entrepreneurship Region Dev 27(9–10):555–573

Hafer RW (2013) Entrepreneurship and state economic growth. J Entrepreneurship Public Policy 2(1):67–79. https://doi.org/10.1080/08985620802332723

Hartog C, Parker S, van Stel A et al (2010) The two-way relationship between entrepreneurship and economic performance. Sci Anal Entrepreneurship SMEs 17:1–45

Hayter CS (2015) Social networks and the success of university spin-offs: toward an agenda for regional growth. Econ Dev Q 29(1):3–13. https://doi.org/10.1177/0891242414566451

Hernández-Carrión C, Camarero-Izquierdo C, Gutiérrez-Cillán J (2020) The internal mechanisms of entrepreneurs' social capital: a multi-network analysis. BRQ Bus Res Quart 23(1):2340944420901047. https://doi.org/10.1177/2340944420901047

Hornik K, Grün B (2011) topicmodels: an r package for fitting topic models. J Stat Softw 40(13):1–30

Hwang H (2006) Planning development: Globalization and the shifting locus of planning. World society and organizational change, Globalization and organization, pp 69–90

IBGE (2019) Sistema de contas regionais: Brasil 2017. IBGE, Rio de Janeiro

INEP (2019a) Censo da Educação Superior. http://inep.gov.br/educacao-superior. Accessed 14 April 2019

INEP (2019b) Indice Geral de Cursos (IGC). http://inep.gov.br/en/indice-geral-de-cursos-igc-. Accessed 14 April 2019

INEP (2019c) Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. http://portal.inep.gov.br/. Accessed 14 April 2019

INEP (2019d) Sistema do censo da educação superior (CENSUP). http://sistemascensosuperior.inep.gov.br/censosuperior_2018. Accessed 14 April 2019

IPEA, PNUD, Pinheiro FJ (2016) Desenvolvimento humano nas macrorregiões brasileiras: 2016. IPEA, Brasília

Ingrid V, Andre JVS (2007) Entrepreneurial diversity and economic growth. In: The entrepreneurial society: how to fill the gap between knowledge and innovation

Jaffe AB (1986) Technological opportunity and spillovers of R &D: evidence from firms' patents, profits, and market Value. Am Econ Rev 76(5):984–1001. https://doi.org/10.3386/w1815

Julia G, Julia B, Fietkiewicz K et al (2018) Transitioning towards a knowledge society: Qatar as a case study. Springer. https://doi.org/10.1007/978-3-319-71195-9

Kacperczyk A (2013) Social influence and entrepreneurship: the effect of university peers on entrepreneurial entry. Organ Sci 24(3):664–683. https://doi.org/10.1287/orsc.1120.0773

Klyver K, Hindle K, Meyer D (2008) Influence of social network structure on entrepreneurship participation-A study of 20 national cultures, Springer International Publishing, Berlin, pp 331–347. https://doi.org/10.1007/s11365-007-0053-0

Kumar KB, van Welsum D (2013) Knowledge-based economies and basing economies on knowledge: Skills a missing link in GCC countries. RAND Corporation. https://www.rand.org/pubs/research_reports/RR188.html

Labra R, Rock JA, Álvarez I (2016) Identifying the key factors of growth in natural resource-driven countries. A look from the knowledge-based economy. Ensayos sobre Política Económica 34(79):78–89. https://doi.org/10.1016/j.espe.2015.12.001

Larson A (1991) Partner networks: leveraging external ties to improve entrepreneurial performance. J Bus Ventur 6(3):173–188. https://doi.org/10.1016/0883-9026(91)90008-2

Johan PL, Karl W, Johan W, Mike W (2017) Location choices of graduate entrepreneurs. Res Policy 46(8):1490–1504. https://doi.org/10.1016/j.respol.2017.07.004

Lechner C, Dowling M (2003) Firm networks: external relationships as sources for the growth and competitiveness of entrepreneurial firms. Entrepreneurship Region Dev 15(1):1–26. https://doi.org/10.1080/08985620210159220

Lee SS, Mirza C (2021) Pursuing innovation as a strategy for improving quality of higher education: challenges and opportunities facing gcc countries. In: Higher education in the Gulf. Routledge, pp 16–31

