An academic perspective on the entrepreneurship policy agenda: themes, geographies and evolution

Arenal, Alberto and Feijoo, Claudio and Moreno, Ana and Armuña, Cristina and Ramos, Sergio

Technical University of Madrid, UNED, Tongji University

5 June 2019

An academic perspective on the entrepreneurship policy agenda: themes, geographies and evolution

•

Alberto Arenal[1], Claudio Feijoo[2], Ana Moreno[3], Sergio Ramos[4], Cristina Armuña[5]

[1]*Technical University of Madrid: José Gutiérrez Abascal Street, 2. 28006 – Madrid (Spain). E-mail: albertoarenal@gmail.com ; a.arenal@alumnos.upm.es /* [4]*National Distance Learning University, Spain - Paseo Senda del Rey, 11. 28040 – Madrid (Spain).* ***(corresponding author)***

[2]*Tongji University – Universidad Politécnica de Madrid: Sino-Spanish Campus, Yifu Building, 1239 Siping Rd, 200092 Shanghai (PR China). E-mail: claudio.feijoo@upm.es*

[3]*Technical University of Madrid, Spain - José Gutiérrez Abascal Street, 2. 28006 – Madrid*

[4]*National Distance Learning University, Spain - Paseo Senda del Rey, 11. 28040 – Madrid (Spain). E-mail: sramos@cee.uned.es*

[5]*Technical University of Madrid, Spain - José Gutiérrez Abascal Street, 2. 28006 – Madrid. E-mail: cristinaag@alumnos.upm.es/* [4]*National Distance Learning University, Spain - Paseo Senda del Rey, 11. 28040 – Madrid (Spain).*

Abstract

Text mining is being increasingly used for the automatic analysis of different corpus of documents, either standalone or complementarily to other bibliometric techniques. The case of academic research into entrepreneurship policy is particularly interesting due to the increasing relevance of the topic and since the knowledge about the evolution of themes in this field is still rather limited. Consequently, this paper analyses the key topics, trends and shifts that have shaped the entrepreneurship policy research agenda to date using text mining techniques, cluster analysis and complementary bibliographic data to examine the evolution of a corpus of 1,048 academic papers focused on entrepreneurial-related policies and published during the period 1990-2016 in ten of the most relevant entrepreneurship journals. The results of the analysis show that inclusion, employment and regulation-related papers have largely dominated the research in the field, evolving from an initial classical approach about the relationship between entrepreneurship and employment to a wider and multidisciplinary perspective, including the relevance of management, geographies, and narrower topics such as agglomeration economics or internationalization instead of previous generic sectorial approaches. Overall, the text mining analysis reveals how entrepreneurship policy research has gained increasing attention and has become both more open, with a growing cooperation among researchers from different affiliations; and more sophisticated, with concepts and themes that moved forward the research agenda closer to the priorites of policies implementation.

Keywords: Text mining; Cluster analysis: Entrepreneurship policy; Entrepreneurship Research.

JEL: L26;

Availability of data: Documents analysed during the current study are available in the Open Science Framework repository, Available at:
https://osf.io/uwph3/?view_only=22d372dcf16e4c799d4726e726069990

Introduction

The emergence of the entrepreneurship as a policy and research field is the logical consequence of the increasing attention on entrepreneurship as a strategic driver of progress during the last quarter of the 20th century (Gilbert et al. 2004; Hart 2003; Stevenson and Lundström 2002). In an economic context aimed at a renewed and innovation-based growth, entrepreneurship started to be considered a key constituent (Z. J. Acs et al. 2012; Audretsch, Bönte, and Keilbach 2008). In turn, public intervention has been included in the field of entrepreneurship as a strategic priority, and it broadly focused on addressing policies oriented to support those who decided to start a business—the entrepreneurs as such—and promoting entrepreneurial culture over the whole society (Verheul et al. 2001). Similarly, research on entrepreneurship policy has grown significantly during the last three decades with remarkable contributions both from the academic (A Lundström and Stevenson 2002; Wennekers, Sander; Thurik 1999) and non-academic worlds (OECD 2010; Ramlogan and Rigby 2012).

However this interest, reseach on entrepreneurship policy is still in its early stages and the knowledge about the evolution of the field is still limited, to the extent it has been defined as under-investigated (Mazzarol and Volery 2015; Rosa 2013). In fact, there is not a comprehensive analysis in order to identify the particular themes that configure the field. As an emerging topic, there have appeared a range of controversies, not only about how the policy intervention should be configured (Lerner 2013a; Minniti 2008) but also about which policy rationale should be its justification (Z. Acs et al. 2016; Shane 2009). Adding it all up, we can say, as Lundstrom and Stevenson (2006, 7;17) noted, that in general there is '*limited knowledge on how entrepreneurship policy is constructed*' and precisely this relative novelty has implied that informed policy based on research is *'quite complex in the field of entrepreneurship policy'*. The consequence is that it is difficult to frame which priorities have led the research agenda, being a major issue both for understanding the scientific structure of the field as a first step to position further research; and also for the use of the research results in practice.

Based on this premises, this paper aims precisely to bridge this gap when mapping research on entrepeneurship policy by providing a retrospective view of the emergence and configuration of the intellectual framework in the field, laying the ground for both newcomers in the discipline and policymakers. In particular, this study uses text mining to explore how entrepreneurship policy concepts and themes have been placed into the research agenda since 1990, seeking for evidences in the path to maturity as a specific research field. In addition, within the framework of the overarching common trends in the research agenda, results from the analysis also help to understand the geographical configuration of the field, in search for particular interests depending on authors´ affiliations. Depicting the evolution of this conceptual framework and identifying key themes and priorities will facilitate both better plans for future works in less developed research areas and also will help to reduce the gap between academia and policymakers.

To these objetives, text mining is applied to a corpus of significant academic papers related to policy and belonging to top ten entrepreneurship journals during the period 1990-2016. First, the exploratory nature of text mining permits to extract new knowledge from large amount of unstructured (text) data. This means that text mining has being recently adopted as an emerging technique for a range of purposes and stakeholders (Fan et al. 2006)[1] as a convenient mean to extract knowledge from large sets of documents. In the particular case of scientific research, text mining techniques are increasingly used either standalone or complementarily to other bibliometric techniques to enhance the traditional literature reviews (Ananiadou et al. 2009; Bragge and Storgårds 2007) , supporting the monitoring of research trends in a particular field (White et al. 2016) or reviewing the evolution of a policy-related academic domain (Gómez-Barroso et al. 2016), among other relevant goals. Second, although bibliometric analysis has been usually applied when reviewing the evolution of research on entrepreneurship – not entrepreneurship policy - and there have been relevant contributions to gather the stock of knowledge and identifying main trends within the field (Busenitz et al. 2014; Ferreira, Reis, and Miranda 2015), within the authors' knowledge there is no prior works using text mining techniques to review the entrepreneurship (policy) field. Hence, as text mining is a new approach when reviewing the evolution entrepreneurship

[1] Text mining is more frequently used from the world of business and industry, for example in customer relationship management (Ngai, Xiu, and Chau 2009), business intelligenge (Ishikiriyama, Miro, and Gomes 2015) or financing security (Seo and Park 2018), among many others; to government, either for policymaking (Ngai and Lee 2016) and security or crime detection (Kontostathis, Edwards, and Leatherman 2010; Zanasi 2009); and also for the academic community, opening up opportunities both for scientific discovery and of course in the field of literature mining (A. M. Cohen and Hersh 2005; Fluck and Hofmann-Apitius 2014).

as a research field and the knowledge about the evolution of themes in the particular case of entrepreneurship policy is still rather limited, from the authors' perspective the paper also contributes illustrating the possibilities of these type of tools to complement bibliometrics when mapping any research field.

With the above goals, this article is structured in five main sections. Section 2 briefly reviews the status od the research on entrepreneurship policy, establishing a background for the results of this paper. Next, the methodology section describes the formation process of the corpus of papers and the text mining workflow. It is followed by the results and discussion section that focus on the cluster analysis of the documents. Finally, Section 5 compiles the main conclusions and limitantions, pointing towards new avenues of research.

Literature review

On his well-known collection *What we talk when we talk about love*, Raymon Carver delineates various types of love through 17 short stories in which love is represented experientally through different daily life characters and situations to almost at the end suggest that we human beings are just beginners at love". With all due caution, there seems to be a similar impression when reviewing the last 25 year of entrepreneurship policy research:in spite of the growing body of knowledge in the field, it is still under-investigated (Mazzarol and Volery 2015; Rosa 2013).

Drawing a parallel with entrepreneurship as research field, it has evolved from been defined as *'fragmented'* and *'lacking of a conceptual framework'* during the early 2000s (Shane and Venkataraman 2000), to be recently considered a maturing discipline (Busenitz et al. 2014; Zahra and Wright 2011). But this has not been the case for entrepreneurship policy. In fact, while it is commonly accepted that public policy can influence entrepreneurship (Verheul et al. 2002; Wennekers and Thurik 1999) and there is no one-size-fits-all strategy to stimulate entrepreneurial activity (Minniti 2008), there have been appeared debates about almost everything else related to entrepreneurship and policy initiatives at all level of governance (Bennett 2014; Stam 2014).

That is not to say that there is no relevant entrepreneurship policy literature. From the seminal works authored by Storey (1994), Verheul (2002), Hart (2003) or Lundstrom & Stevenson (2002; 2005), among others, there have been a large number of studies on the topic, ranging from the analysis of specific policy instruments applied to different units of entrepreneurship and/or policy stages and geographies (Z. Acs et al. 2016; Autio and Rannikko 2016; Dennis Jr. 2011) to studies about the potentiality of the public intervention (Figueroa-Armijos and Johnson 2016; Grimm 2006; Van Stel, Storey, and Thurik 2007) or deeper theoretical reviews of the field (Audretsch, Grilo, and Thurik 2007; Bennett 2014)**.** In addition, policies have been also acknowledged as part of the relevant context for entrepreneurship (Welter 2011) and the policy dimension has been identified as one of the key elements of the entrepreneurial ecosystem (Mason,Colin; Brown 2014; Stam 2015).

Entrepreneurship policy has also received remarkable updates from international organizations and institutions such as the OECD[2], European Commission[3] or UNCTAD[4] have developed initiatives to promote. Along these lines and as a main example, the Innovation Policy Platform (IPP) is an on-going initiative—developed by the World Bank and the OECD—with the aim of facilitating knowledge on how innovation systems operate. Moreover, it is a space where institutional users from different regions can share good practices. It contains a specific section about the policy

[2] In 2007, the OECD published a Framework for the Evaluation of SME and Entrepreneurship Policies and Programmes, aiming to help policymakers a practical guide to improve the evaluation stage of entrepreneurship programmes. This publication derived from a recommendation of the 2nd OECD SME Ministerial Conference celebrated in Istambul in 2004. The 3rd OECD SME Ministerial Conference took place in 2018 in Mexico and further recommendations are expected. Available at:
https://read.oecd-ilibrary.org/industry-and-services/oecd-framework-for-the-evaluation-of-sme-and-entrepreneurship-policies-and-programmes_9789264040090-en#page1

[3] The EC and the OECD has launched in 2018 the 'Better Entrepreneurship Policy Tool' to support the inclusive and entrepreneurship including a self-assesment tool and guidance to improve the design of these policies.

[4] UNCTAD developed in 2018 the 'Entrepreneurship Policy Framework' with the aim of support the promotion of entrepreneurship in developing countries. Available at: https://unctad.org/en/Pages/DIAE/Entrepreneurship/Entrepreneurship-Policy-Framework-and-Implementation-Guidance.aspx

rationales and objectives for innovative entrepreneurship, a comprehensive explanation about the main market and system failures, and specific scenarios that require policy intervention[5].

Despite these contributions, the construction of the entrepreneurship policy body of knowledge is still in progress.. At this point, the main motivation of this paper is not to advice about the future configuration entrepreneurship policy as a research topic, but to unwrap the past by unveiling how key themes of the entrepreneurship policy research agenda have emerged, evolved and / or declined over time as a foundation on which to build further developments.

Methodology

This study uses text mining to analyse a collection of scientific papers related to entrepreneurship policy. Text mining and/or text data mining are a category of data mining techniques. Like data mining, text mining aims to discover new knowledge or hidden patterns from large amounts of data (Delen and Crossland, 2008). The particularity is that text mining processes natural language text or textual content instead of structured regular data, so it requires some pre-processing tasks. The most common goals of text mining are natural language processing and text representation, concept/entity extraction, search and information retrieval, web mining, information extraction, document classification/ categorisation, document clustering and sentiment analysis (Miner et al., 2012). This paper concentrates on concept extraction, document clustering and document classification. Particularly, results within the paper extract the key concepts and provide a basis for the categorisation of main themes that have shaped the research agenda on entrepreneurship policy of the journals under study over time.

This methodology section describes how text mining is applied to achieve these objectives: (i) the text data collection and the formation of the corpus of academic papers; (ii) the pre-processing tasks to transform the unstructured text data; and (iii) the specific text mining techniques applied to conduct the analyses. For the selection of methods and techniques applied in the analysis, robustness and maturity have prevailed as criteria over other most recent and less tested options.

The analyses within the paper have been conducted using the text analytics extension of RapidMiner Studio,[6] an open source software written in Java that supports a range of applications related to data science, complemented by Microsoft Excel. RapidMiner is one of the most popular tools in the field of data and text analytics. Among other advantages are the ease of use and the quality of the official documentation, the huge adoption rate, and the support provided by the online community. In addition, a free version is available under copyleft licensing to process up to 10,000 data rows.

Text collection and definition of the corpus

The analysis considers a representative selection of literature that explicitly refers to the views of academia on entrepreneurship policy, named the Academia corpus. In particular, the corpus consists of a selection of 1,048 articles from ten significant academic journals in the field during the period 1990–2016. The process to form the corpus is detailed below.

The Academia corpus is formed by all the policy-based articles from ten academic journals related to entrepreneurship research which are being published from at least the 1990s onwards and have been always ranked in the first or second quartile of the Scopus index. Although all of the entrepreneurship policy literature was not analysed, the ten journals included are considered to be representative of mainstream research on the field entrepreneurship. The distribution of papers per journal is shown in Table 1.

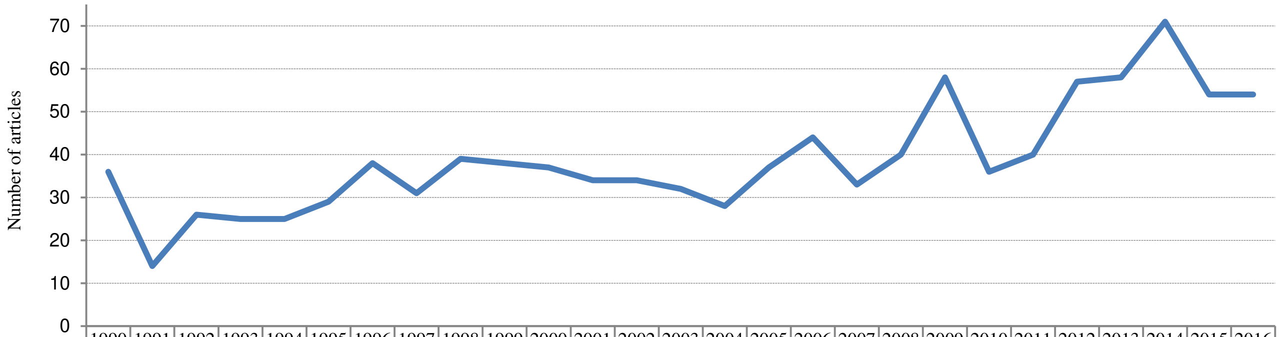
An initial search[7] in these journals generated 1,190 items related to policy—1,048 articles, 92 reviews, 26 articles in press, 15 conference papers, and 9 editorials and errata. Following a similar approach than other text mining studies

[5] Available at: https://www.innovationpolicyplatform.org/

[6] Documentation for RapidMiner Studio and a free version of the software under copyleft licensing can be downloaded at www.rapid-i.com

[7] Results from a query in Scopus database on 30 March 2017. In particular, the query used to retrieve all the documents related to policy and published in those ten journals during 1990-2016, was:

(Basole et al., 2013; Chang and Katrichis, 2016; Ngai and Lee, 2016), the analysis in this paper considers only the 1,048 scientific articles, corresponding to 1,339 author affiliations, excluding the other items to guarantee higher level of rigour.



	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Articles	36	14	26	25	25	29	38	31	39	38	37	34	34	32	28	37	44	33	40	58	36	40	57	58	71	54	54

Fig 1 shows the distribution of published articles per year. The evolution of the number of papers over the years indicates a growing number of articles in the journals under analysis, especially since the mid-2000s.

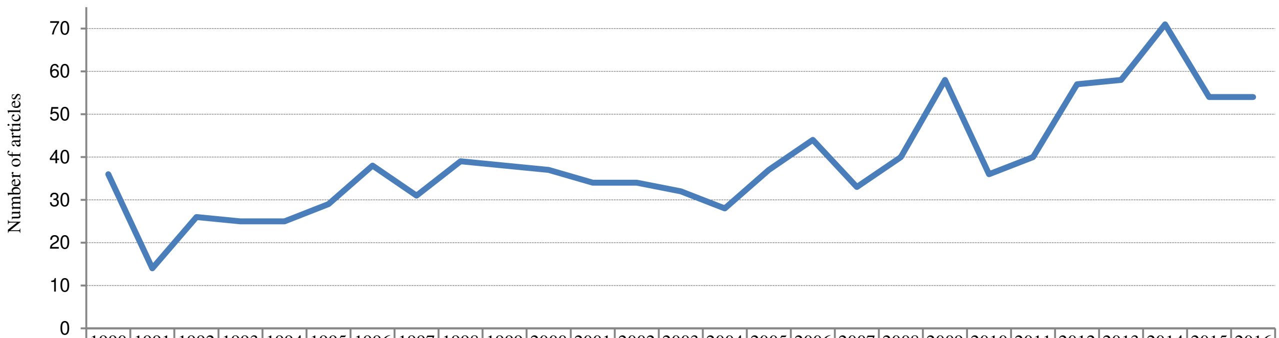
Table 1 - Journals included in the sample (1990–2016).

Journal	Number of articles in the sample	Scopus SJR	Scopus H Index
Technovation	292	1,738	102
Small Business Economics	240	2,116	98
Entrepreneurship and Regional Development	174	1,581	69
Journal of Business Venturing	85	5,513	140
International Business Review	82	1,180	73
Organization Science	49	7,151	196
Journal of Small Business Management	42	1,504	85
Academy of Management Journal	37	11,053	266
Journal of Management	33	6,353	176

[TITLE-ABS-KEY (policy) AND (LIMIT-TO (DOCTYPE , "ar ")) AND (LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012) OR LIMIT-TO (PUBYEAR , 2011) OR LIMIT-TO (PUBYEAR , 2010) OR LIMIT-TO (PUBYEAR , 2009) OR LIMIT-TO (PUBYEAR , 2008) OR LIMIT-TO (PUBYEAR , 2007) OR LIMIT-TO (PUBYEAR , 2006) OR LIMIT-TO (PUBYEAR , 2005) OR LIMIT-TO (PUBYEAR , 2004) OR LIMIT-TO (PUBYEAR , 2003) OR LIMIT-TO (PUBYEAR , 2002) OR LIMIT-TO (PUBYEAR , 2001) OR LIMIT-TO (PUBYEAR , 2000) OR LIMIT-TO (PUBYEAR , 1999) OR LIMIT-TO (PUBYEAR , 1998) OR LIMIT-TO (PUBYEAR , 1997) OR LIMIT-TO (PUBYEAR , 1996) OR LIMIT-TO (PUBYEAR , 1995) OR LIMIT-TO (PUBYEAR , 1994) OR LIMIT-TO (PUBYEAR , 1993) OR LIMIT-TO (PUBYEAR , 1992) OR LIMIT-TO (PUBYEAR , 1991) OR LIMIT-TO (PUBYEAR , 1990)) AND (LIMIT-TO (EXACTSRCTITLE , "Technovation ") OR LIMIT-TO (EXACTSRCTITLE , " Small Business Economics ") OR LIMIT-TO (EXACTSRCTITLE , " Entrepreneurship And Regional Development ") OR LIMIT-TO (EXACTSRCTITLE , " Journal Of Business Venturing ") OR LIMIT-TO (EXACTSRCTITLE , " International Business Review ") OR LIMIT-TO (EXACTSRCTITLE , " Organization Science ") OR LIMIT-TO (EXACTSRCTITLE , " Journal Of Small Business Management ") OR LIMIT-TO (EXACTSRCTITLE , " Academy Of Management Journal ") OR LIMIT-TO (EXACTSRCTITLE , " Journal Of Management ") OR LIMIT-TO (EXACTSRCTITLE , " Academy Of Management Review "))]

Academy of Management Review	14	9,056	229

Source: Scopus SJR and Scopus H-Index retrieved from SCImago Journal Rank 2015. For the purpose of the analysis, journals under study have been always placed within the Q1 and Q2 in their respective subject categories since 1990s. More information available at: https://www.scimagojr.com/journalrank.php?year=2015



	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Articles	36	14	26	25	25	29	38	31	39	38	37	34	34	32	28	37	44	33	40	58	36	40	57	58	71	54	54

Fig 1-Evolution of the total number of articles related to entrepreneurship policy within the ten journals analysed

The corpus derived from these articles consists of merging document titles, abstracts and keywords that are the most common summaries in scientific papers (HaCohen-Kerner, 2003). This approach for the analysis of academic papers is a standard and broadly used approach in text data mining and bibliometric studies (Felizardo et al., 2011; Glanzel, 2012; Liu et al., 2010). It is also customary not to include the conclusions section, as it is thought that the title describes the main theme of the paper, the abstract is a good summary of the background and the results, and keywords consolidate the essence of the content. In addition, the editorial and peer review process of these journals ensures their quality, so this choice is a fair representation of the articles.

Pre-processing and stemming processes

The corpus comes from electronic textual documents: they are unstructured content requiring at least two sets of preparatory processes before the text mining analysis, specifically, a set of pre-processing tasks and then a stemming process.

Following standard practices in text mining (Miner et al., 2012), the pre-processing tasks in this paper include: (i) tokenisation of the document to chop it into indivisible words/phrases pieces or 'tokens'. These tokens can be defined as a sequence of characters in a particular text that are gathered as semantic units for processing, typically forming a term or a word; (ii) data cleaning to remove certain characters without meaning, such as punctuation, non-alphanumeric characters and numbers; (iii) normalisation, to convert words to lower case; (iv) filtering out general English stop words, such as articles, pronouns or prepositions, since they are the most common words and contribute little to the overall meaning. To this end, it is necessary to compare the set of tokens to a predefined dictionary;[8] and (v) filtering out those tokens that contain fewer than a certain number of characters, in the case of this paper three characters. The output of these five preparatory steps is a tokenised corpus that, in turn, needs to be stemmed.

[8] This pre-built list is provided by RapidMiner and is formed of 452 English common terms such as 'the', 'now' or 'yourself' that are not relevant for the analysis. The whole list is configured as a Java class of the library WV Tool. See: https://sourceforge.net/projects/wvtool

Stemming is the process of removing affixes from a word and grouping those words with a common root to obtain stem words. As an example, singular and plural nouns and verbs with different conjugations are reduced respectively to the same root (stems). The analysis within this paper applies the Snowball English Porter stemmer, an advanced version of the Porter stemmer (Porter, 2001). This is one of the most common stemming algorithms in the field of text mining. It is basically a truncating method based on the idea that the suffixes in the English language are mostly composed by grouping smaller and simpler suffixes and, for this reason, they can be removed under certain conditions while remaining meaningful.

Pre-processing for the Academia corpus amounted initially to 11,902 root terms, which were reduced to 7,722 tokens after stemming.

Data Analysis

Preliminary results

After these preparatory processes, the text mining workflow begins. The first step is to convert the unstructured text data into structured data by creating the word vectors that numerically define the documents from the tokens. To this aim, this paper uses the Term Frequency-Inverse Document Frequency (TF-IDF), a statistic that measures the relative relevance of a word in a document from a corpus. The TF-IDF value for each word in a document is proportional to the number of times a word appears in the document but it is offset by the percentage of documents in the corpus that contain the word. This helps to evaluate the real relevance of a word considering that, in general, some words appear more frequently than others but not necessary more meaningful. As a result, the process obtains a matrix formed by word vectors representing the relationship between words and documents through their TF-IDF values.

Then, the first analysis within the paper is the examination of most repeated concepts within the corpus Academia as a whole considering the whole period 1990-2016 and also in five-year terms. As additional analysis, the paper examines the most relevant words considering three regional sub-corpuses, distinguishing those articles authored by SE Asia, EU and US, respectively.

Cluster analysis

The second step of the text mining process in this paper is the cluster analysis of the corpus. In general, clustering is one of the most common processes in data mining analysis. In the particular case of text mining, text clustering aims to divide a collection of documents into coherent groups with similar content–clusters-(Hotho, Nürnberger, and Paaß 2005; Manning, Ragahvan, and Schutze 2009). To do it, clustering algorithms intend to optimize the resulting partition by both minimizing the inter-cluster similarity and mazimizing the intra-cluster similarity. So there are two key aspects to decide to conduct a cluster analysis: the similarity function and the clustering algorithm itself.

Similarity is quantified by using a similarity function. For the purpose of the analysis, this paper has opted by the Euclidean distance, which is one of the most popular and frequently used techniques for normalized vectors (Hotho, Nürnberger, and Paaß 2005; Jain 2010), The clustering algorithm defines how to classify the whole collection of documents into clusters. From the range of clustering algorithms that could be used and considering the size of the sample, this paper has chosen the standard k-means, a partitioning clustering algorithm widely used in data mining that presents an optimal trade-off in terms of effectiveness (clustering performance (effectiveness) and eficiency (computational requirements).

The funcioting of k-means could be summarized as follows. It begins by choosing *k* initial centroids or cluster centers and then documents (items of the collection) are assigned to the nearest of these k centroids, aiming to optimize the similarity function, in this case, the goal of k-means is to minimize the average squared Euclidean distance of documents from their respective cluster centers or centroids.Then, new centroids are calculated again according to the initial distribution. The process is repeated a number of times until there are no changes among the cluster centroids. As the number of these iterations is always limited because of the computational resources, the initial assignation of centroids (seeding) is very relevant in the case of k-means, For this reason, this paper has used a randomised seeding technique with a maximum of fifteen iterations of the algorithm.