Light I (1984) Immigrant and ethnic enterprise in North America. Ethn Racial Stud 7(2):195–216. https://doi.org/10.1080/01419870.1984.9993441

Light I, Dana LP (2013) Boundaries of social capital in entrepreneurship. Entrep Theory Pract 37(3):603–624

Lima A, Musolesi M (2012) Spatial dissemination metrics for location-based social networks. In: Proceedings of the 2012 ACM conference on ubiquitous computing, pp 972–979, https://doi.org/10.1145/2370216.2370429

Lyu L, Wu W, Hu H et al (2019) An evolving regional innovation network: collaboration among industry, university, and research institution in china's first technology hub. J Technol Transf 44(3):659–680

Marvel MR, Davis JL, Sproul CR (2016) Human capital and entrepreneurship research: a critical review and future directions. Entrep Theory Pract 40(3):599–626

McCowan T (2007) Expansion without equity: an analysis of current policy on access to higher education in Brazil. High Educ 53:579–598. https://doi.org/10.1007/s10734-005-0097-4

Claudio and Olmo (2007) Why so many local entrepreneurs? Rev Econ Stat 89(4):615–633. https://doi.org/10.1162/rest.89.4.615

Minguillo D, Thelwall M (2015) Research excellence and university-industry collaboration in UK science parks. Res Eval 24(2):181–196

Mitra J, Formica P (1997) Innovation and economic development: University-enterprise partnerships in action. Oak Tree Press

Morris MH, Neumeyer X, Jang Y et al (2018) Distinguishing types of entrepreneurial ventures: an identity-based perspective. J Small Bus Manage 56(3):453–474

Msigwa FM (2016) Widening participation in higher education: a social justice analysis of student loans in Tanzania. High Educ 72(4):541–556. https://doi.org/10.1007/s10734-016-0037-5

Nahapiet J, Ghoshal S (1998) Social capital, intellectual capital, and the organizational advantage. Acad Manag Rev 23(2):242–266. https://doi.org/10.5465/amr.1998.533225

Nations Educational, Scientific and Cultural Organization (UNESCO) (2016) Education 2030: towards inclusive and equitable quality education and lifelong learning for all. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000245656

Neumeyer X, Santos SC, Caetano A et al (2019) Entrepreneurship ecosystems and women entrepreneurs: a social capital and network approach. Small Bus Econ 53(2):475–489

Newman ME (2003) Mixing patterns in networks. Phys Rev E 67(2):026126. https://doi.org/10.1103/PhysRevE.67.026126

Newman M (2010) Networks: an introduction. Oxford University Press Inc, New York, NY. https://doi.org/10.1093/acprof:oso/9780199206650.001.0001

Nuscheler D (2016) Regularly change a running system! An analysis of stage-specific criteria for attracting venture capital and changing the likelihood for getting funded. http://ifabs.org/assets/stores/1206/userfiles/3IFABS%20Best%20Poster%20Award%20-%20Daniela%20Nuscheler,%20TU%20Dortmund%20University,%20DE.pdf

OECD (2018) Rethinking quality assurance for higher education in Brazil. https://doi.org/10.1787/9789264309050-en

Owen-Smith J, Powell W (2004) Knowledge networks as channels and conduits: the effects of spillovers in the Boston biotechnology community. Org Sci 15:5–21. https://doi.org/10.1287/orsc.1030.0054

Perotti V, Yu Y (2015) Startup tribes: social network ties that support success in new firms. In: AMCIS 2015 proceedings. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.852.4582 &rep=rep1 &type=pdf

Porter K, Whittington KB, Powell WW (2005) The institutional embeddedness of high-tech regions: relational foundations of the Boston biotechnology community. Clust Netw Innov 261:296

Renzulli LA, Aldrich H (2005) Who can you turn to? Tie activation within core business discussion networks. Soc Forces 84(1):323–341. https://doi.org/10.1353/sof.2005.0122

Rosenberg N, Nelson R (1994) American universities and technical advance in industry. Res Policy 23(3):323–348. https://doi.org/10.1016/0048-7333(94)90042-6