As partitioning algorithm, the main issue of k-means is that the number of clusters (k) is an input parameter that must be assigned by the researcher. There are some methods to relieve the potential bias of this a priori choice. In this case, the paper has combined both qualitative evaluation of the results for the k values from four to thirteen with the Davies-Bouldin criterion, that is, calculating the values for the Davies-Bouldin index[9] of these clustering results, considering that the lower value of Davies-Bouldin index means a better clustering result.

.

[9] The Davies-Boulding index (Davies and Bouldin 1979) is a metric to evaluate clustering partitions, suitable for those cases as k-means where the value of k is not known a priori (Arbelaitz et al. 2013)

Results of the analysis and discussion

Descriptive statistics

For the purpose of the analysis, corpus Academia is also splitted to include a geographical dimension in the study, particulary to analyse those papers authored by researchers affiliated in the European Union (EU), in the United States (US) and in the South and Eastern Asia (SE Asia). This analysis is only a snapshot of the whole entrepreneurship research in the considered geographies and it is obviously limited by the sample and the fact that authors´ current affiliation is only one of the aspects of the context, which is much more complex. But it aims to give some additional insights to enrich the debate about how have been configured the research on entrepreneurship policy per region since the early 1990s and the contributions to the whole research picture.

From the whole corpus Academia, there are 558 documents corresponding to EU affiliations [10], 340 documents corresponding to US affiliations and 118 documents with SE Asia affiliations[11]. Considering that the total sample is formed by 1,048 articles, European Union largely leads the research in this field for the period under study with EU affiliated authoring the 53.2% of the papers, followed by US affiliated authors participating in almost one third of the total amount of papers and finally South and Eastern Asia affiliated authors taking part in the 11.3%.

Fig. 1 shows how regional production of papers has evolved since 1990. Those papers authored by EU and US affiliated researchers present a similar distribution, gathering around a quarter of the documents during the 1990s decade, a third during the 2000s and the rest since 2010s. On the other hand, the distribution of SE Asia affiliated papers is lower during the 1990s to explode during the initial lustrum of the 2000s; and then presenting a similar balance than their EU and US counterparts during 2005-2014 to finally present a decrease during the last period.

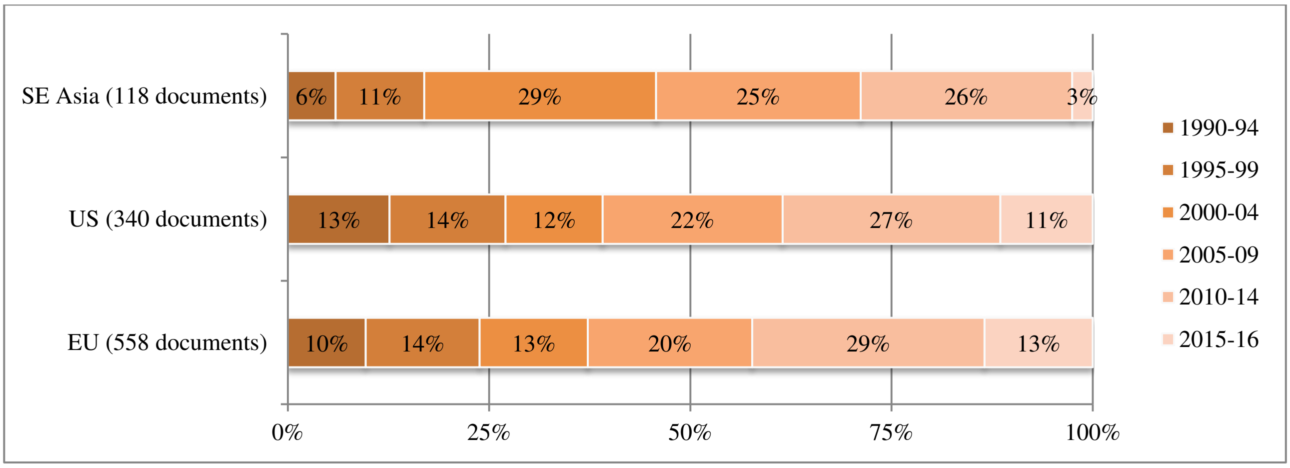


Fig. 1 Regional distribution of papers in SE Asia, the EU and US sub-corpuses per five-year terms (1990-2016)

Fig. 2 shows the distribution of papers taking into account the different combinations of authors´ affiliations for the three regions under study. As a brief summary of the main results, the percentage of SE Asia with one unique country-affiliation is around 63% while this percentage rises up to around 70% in those papers affiliated in the EU and US there is only one paper co-authored simultaneously by SE Asia, EU and US affiliated authors. Going into the detail about inter-

[10] In the case of EU, United Kingdom largely leads the research in the region for the period under study with more than 35% of the total number of affiliations, followed by Netherlands, Spain, Italy, Germany, Sweden and France, all of them together adding up to almost half of the EU authors´ affiliations. Then follows a group of six countries with at least ten affiliated papers (Belgium, Finland, Denmark, Greece and Austria), and finally a long tail of countries up to complete the twenty four countries with at least one affiliated paper. Slovakia, Latvia, Lithuania and Malta do not count with any affiliated researcher within the Corpus during 1990-2016.

[11] In particular, papers in the sample corresponds to authors affiliated in China, Hong Kong, India, Japan, Malaysia, Philippines, Singapore, South Korea, Taiwan and Thailand. South Korea and Japan affiliated authors add the 20% and the 18% respectively of the total production in the region followed by China (14%), India (13%), Singapore (10%) and Taiwan (9%), Thailand (4.8%), Malaysia (3%) and Philippines with one unique affiliated paper (0.8%).

regional and intra-regional collaborations, it should be noted that there are hardly two papers co-authored by different SE Asia affiliations while papers with SE Asia – US and SE Asia – EU represent respectively a 13.6% and a 15.3% of the regional sample. In the case of EU affiliated papers, there are a wider number of papers co-authored between EU and US affiliated authors than those elaborated only with other EU counterparts. Similarly, in the case of those US affiliated papers, collaboration with EU affiliated researchers represent a 17.9% of the total number of papers. Finally, results show that the percentage of collaborations with other affiliated authors is similar in the case of the three regions under study, with a slightly lower rate in the case of SE Asia papers.

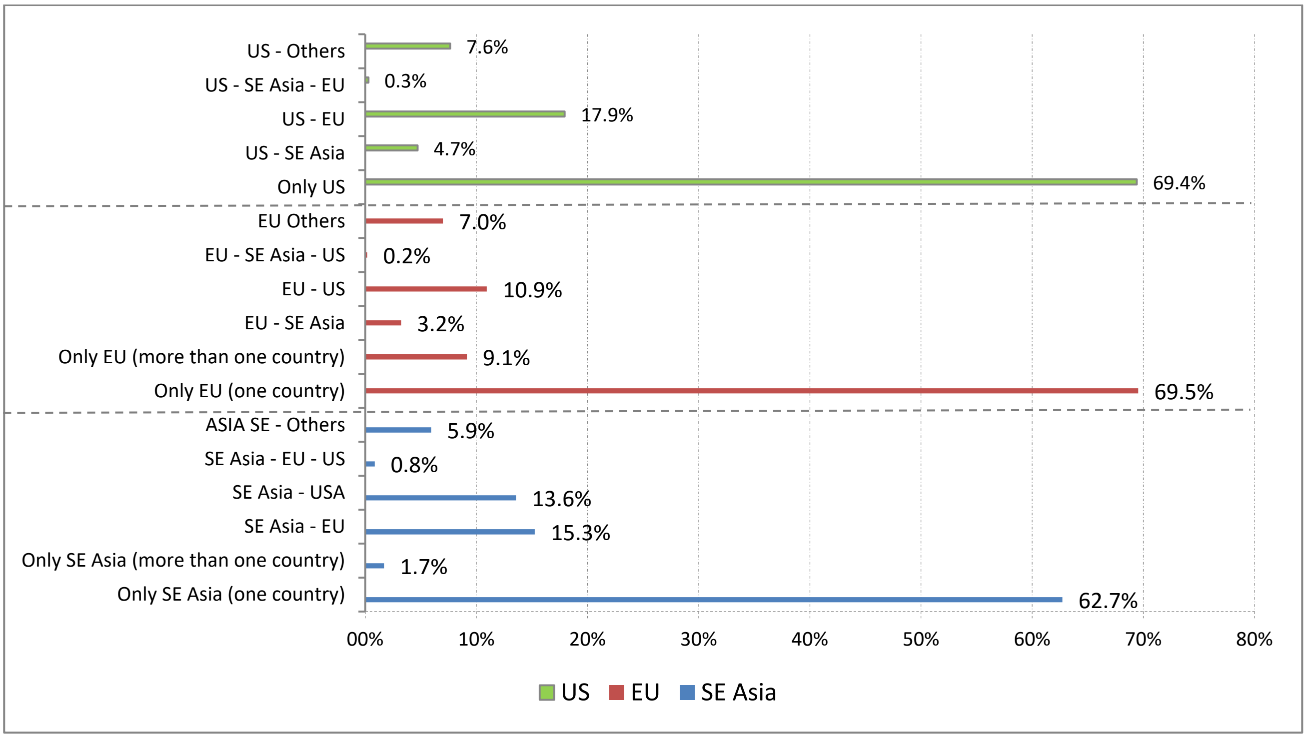


Fig. 2 Regional distribution of papers per type of affiliation within SE Asia, EU and US sub-corpuses (1990-2016)

Table 2 shows the total distribution of papers per five-year terms in each region compared to the evolution of the different types of collaborations during the same periods. Results are rather homogenous, with a peak of collaborations of SE Asia-EU, SE Asia-US, EU-US and even EU inwards during 2005-2014. In the case of EU-US collaborations, this tendency continues to be incremental during the last two years, with around a 25 of the 61 shared papers published in 2015 or 2016.

	Total number of papers	1990-94	1995-99	2000-04	2005-09	2010-14	2015-16
SE Asia	**118**	**6%**	**11%**	**29%**	**25%**	**26%**	**3%**
Only SE Asia (one country)	74	7%	12%	38%	18%	24%	1%
Only SE Asia (more than one country)	2	0%	0%	100%	0%	0%	0%
SE Asia - EU	18	0%	11%	11%	33%	44%	0%
SE Asia - US	16	0%	6%	13%	44%	31%	6%
SE Asia - EU - US	1	0%	0%	0%	100%	0%	0%
SE ASIA - Others	7	29%	14%	0%	43%	0%	14%
EU	**558**	**10%**	**14%**	**13%**	**20%**	**29%**	**13%**
Only EU (one country)	388	13%	18%	16%	19%	24%	10%
Only EU (more than one country)	51	2%	4%	6%	24%	41%	24%
EU – SE Asia	18	0%	11%	11%	33%	44%	0%
EU – US	61	2%	7%	5%	21%	41%	25%

EU – SE Asia - US	1	0%	0%	0%	100%	0%	0%
EU Others	39	3%	8%	8%	21%	36%	26%
US affiliated papers	**340**	**13%**	**14%**	**12%**	**22%**	**27%**	**11%**
Only US	236	18%	18%	14%	21%	22%	7%
US – SE Asia	16	0%	6%	13%	44%	31%	6%
US - EU	61	2%	7%	5%	21%	41%	25%
US – SE Asia- EU	1	0%	0%	0%	100%	0%	0%
US - Others	26	0%	8%	15%	19%	35%	23%

Table 2 Distribution of SE Asia, EU and US affiliated papers by type of collaboration, in five-year terms

Therefore, SE Asia affiliated authors seem to be more inclined to collaborate with the outside world, but they tend to collaborate more with authors from US or EU countries than with surrounding countries. In addition, EU and US affiliated papers represent the most common external collaboration, in the case of the EU even with a higher representation than those papers with only EU authors. From a chronological dimension, there is a common turning point to the increasing of collaboration among different affiliations since the mid-2000s with the highest rate of co-authoring both outward and inward during the period 2010-2014.

Preliminary analysis

The preliminary analysis consists of the absolute word frequency, measured as the total occurrences of a word within corpus Academia after having pruned out irrelevant elements and consolidating those terms with similar meaning by using the Porter stemmer as explained in previous methodology section.
Fig. 3 displays the most frequent words in this corpus. Terms related to the business perspective are placed in high positions, with *firm*, *enterprise* and *company* having the highest number of occurrences; while *entrepreneur* and *business* are the 3rd and 4th most frequent words, respectively. The term *policy* ranks 2nd. Altogether, these terms are consistent with the aim of this research paper. In a similar manner, *technology* and *innovation* appear at the 6th and 7th position, respectively. Along the same lines, *knowledge* is at 18th, *science* at 22nd and *data* emerges at 36th. Intertwined with technology are those themes linked to finance and socio-economics, with *capital*, *venture*, *investment* and *finance* placed at 23rd, 26th, 28th and 31st, respectively. Furthermore, *development* outlined at 5th, with *economy* at 8th, *industry* at 9th, *growth* at 14th, *market* at 15th, *social* at 40th and *employment* at 34th.

An interesting insight is the position of *size* at 49th, showing the interest of academia on firms' size as a key aspect of the contribution of entrepreneurship to productivity and innovation. The institutional perspective includes *govern* at 16th, *institutional* at 17th, *network* at 27th, *system* at 33rd and *competition* at 41st. Collectively, the technological, socio-economic and institutional viewpoints, with no clear pre-eminence among them, seem to point toward multiple perspectives that are required to study entrepreneurial policy from the academic perspective. However, there are no significant references to terms related to education as mentioned above. As an additional note from a geographic viewpoint, *region* is placed at 12th, the aggrupation of *nation* and *country* is at 13th, while *international* is only located at 30th. Such terms are an early indication that most of the academic analysis refers to specific geographical areas, or is at least influenced by such.

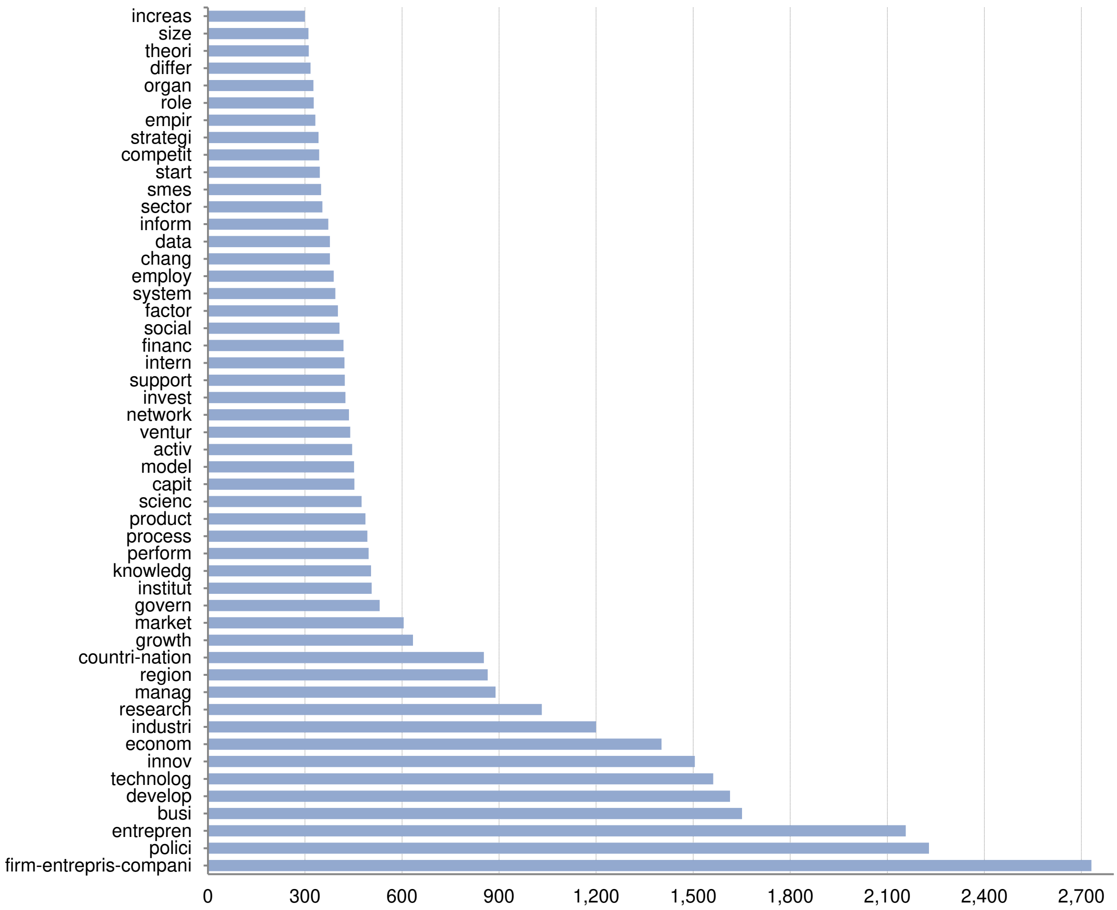


Fig. 3 Top 50 frequent word stems within the corpus Academia by absolute number of occurrences (1990–2016)

An analysis was applied to each five-year period, since 1990–2016, as a relevant input of the study. The results are available in the Apendix (Figs. 1–6).

Next, as in the case of the Global Academia Corpus, the paper analyses the absolute word frequency within the three regional sub-corpuses, distinguishing those articles authored by SE Asia, EU and US, respectively, and compares them with the results of the global picture and also related to each other.

Southern and East Asia Academia Sub-Corpus

Fig. 4 shows the most frequent terms in the Southern and East Asia Academia Sub-Corpus. In this case, *technology* is the most repeated term, rising from the 6th position in the Global Academia Corpus. Similarly, *innovation* also goes up to the 5th position from the 7th when considering the whole corpus.

On the contrary, the three most repeated terms in the whole corpus decline here. Concretely, *firm-enterprise-compani, policy* and *entrepreneur* decrease to the 2nd, 3rd and 10th positions respectively. Looking at other common agents of entrepreneurship, *smes* scales from the 40th in the whole corpus to the 28th position within the Southern and East Asia affiliated papers, while *startup* declines to the 243rd position here from the 40th in the Global Academia Corpus.

Growth, *social* and *employment* go down dramatically within de Asian Sub-Corpus, ranking 33rd, 79th and 103rd positions respectively while they are the 14th, 32nd and 35th most repeated terms in the Global Academia Corpus. In addition, *industry* raises three positions to the 6th but *economy* decreases to the 8th from the 9th in the whole corpus and looking at the business dimension, *business* declines from 3rd to the 16th and *management* keeps at the 11th position.

The contribution of terms related to research and knowledge is prominent in those papers authored by Southern and East Asia affiliated researchers. *Research*, *knowledge*, *transfer* and *university* go up to 8th, 18th, 48th and 67th positions respectively from the previous 10th, 18th, 64th and 80th, while *science* keeps the 22nd position. From the financial dimension, *venture* and *investment* rise here to the 25th and 29th from the 26th and 19th in the whole corpus, while finance decreases to the 72nd.

From an institutional perspective there are also differences comparing to the Global Academia Corpus since *govern*, *system* and *competition* rank 13th, 14th and 16th positions here, raising from the previous 16th, 34th and 42nd .

Finally, there are interesting insights when looking at the geographical dimension. *Country-nation* rises to the 7th position from the 13th in the whole corpus, but *region* declines to the 29th from the 12th. This is reinforced with the high positions of the specific names of the Asian countries under study, for instance. China and Korea are in the top 50. Besides, *international* and *foreign* also go up to the 24th and 42nd positions. Focusing on those terms related to context/location, both *network* and *cluster* rise to the 26 and 46th positions from the 27th and 57th within the whole corpus.

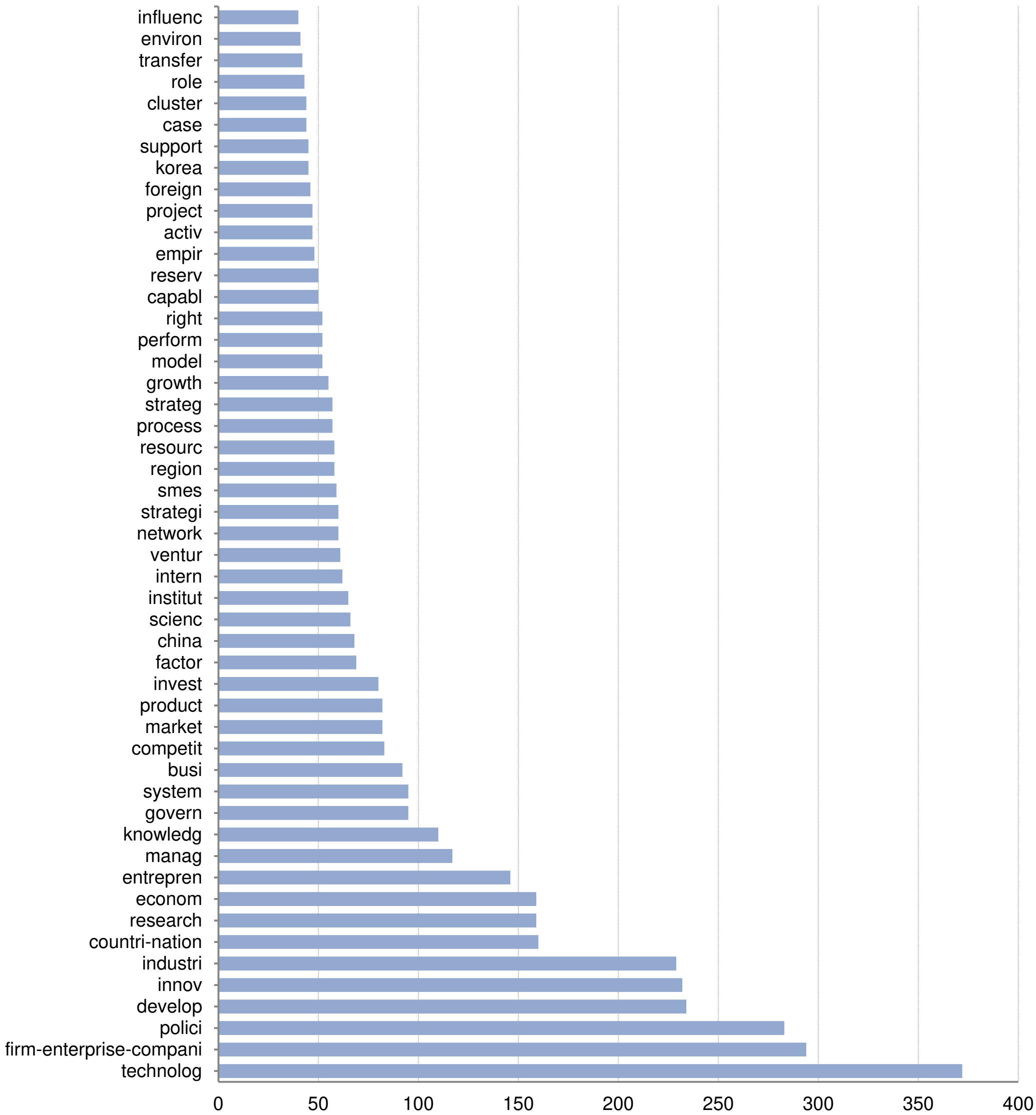


Fig. 4 Top 50 frequent word stems within the corpus SE Asia Academia Sub-Corpus by absolute number of occurrences

EU Academia Sub-Corpus

Fig. 5 displays the most frequent terms in the EU-Academia Sub-Corpus. Words in the top ten are very similar to the Global Academia Corpus with the minor exceptions of *innovation*, which rises from 7th up to 4th position when considering only papers with EU authors; and also *technology* that goes down from 6th position considering all the papers to the 8th position when considering papers with just EU authors.

A more detailed analysis about the performance of entrepreneurial agents in the corpuses offers interesting insights. While there are no differences for the 1st position of *firm-enterprise-company* and 3rd position of *entrepreneur*, positions of the terms *startup* and overall *smes* are higher when considering papers with authors with EU affiliations. However, *size* is at 49th within the analysis of the whole Academia corpus and at 66th from the EU Academia point of view, showing a potential less interest from authors affiliated to EU institutions on firm´s size in the context of entrepreneurship policies.

Other significant differences within the most frequent words are *employment* that ranks at 27th in the EU Academia Corpus and 35th at Global Academia Corpus and *knowledge*, *science* and *transfer*, ranking 16th, 18th and 44th in the EU Academia and 18th, 22nd and 64th respectively within the whole Academia Corpus. Similarly, terms related to finance are more relevant for EU Academia, with *venture*, *investment* and *finance* at 22nd, 23rd and 24th when considering papers related to EU authors and the same terms at 26th, 28th and 31st, respectively within the whole Academia corpus.

On the contrary, the business dimension underperforms when analysing the EU Academia comparing to the global analysis, with *business* and *management* ranking 5th and 14th versus 4th and 11th within the whole corpus

Regarding to institutional perspective, terms such us *govern*, *system*, *competition* ranks at 16th, 34th and 42nd in Global Academia Corpus, while the same words descend to 34th, 35th and 61st within the EU Academia Corpus.

The comparative analysis of words related to geography and context also offers interesting differences between EU Academia and Global Academia. While terms referring to geographical level such as *international*, *nation-country* and *region* are in slightly in an upper position in the case of EU Academia, differences are bigger when considering both generic terms related to context and/or location such as *context*, *agglomeration* or *geography* and also a greater level of detail with terms such *local*, *cluster*, *network* or *relationship*.