Rubens N, Russell M, Perez R, et al (2011) Alumni network analysis. In: 2011 IEEE global engineering education conference (EDUCON), pp 606–611. https://doi.org/10.1109/EDUCON.2011.5773200

Martin R, Howard EA, Nancy MC (2003) The structure of founding teams: homophily, strong ties, and isolation among US entrepreneurs. Am Sociol Rev 68(2):195–222. https://doi.org/10.2307/1519766

Ruef M (2010) The entrepreneurial group: social identities, relations, and collective action. Princeton University Press, Princeton, Nova Jersey, USA

Sanderson T (2017) Education remains the catalyst for Brazil's staggering inequality. https://brazilian.report/society/2017/11/06/education-brazil-staggering-inequality

Saxenian A (2018) Inside-out: regional networks and industrial adaptation in silicon valley and route 128. In: The sociology of economic life. Routledge, pp 357–374

Schaeffer PR, Fischer B, Queiroz S (2018) Beyond education: the role of research universities in innovation ecosystems. Foresight STI Gov 12(2):50–61

Simone Affonso da Silva (2017) Regional inequalities in Brazil: divergent readings on their origin and public policy design. EchoGeo. https://journals.openedition.org/echogeo/15060

Soetanto D, Van Geenhuizen M (2015) Getting the right balance: university networks' influence on spin-offs' attraction of funding for innovation. Technovation 36:26–38

Olav S (2018) Social networks and the geography of entrepreneurship. Small Bus Econ 51:527–537. https://doi.org/10.1007/s11187-018-0076-7

Sorenson O (2018) Social networks and the geography of entrepreneurship. Small Bus Econ 51(3):527–537

Toby ES, Waverly WD (2006) When do scientists become entrepreneurs? The social structural antecedents of commercial activity in the academic life sciences. Am J Sociol 112(1):97–144. https://doi.org/10.1086/502691

Tata A, Martinez DL, Garcia D et al (2017) The psycholinguistics of entrepreneurship. J Bus Ventur Insights 7:38–44. https://doi.org/10.1016/j.jbvi.2017.02.001

Tata A, Laureiro Martinez D, Brusoni S (2016) Don't look back? The effect of attention to time and self on startup funding. Acad Manag Proc 2016(1):13,926. https://doi.org/10.5465/ambpp.2016.13926abstract

Thurik R, Wennekers S (2004) Entrepreneurship, small business and economic growth. J Small Bus Enterp Dev 11(1):140–149. https://doi.org/10.1108/14626000410519173

Times Higher Education (2019) The World University Rankings. https://www.timeshighereducation.com/world-university-rankings/2019/world-ranking#!/page/0/length/25/sort_by/rank/sort_order/asc/cols/stats. Accessed 14 April 2019

Tolbert PS, David RJ, Sine WD (2011) Studying choice and change: the intersection of institutional theory and entrepreneurship research. Organ Sci 22(5):1332–1344

Nirnaya T, Pertti S, Ganesh B, Markku O, Kari L (2019) Insights into startup ecosystems through exploration of multi-vocal literature. Inf Softw Technol 105:56–77

Unger JM, Rauch A, Frese M et al (2011) Human capital and entrepreneurial success: a meta-analytical review. J Bus Ventur 26(3):341–358

Valliere D, Peterson R (2009) Entrepreneurship and economic growth: evidence from emerging and developed countries. Entrepreneurship Region Dev 21(5–6):459–480

Vivarelli M (2013) Is entrepreneurship necessarily good? Microeconomic evidence from developed and developing countries. Ind Corp Chang 22(6):1453–1495. https://doi.org/10.1093/icc/dtt005

Wennekers S, Van Wennekers A, Thurik R et al (2005) Nascent entrepreneurship and the level of economic development. Small Bus Econ 24(3):293–309. https://doi.org/10.1007/s11187-007-9066-x

Wennekers S, van Stel A, Carree M, et al (2010) The relationship between entrepreneurship and economic development: is it u-shaped? Foundations and trends ®in entrepreneurship 6(3):167–237. https://doi.org/10.1561/0300000023

Zimmer C, Aldrich H (1987) Resource mobilization through ethnic networks: Kinship and friendship ties of shopkeepers in England. Sociol Perspect 30(4):422–445. https://doi.org/10.2307/1389212