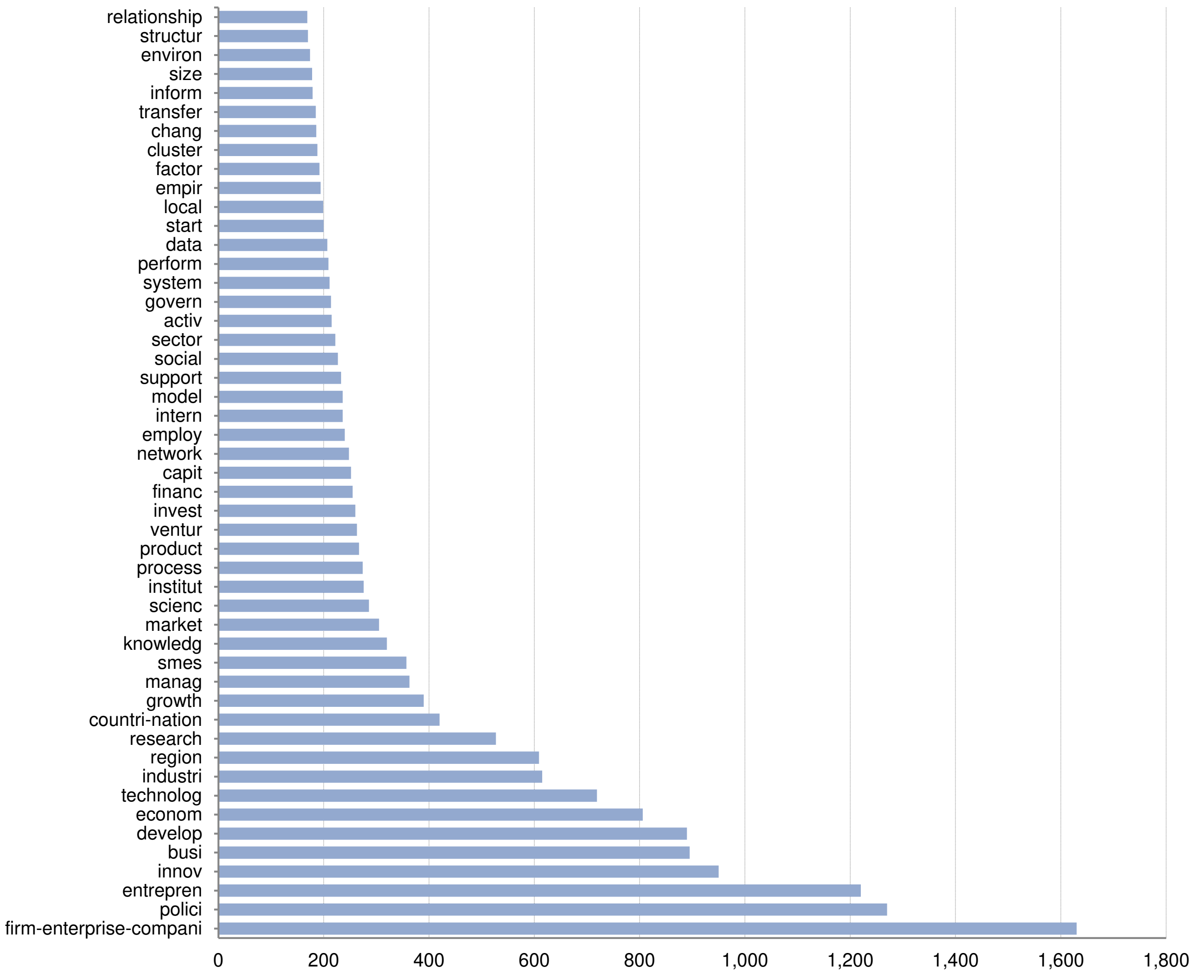


Fig. 5 Top 50 frequent word stems within the corpus EU Academia Sub-Corpus by absolute number of occurrences

US Academia Sub-Corpus

Fig. 6 shows the most frequent terms in the US-Academia Sub-Corpus. *Entrepreneur* replaces *firm-enterprise-compani* as the most repeated word within the papers authored by US affiliated researchers, what is reinforced by the rise of individual which is here in the 36th position from around the top 100 within the Global Academia Corpus. On the contrary, while *startup* slightly falls from 44th to 40th, the role of *smes* comes down dramatically here to 292nd position from 41st within the Global Academia Corpus

Within the top ten most repeated words is also remarkable that *policy* decreases now from 2nd position to 4th, and both *technology* and *innovation* go down from 6th and 7th when considering the whole Academia Corpus to 7th and 11th respectively when analysing those papers authored by US affiliated researchers. On the other hand, *business*, *economy* and *management* raise here to 3rd, 5th, 8th position from 4th, 8th and 11th respectively in the Global Academia Corpus. Similarly, *market* continues this tendency gaining a position and becomes the 14th most repeated word while *growth* descends from 14th in the global corpus to 19th within US affiliated papers.

While terms such as *employment*, *social* and *support* keep the same positions in both US Sub-Corpus and in the Global Academia Corpus, there are significant rises in the case of *environment* and *culture*, placed at 26th and 57th when considering papers authored by US affiliated authors.

There are divergences when analysing terms related to finance: while *finance* and *investment* comes down to 46th and 52nd considering those US affiliated papers from 31st and 28th respectively within the Global Academia Corpus, *venture* goes up now to 13th position from 26th in the global analysis. Similarly, *research* and *university* rank here 9th and 41st and 10th and 80th in the Global Academia Corpus, *science* and *knowledge* decrease to 27th and 45th within the US Academia Sub-Corpus from 16th and 18th when revising the whole Academia corpus.

From a geographical perspective, there are significant differences between the Global Academia Corpus and the US Sub-Corpus. Country-nation and international rank here 12th and 25th, going up from the 13th and 30th position within the global analysis. The rise of *global* strengthens the international dimension of the analysis in US affiliated papers. Coherently with the domestic administrative organization in US, *region* ranks 54th position here, decreasing from 12th within the Global Academia Corpus, but *state* raises to 40th position while ranks 110th in the global analysis. Looking at those terms related to context and/or location such as cluster, local, agglomeration or network, all of them rank lower positions within the US Sub-Corpus than in the Global Academia corpus.

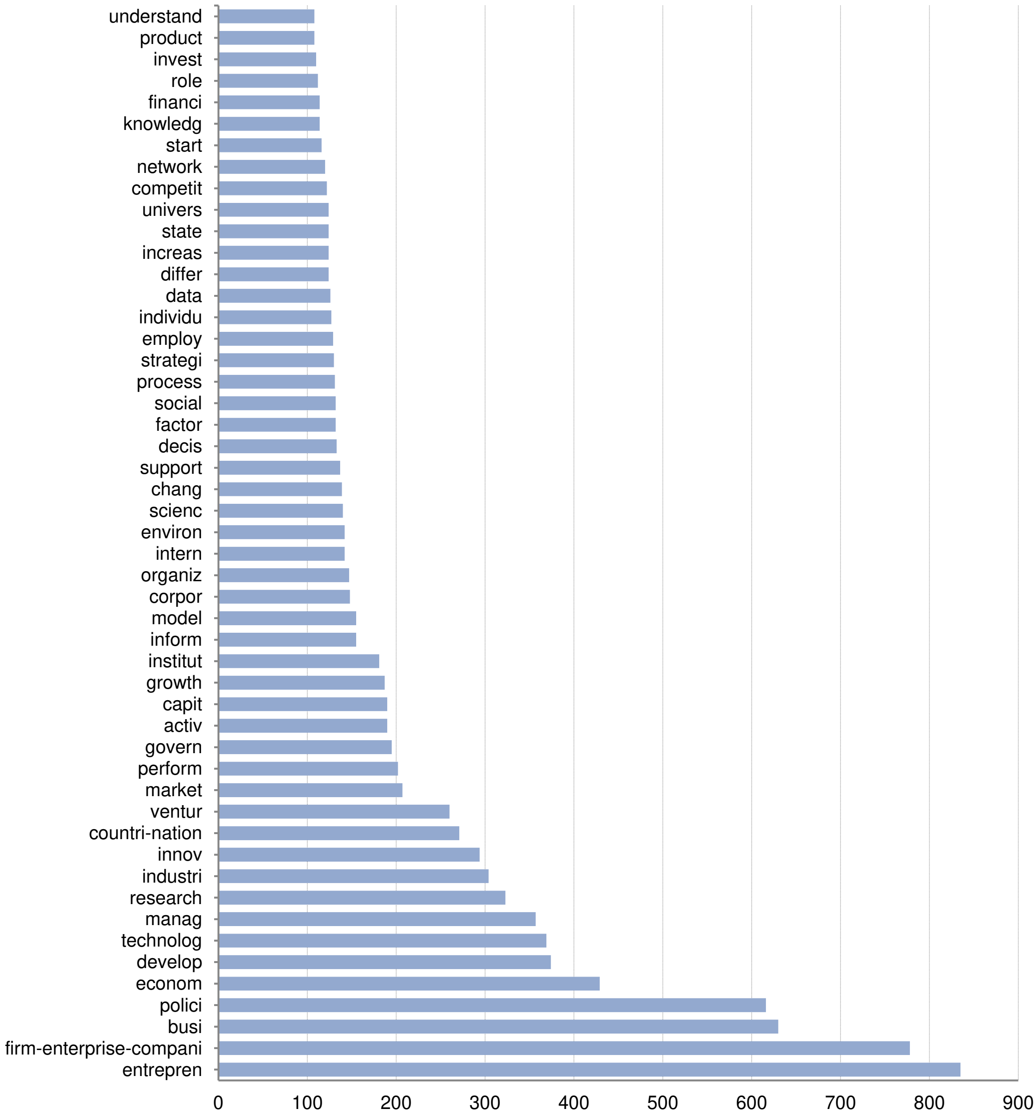


Fig. 6 Top 50 frequent word stems within the corpus US Academia Sub-Corpus by absolute number of occurrences

Cluster analysis

The next stage in the analysis is conducting thek-means clustering... As mentioned in the methodology section, the number of clusters (k) should be specified a priori by the researcher. , so Hence, the analysisstarts with the evaluation of the results of the Davies-Bouldin criterion for the k-means clustering results for k values from four to thirteen and then the qualitative examination of the cluster results.

Corpus Global Academia

The Davies-Bouldin index presents the best (minimum) values for nine, eleven and twelve number of clusters using the K-Means clustering algorithm. Therefore, in the following, this paper studies the results from four to twelve clusters.

When the number of clusters is four, *Innovation–Technology–Research*, *Geography/Location*, entrepreneurial process (i.e., *Agents–Finance–Venture Capital–Regulation–Employment*) and *Management* represent the main themes. A new theme, *Internationalization* (*Investment–Export*), emerged in the case of five clusters. Then, the six-cluster analysis separates the general theme of *Agents–Finance–Venture Capital–Regulation–Employment* into issues related to *Agents–Inclusion* and *Employment–Regulation–Finance–Venture Capital*. Seven clusters introduce *Market–Sectors* as a new topic. The eight-cluster partition splits the *Internationalization* theme in two topics—*Investment* and *Export*. In addition, the eight-cluster partition introduces a new specific theme, *Research–Knowledge*, which was separated from *Innovation–Technology*. The nine-cluster partition added a new topic, *Support*. In the ten-cluster partition, *Geography/Location* is divided in *Agglomeration*, which is mainly related to *Innovation–Technology*. Also, the theme *Management* emerges again and adds a new topic, *Policy*. The eleven-cluster partition shows three *Geography/Location* related clusters: (i) *Agglomeration*; (ii) *Lifecycle*; and (iii) *Innovation–Technology* or knowledge-based *Industry*. The twelve-cluster partition summarises and groups the themes as follows: (i) *Inclusion*; (ii) *Research–Knowledge*; (iii) *Internationalization–Investment*; (iv) *Geography/Location–Lifecycle*; (v) *Geography/Location–Industry*; (vi) *Internationalization–Export*; (vii) *Support*; (viii) *Innovation–Technology*; (ix) *Employment–Regulation–Finance*; (x) *Geography/Location–Agglomeration*; (xi) *Management*; and (xii) *Venture Capital*. From the authors' perspective, this twelve-cluster partition is a rather meaningful combination of the main themes in entrepreneurship research from academia.

Results of the analysis for the twelve-cluster partition of the corpus Academia are shown in Fig. 1, with the defining topics for each cluster, and in Table 1 with a breakdown of the number of papers per cluster. *Inclusion* leads the ranking on the number of papers, followed closely by *Employment–Regulation–Finance*; and, at a distance, by *Management* and *Research–Knowledge*. The results of the other cluster partitions are shown in Tables 10-18 in the Online Supplemental Data.

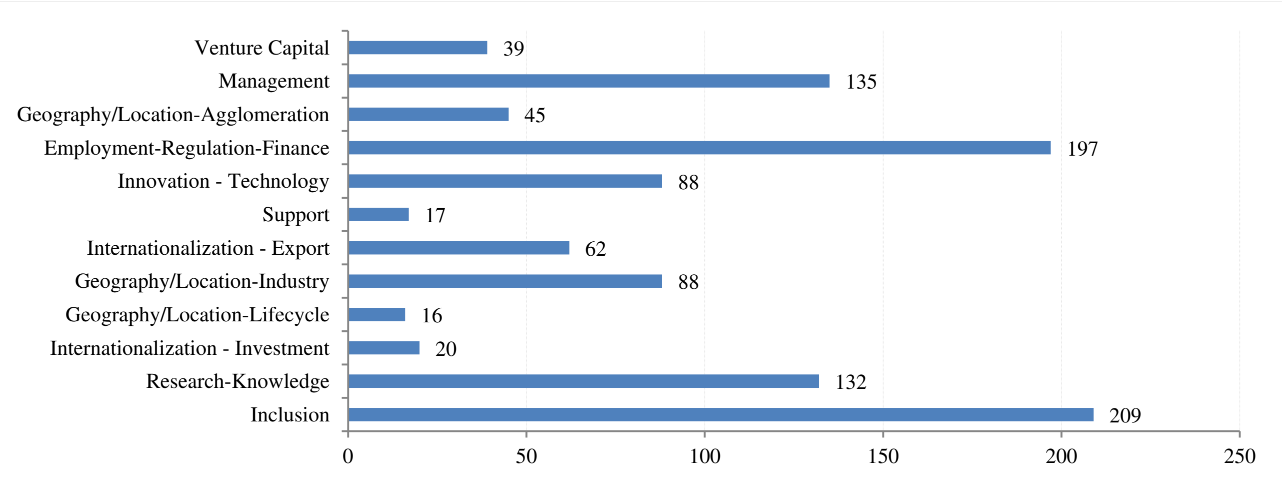


Fig. 1 Distribution of papers for the twelve-cluster taxonomy (Corpus Academia, 1990–2016)

Next, this paper has analysed the evolution of the twelve clusters during the study period, 1990–2016, by comparing the number of papers corresponding to each theme every five-years—with the exception of the period 2015–2016, as shown the Fig. 2. In particular, the main changes within the evolution of clusters are outlined in the following.

Inclusion only starts to be truly relevant in 2005, and especially after 2010 when topics like women, social entrepreneurship or immigrants gain momentum. Likewise, the relevance of *Management* increased in 2005 and presented a relative maximum in 2015–2016. The relevance of *Research–Knowledge* remained constant since 1995, with a peak in 2010-14; whereas, the number of papers related to *Innovation–Technology* made a quantitative leap in 2005. *Support* is, to some extent, related to research and scientific activities and reached a maximum in 2000–2004. Likewise, the behaviour of *Internationalization–Investment* and *Internationalization–Export* was very similar until 2010. Then, *Investment*, which is related with expansion by means of establishing abroad, reached its maximum in 2010–2014; however, the tendency of *Export*, with a focus on selling abroad and China, made the maximum shift in 2015–2016.

Location-related clusters also show significant differences. For instance, *Geography/Location–Sector* has a national dimension and reached its maximum in 1990–1995; however, it declined, especially in 2005. *Geography/Location–Lifecycle* gained relevance in 2000 and remained relatively stable until 2014—with the anecdotic result that there is no related paper in the period 2015-2016. The number of articles related to *Geography/Location–Agglomeration* started to grow intensely in 2005, even enjoying a relative maximum in 2015–2016, when regional dimension and the importance of connections were perceived as key features of entrepreneurship. Finally, the number of papers about *Employment–Regulation–Finance* was at its highest in 2010, just after a decline in 2005–2009, which coincides with the financial crisis. However, *Venture Capital* displayed a maximum in 1990–1994 and 2010–2014.

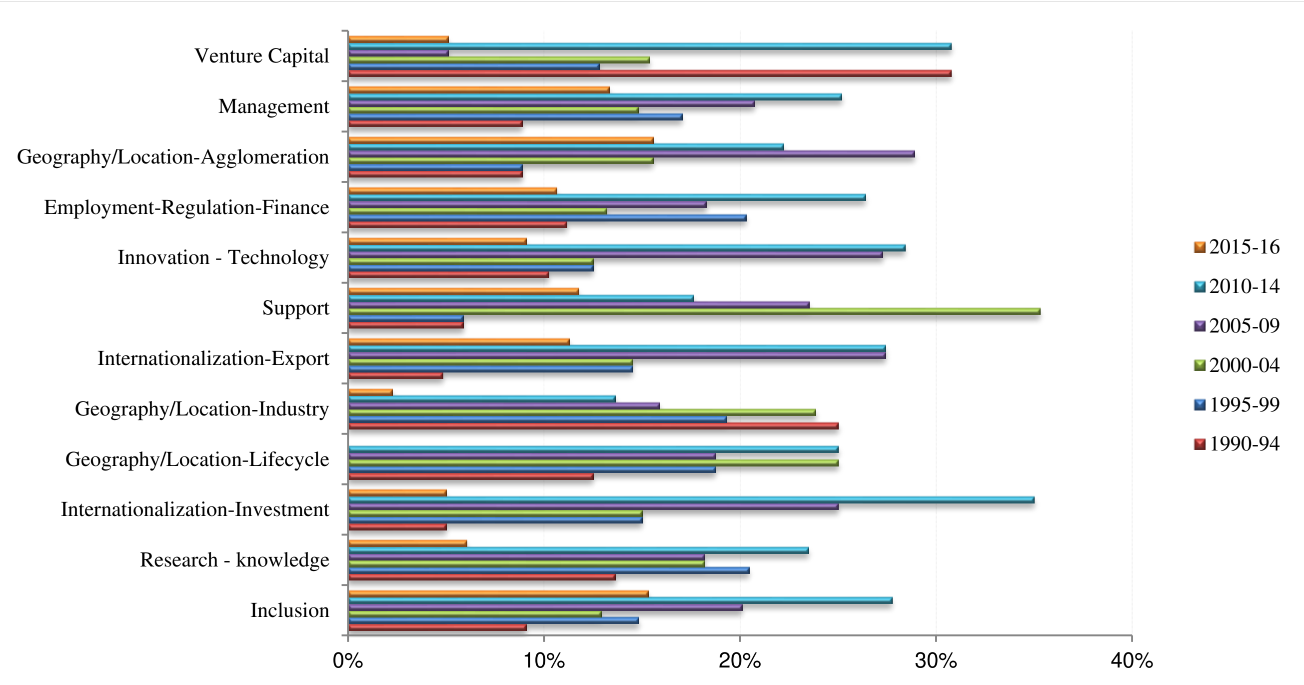


Fig. 2 Evolution of the relative weight of the number of papers in the twelve-cluster taxonomy for each five-year period (Corpus Academia, 1990–2016)

Table 1 Top 20 relevant word stems for the twelve-cluster analysis of the corpus Academia (1990–2016)

Order	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
	Inclusion	*Research - Knowledge*	*Internationalization – Investment*	*Geography / Location – Life cycle*	*Geography/ Location - Industry*	*Internationalization – Export*	*Support*	*Innovation - Technology*	*Employment – Regulation - Finance*	*Geography/ Location – Agglomeration*	*Management*	*Venture Capital*
1	entrepren	technolog	subsidiari	cycl	cluster	export	incub	innov	employ	network	organiz	ventur
2	region	transfer	multin	life	technolog	foreign	park	system	smes	knowledg	polit	capit
3	busi	univers	knowledg	technolog	biotechnolog	invest	scienc	region	firm	region	manag	invest
4	rural	patent	foreign	energi	industri	domest	technolog	smes	growth	tie	organ	fund
5	women	industri	japanes	metal	telecommun	china	univers	product	tax	innov	corpor	vc
6	develop	research	invest	product	electron	spillov	special	firm	credit	spillov	strategi	equiti
7	educ	innov	local	cluster	nation	enterpris	locat	process	subsidi	learn	environment	inform
8	econom	spin	host	recycl	competit	intern	innov	model	size	entrepren	respons	decis
9	social	project	expatri	local	govern	direct	tenant	technolog	self	social	institut	financ
10	start	knowledg	risk	industri	innov	firm	multimedia	enterpris	job	firm	theori	bootstrap
11	ventur	commerci	polit	stage	internet	countri	servic	design	busi	technolog	work	corpor
12	enterpris	intellectu	divest	curv	global	outward	yacht	nation	guarante	growth	famili	entrepren
13	activ	properti	oversea	nanotechnolog	korea	trade	firm	manufactur	market	bridg	firm	market
14	immigr	manag	acquisit	market	region	perform	base	knowledg	bank	endogen	perform	criteria
15	cultur	develop	intern	innov	countri	review	linkag	instrument	financi	organ	cultur	angel
16	institut	govern	typolog	simul	firm	young	virtual	develop	rate	collabor	strateg	vcs
17	growth	collabor	strateg	cast	develop	manufactur	manag	energi	effect	communiti	employe	prefer
18	support	learn	embedded	effici	sector	host	analog	learn	loan	cluster	human	human
19	famili	institut	review	constraint	system	economi	custom	centr	financ	industri	govern	financi
20	femal	capabl	frontier	accumul	network	busi	strait	research	start	structur	resourc	compani

Cluster analysis from a SE Asia, EU and US regional perspective

The analysis of the particularities of the topics of those research papers on entrepreneurship policy from SE Asia, EU and US affiliations also draws additional interesting regional insights. Table 3 and Fig. 7 shows the percentage of SE Asia, EU and US affiliated papers included in each cluster and also for the whole Academia Corpus.

The comparison of these results displays that the percentage of SE Asia affiliated papers related to *Inclusion* and *Employment-Regulation-Finance* are much lower than both the general case and also with respect to EU and/or US affiliated documents. On the other hand, SE Asia authored papers present a higher rate of papers in *Research-Knowledge*, *Internationalization–Investment* and *Internationalization–Export* clusters.

Likewise, almost a quarter of the US affiliated papers belong to the *Management* cluster, surpassing widely the participation of authors from the rest of affiliations. In addition, the percentage of US affiliated papers related to *Venture Capital* is higher than in the rest of the cases, but the participation of US affiliated authors in *Support* and *Innovation-Technology* papers is limited.

Finally, the share of EU affiliated papers related to Management is lower than in the rest of the cases, while EU authors tend to increase their presence with respect to the general case in those papers belonging to *Employment-Regulation-Finance* and *Support.*

	SE Asia (118 papers)	**EU (558 papers)**	**US (340 papers)**	**GLOBAL (1048 papers)**
C1 - Inclusion	9.3%	19.9%	22.4%	19.9%
C2 -Research - knowledge	26.3%	12.5%	10.0%	12.6%
C3 - Internationalization-Investment	5.9%	1.4%	1.8%	1.9%
C4 - Geography/Location-Lifecycle	1.7%	1.4%	0.9%	1.5%
C5 - Geography/Location-Industry	10.2%	6.6%	7.1%	8.4%
C6 - Internationalization-Export	11.0%	7.3%	4.1%	5.9%
C7 - Support	1.7%	2.3%	0.3%	1.6%
C8 - Innovation - Technology	9.3%	10.7%	2.9%	8.4%
C9 - Employment-Regulation-Finance	6.8%	22.4%	16.5%	18.8%
C10 - Geography/Location-Agglomeration	5.9%	4.7%	4.1%	4.3%
C11 - Management	9.3%	6.8%	24.7%	12.9%
C12 - Venture Capital	2.5%	3.8%	5.3%	3.7%

Table 3 Percentage of SE Asia/ EU /US-affiliated papers per cluster

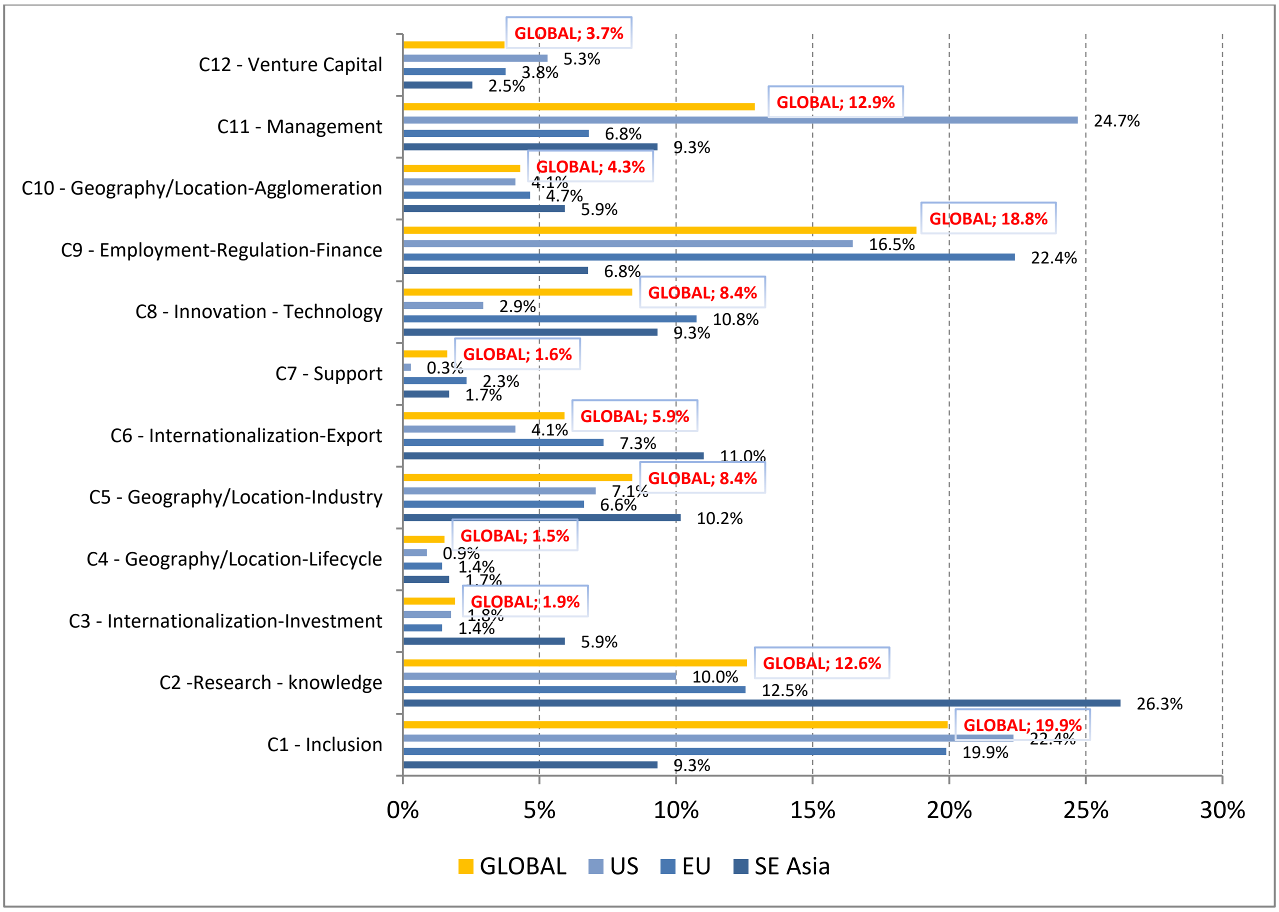


Fig. 7 Distribution / Percentage of papers per cluster within the Global Academia Corpus and within SE Asia, EU and US Sub-Corpuses

Similarly, Table 4 shows how papers authored in collaboration among SE Asia, EU and US affiliated researchers are distributed within the twelve clusters[12]. As previously mentioned, EU-US is the most frequent collaboration with 61 co-authored papers. From those, there are 14 papers related to *Inclusion* and 10 papers related to *Employment-Regulation-Finance*, which is coherent with the general interest of EU and US researchers separately. In addition, there are 13 papers related to *Management*, in spite of in general EU authors are not inclined to participate in this topic.

From the 18 and 16 co-authored papers between SE Asia – EU and SE Asia -US respectively, it is remarkable high percentage of papers related to *Research-Knowledge* since EU and US, much higher than the rate of papers in this topic when there are no SE Asia co-authors. In addition, 6 of the 16 papers shared by SE Asia –US affiliated researchers belong to Management and 6 of the 18 SE Asia – EU co-authored papers are related to Internationalization – Export.

Finally, focusing on the collaboration inwards the EU there is only two papers related to Innovation – Technology, while *Inclusion* and *Employment-Regulation-Finance* amount 11 and 12 papers respectively, adding up a shared rate of around the 45% of the total papers in this category. Besides, the proportion of papers related to Geography/Location – Agglomeration in this case doubles the rate corresponding both the general EU Sub – Corpus and the Global Academia Corpus

	SE Asia - EU	**SE Asia - US**	**EU - US**	**EU inwards**
C1 - Inclusion	1	1	14	11

[12] There is an unique paper shared at the time by SE Asia, EU and US affiliated authors and it belongs to the *Research-Knowledge* cluster.

C2 -Research - knowledge	4	4	8	7
C3 - Internationalization-Investment	1	1	0	1
C4 - Geography/Location-Lifecycle	0	0	0	0
C5 - Geography/Location-Industry	3	1	2	3
C6 - Internationalization-Export	6	1	3	4
C7 - Support	0	0	1	1
C8 - Innovation - Technology	1	0	2	2
C9 - Employment-Regulation-Finance	0	1	10	12
C10 - Geography/Location-Agglomeration	0	1	4	5
C11 - Management	1	6	13	3
C12 - Venture Capital	1	0	4	2
TOTAL	**18**	**16**	**61**	**51**

Table 4 : Number of SE Asia / EU / US co-authored papers per cluster

Conclusions

As a main contribution, this paper delineates a categorization of the key concepts and themes that have shaped the entrepreneurship policy research agenda and their evolution during 1990-2016; and it also adds some insights about the geographical configuration of the discipline and its evolution over time. Thus, this study can potentially benefit both researchers and policymakers in the field by identifying main trends and research gaps. An additional purpose of the paper is to show the potential of text mining to complement other approaches based on bibliometrics when reviewing the field of entrepreneurship – and also in other disciplines of social sciences or policy studies-.-.

In general, the analyses and results gathered in the paper show that the entrepreneurship policy research agenda has gained both attention, with a growing amount of articles in top journals especially since mid-2000s; and also maturity including more sophisticated and complex themes and concepts over the years. Results from the text mining represent a synthesis of the conceptual framework of the discipline, which has been largely dominated by studies related to inclusion, employment or regulation. Regarding the themes addressed, the distribution of papers within the different clusters over time shows that research has been expanded and evolved from an initial classical approach about the relationship between entrepreneurship and employment to a wider and multidisciplinary scope not directly linked to conventional policy, including management issues and advance views on agglomeration instead narrower sectorial approaches. With this comprehensive view on the evolution of priorities that have shaped the research in the field, it could be also concluded that research themes have become more mainstream over time, which could be interpreted as an effort to approximate to the practice.

On the geographical perspective, the regional analysis shows that researchers affiliated in US and in the EU dominate the scientific production in top journals in the field. In the case of EU, there is a major influence of the UK affiliated researchers and the contributions of Netherlands and Sweden are also significant in relative terms. This is coherent with previous studies about entrepreneurship as an independent field in social sciences, clearly dominated by Anglo-Saxon countries, particularly by those from North America and the UK, with the relative solid position of Nordic countries and Netherlands (Meyer et al. 2014).These results would also go along with what was mentioned specifically by Welter and Lasch in relation to the national or international orientation of the European research communities in the entrepreneurship field and how aspects as language and the size of national scenes might be playing a role in the case of Germany or France (Welter and Lasch 2008). Despite the advances, the internationalizacion of the research community in the field is still limited. Besides, results from this paper display that themes and topics addressed across different regional affiliations are rather heterogeneous, probably linked to the diverse development priorities. In authors' view, one of the most significant findings is the additional level of complexity that these amalgam of prorities implies to the transfer of knowledge from the academia to policymakers and viceversa. At the same time, collaboration outwards seem to encourage the diversification of themes to research, so promoting connections among researchers from different affiliations should be a challenging priority that arises to enrich the field.

As an additional conclusion, from the authors´ perspective results in this paper also show how text mining could help to frame the evolution of the academic discourse on a field. The exploratory nature of text mining permit to obtain new knowledge and reveal hidden patterns from large amount of documets / text data, which represent an opportunity to complement other qualitative reviews. In this sense, further works could examine alternative corpora of texts such as policy documents to advance in the contribution to the study of relationships between research and policy, looking for patterns, coincidences and potential linkages between academic and policy approaches, especially considering that entrepreneurship research in some geographies like Europe has depended deeply on policy (Landström 2015; Rosa 2013). This could add new elements to the debate about the use of scientific knowledge in policymaking, taking into account that entrepreneurship policy appropriateness and effectiveness have been largely controversial and in addition the scope, methods and outcomes of entrepreneurship research need a shift to be useful for policymakers (Zahra and Wright 2011, 67)

Ultimately, analyses and results gathered in this paper are obviously a simplification about the evolution of the research agenda for entrepreneurship policy mainly due to the nature of the sample. First of all, this study only considers papers in top entrepreneurship journals in English included in Scopus, but other important databases and/or dissemination channels such as book chapters, conference proceedings or different types of events or even other types of journals not mainly dedicated to entrepreneurship studies could be considered to obtain new insights that would complement the results obtained in this study. In addition, as occurs with all statistical analysis, there are also some limitations related to the relative dependency on the subjectivity of the researcher/s performing the text mining analysis. In this case, the choice of the algorithm and parameters for the clustering and the interpretation of the results are not absolutely neutral. Precisely to cope with this potential bias, the analayis in this paper introduced the Davies-Bouldin index, as an objective way to evalualete the appropriateness of the various clustering divisions.

However these limitations, this paper expects to contribute to enrich the existing literature on the field through the identification of the main threads that have configured the research priorities, putting them in a temporal and geographical framework and unveiling main trends and shifts within the academic discourse. All in all, research on entrepreneurship policy has evolved through topics increasingly closer to those used in practice and a growing community of international researchers willing to cooperate outwards and moving away from the ivory tower.

Appendix - additional results from text mining analysis

A1. Preliminary results - most frequent word stems per five-year periods for the corpus Academia

Figures s1-s6 show the most repeated word stems in periods of five years since 1990 except for the last period (2015-2016) which obviously only comprises two years.

First, it is interesting to examine the relative position between the aggregration 'firm-enterprise-company' and 'entrepreneur': the former leads until 2005-09, but since then 'entrepreneur' have surpassed it, even becoming the most frequent word in 2015-16. In a similar pattern 'SME' has escalated positions over the years. For instance the term 'SME' appeared at 148th position in 1990-1995 but then has gone up within the Top 50, being at 17th in 1995-1999 (the highest), and then keeping a relatively constant profile at 39th in 2000-2004, 21st in 2005-2009, 25th in 2010-2014 and finally rising again up to 17th in 2015-2016, a similar pattern to that of 'startup' (from 56th in 2000-04 to 51st in 2015-16).

The evolution of 'technology' and 'innovation' is also relevant. 'Technology' starts at 2nd position in 1990-04 and 3rd in 1994-1999 and in 2000-2004 and then begins to decline moving to 9th in 2005-2009, 7th in 2010-2014 and 20th in 2015-2016. On the contrary, 'innovation' departs from 8th in 1990-94 and 10th in 1995-1999, but then it increases its relevance positioning at 7th in 2000-04, 4th in 2005-2009 and 2010-2014 and 5th in 2015-2016. 'Data' and 'university' present a similar pattern as 'innovation'. 'Data' starts at 86th in 1990-94 and then goes up to 49th and 48th in 1995-99, 39th in 2005-09 and 26th in 2010-14 and 2015-16. Particularly interesting is the emergence of 'university': the term was far from the top 50 during the 1990s and then ranking 56th in 2000-04 and 43rd, 44th and 43rd in 2005-09, 2010-14 and 2015-16, respectively.

'Social' in another interesting case. The term started out in 1990-95 at 124th, in 1996-00 at 46th and in 2000-04 at 61st. However, since 2005-09 'social' has gone up at 25th and it has continued being more and more frequent in the next periods, at 22nd in 2010-14 and at 15th in 2015-2016. To this regard note that most of the terms related with economics remain basically sTable s, with the own term 'economy' staying within the top ten for all the periods. Maybe the little exception is 'industry' that has declined from 4th position in 1990-94 and 8th position in 1995-1999, 2000-2004 and 2005-2009 to 12nd in 2010-2014 and 16th in 2015-2016.

As previously stated, 'education' is far to be among the most frequent terms during all the considered periods, although its best rankings are during the last years of the overall period being at 59th in 2010-14 and 74th in 2015-16. Concepts related to education, such as 'skill' or 'train' are not either relevant with the unique exceptions of 'learn', that is relatively relevant during 1995-99 (61st) and 2000-04 (59th), and the case of 'university' that is in the top 50 since 2005 as mentioned above.

Regarding geographical considerations, 'country-nation' and 'region' stay well above 'international' for the whole period with slight variations between them: after an initial period that are almost equally frequent at 15th and 14th positions, 'country-nation' dominates from 1995-99 up to 2005-09, and then 'region' overtakes along the period 2010-2016, with 'region' at 10th position in 2010-14 and 9th in 2015-2016 while 'country' goes down at 13th position in 2010-14 and at 11th position in 2015-2016.

Terms related with ecosystem, this is terms related with entrepreneurial activities context and conditions including interactions with the rest of the agents, display interesting evolution patterns. 'Environment' and 'system' both rank their lowest in the initial period of 1990-94 (100th and 49th, respectively), being 1995-99 the period with the highest positon for 'environment' (at 31st) and 2005-09 for 'system' (at 26th). 'Network' and 'cluster' spiked at 2005-09 (16th and 32nd, respectively) with the latter almost unheard of in the period 1990-99. Note also that the very term 'ecosystem' is virtually missing up to 2015-2016, when it has appeared at 269th position.

Considering ecosystem domains, as as previously described, 'policy' keeps a position at the top three for all the considered periods due to be the focus of the research. 'Finance' also appears in the Top 50 for all the considered periods: it starts at 35th in 1990-94 and then rises up to 28th, thereafter declining to 32th in 2000-2004 and 41st in 2005-09 and finally it rises at the top 30 during the last years, being 28th and 27th in 2010-14 and 2015-16 respectively. Then 'market' is very relevant for all the periods remaining at the top 50 with the side note of going down to 36th in 2010-2014 and to 82nd in 2015-2016. 'Culture' behaviour is irregular: it ranks 224th in 1990-1994 and then moves up to the Top 200 during the following four five-year periods, reaching 63rd position, its highest rank, at 2015-2016. Finally, 'human' is far from the leading positions being the 92nd term at 2010-2014 as its best ranking.

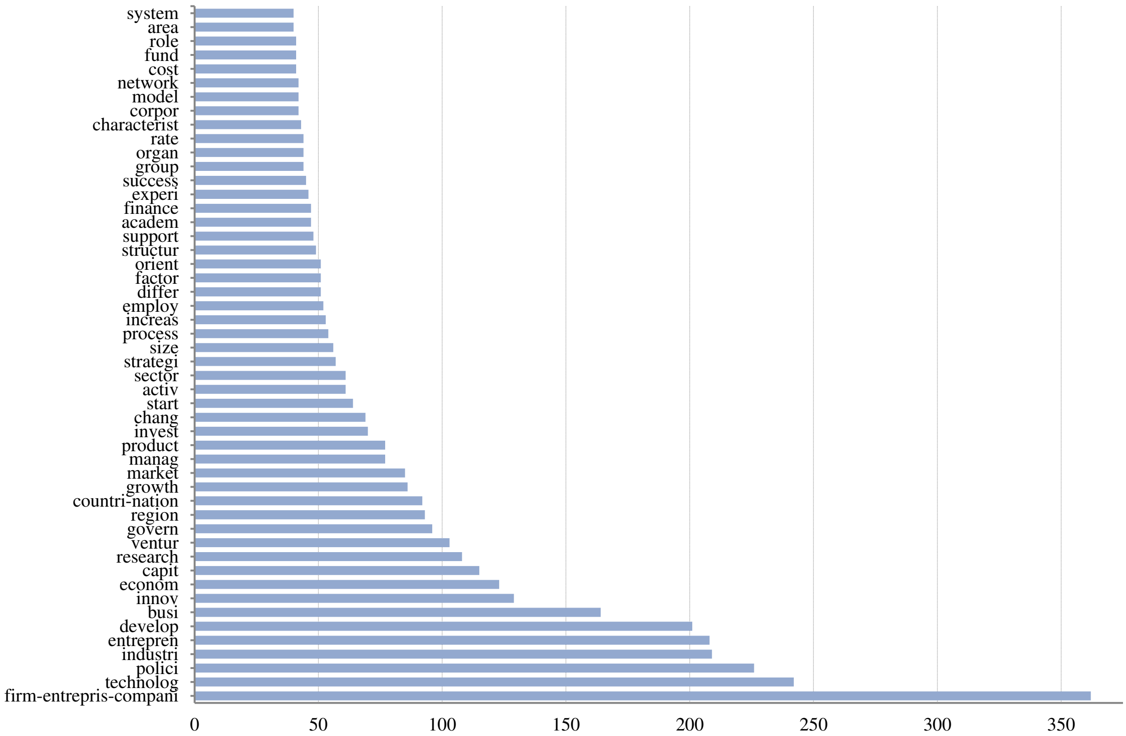


Figure s8 - 50 most frequent word stems for the corpus Academia in 1990-1994

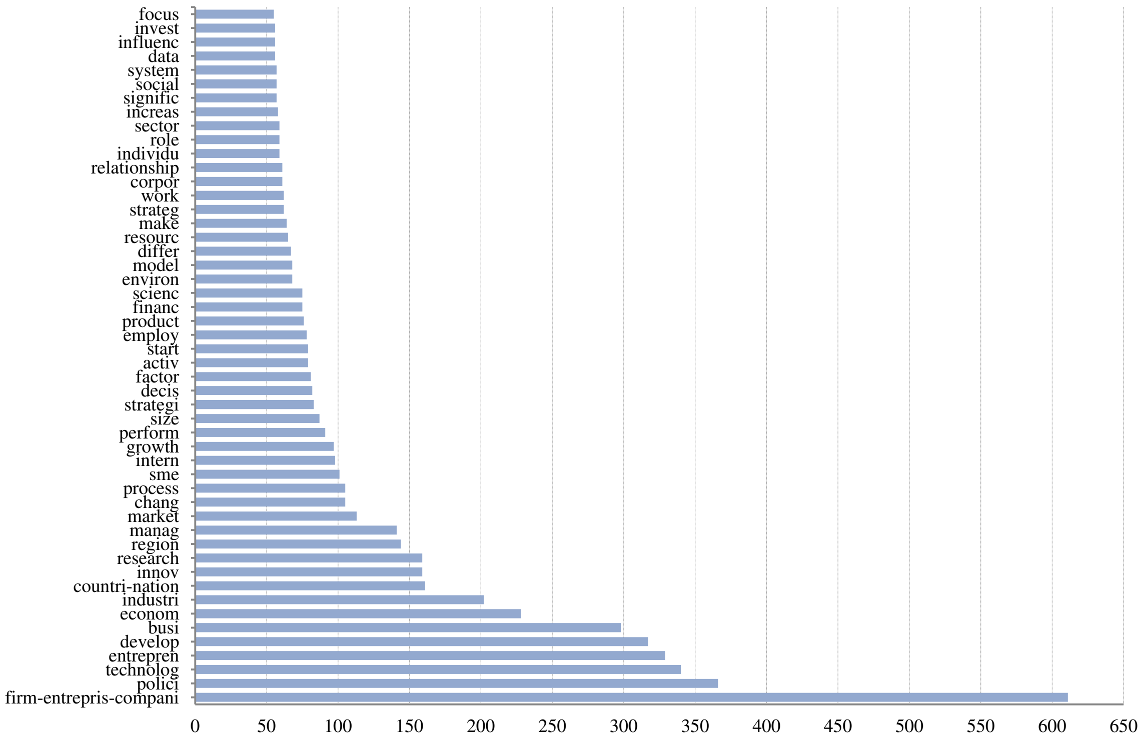


Figure s9 - 50 most frequent word stems for the corpus Academia in 1995-1999

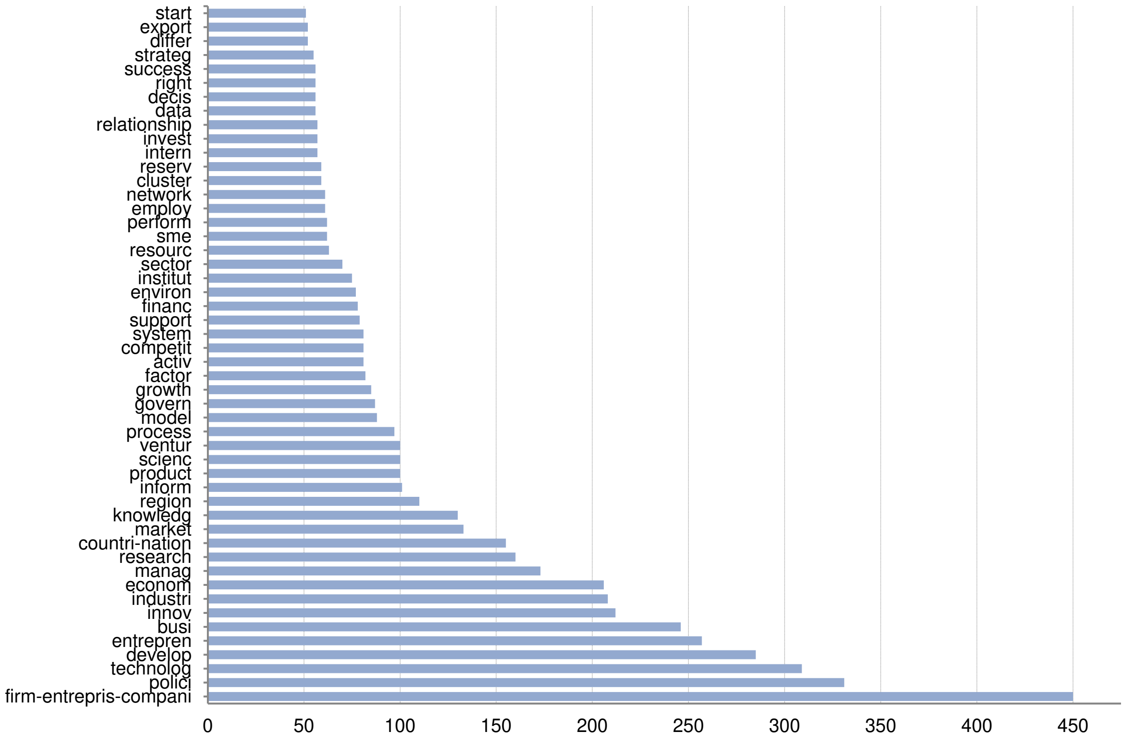


Figure s10 - 50 most frequent word stems for the corpus Academia in 2000-2004

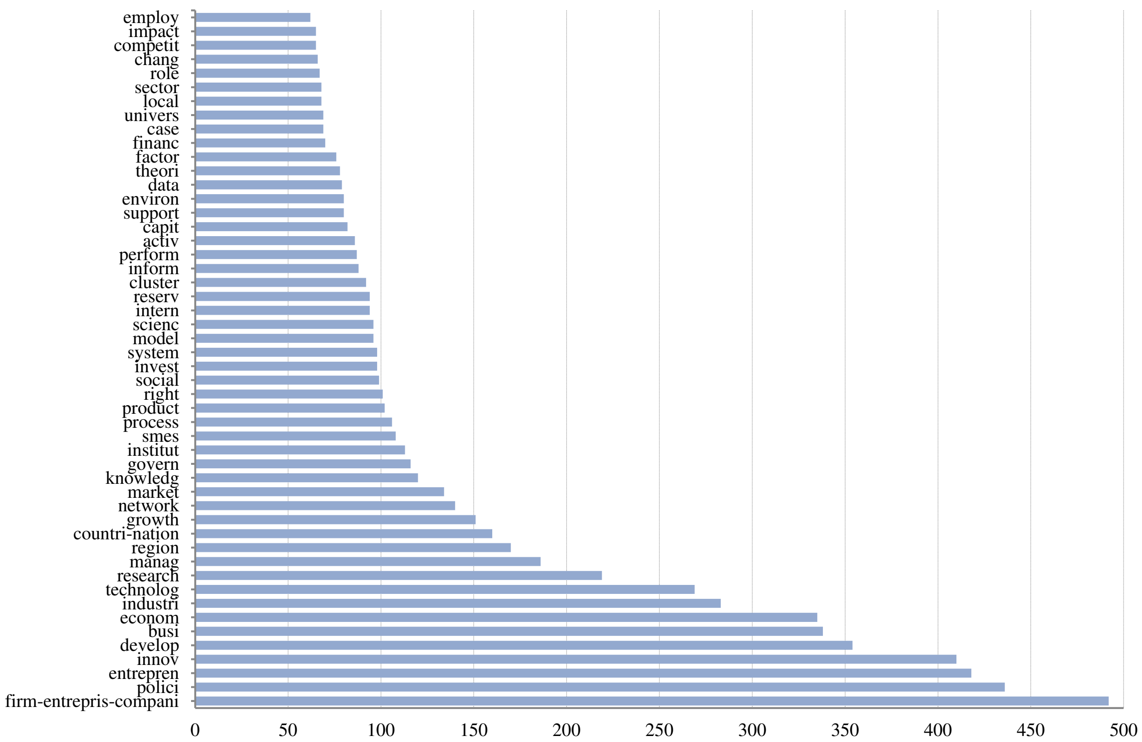


Figure s11 50 most frequent word stems for the corpus Academia in 2005-2009

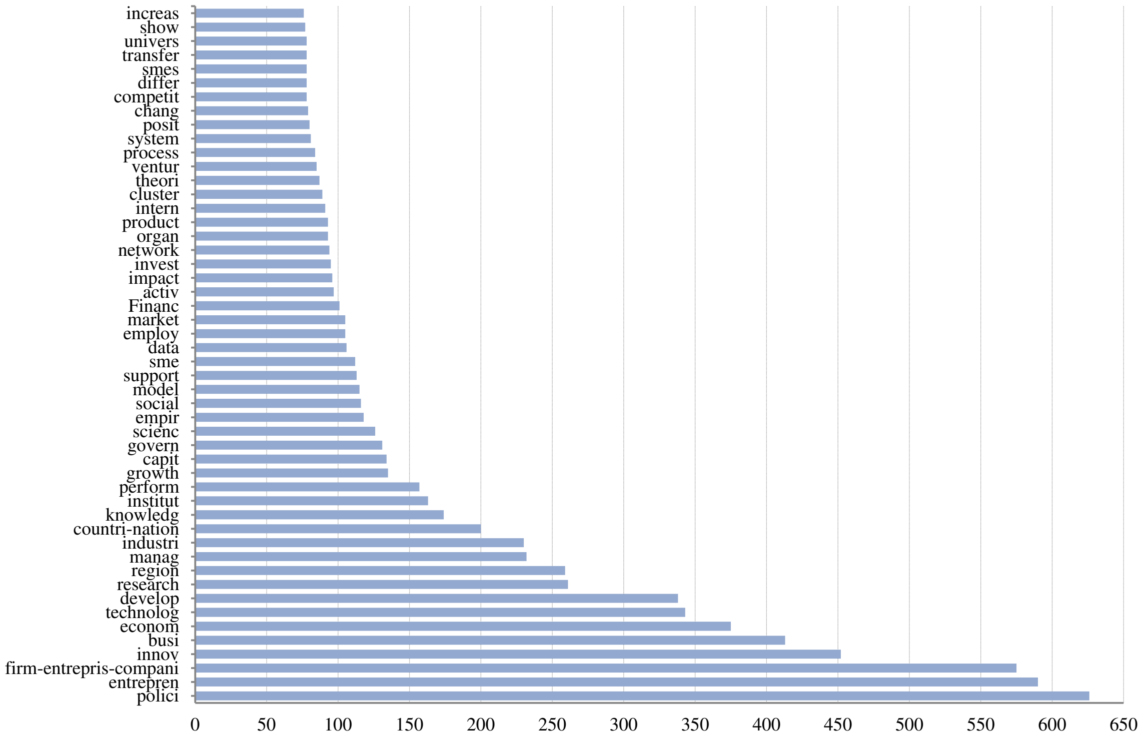


Figure s12 - 50 most frequent word stems for the corpus Academia in 2010-2014

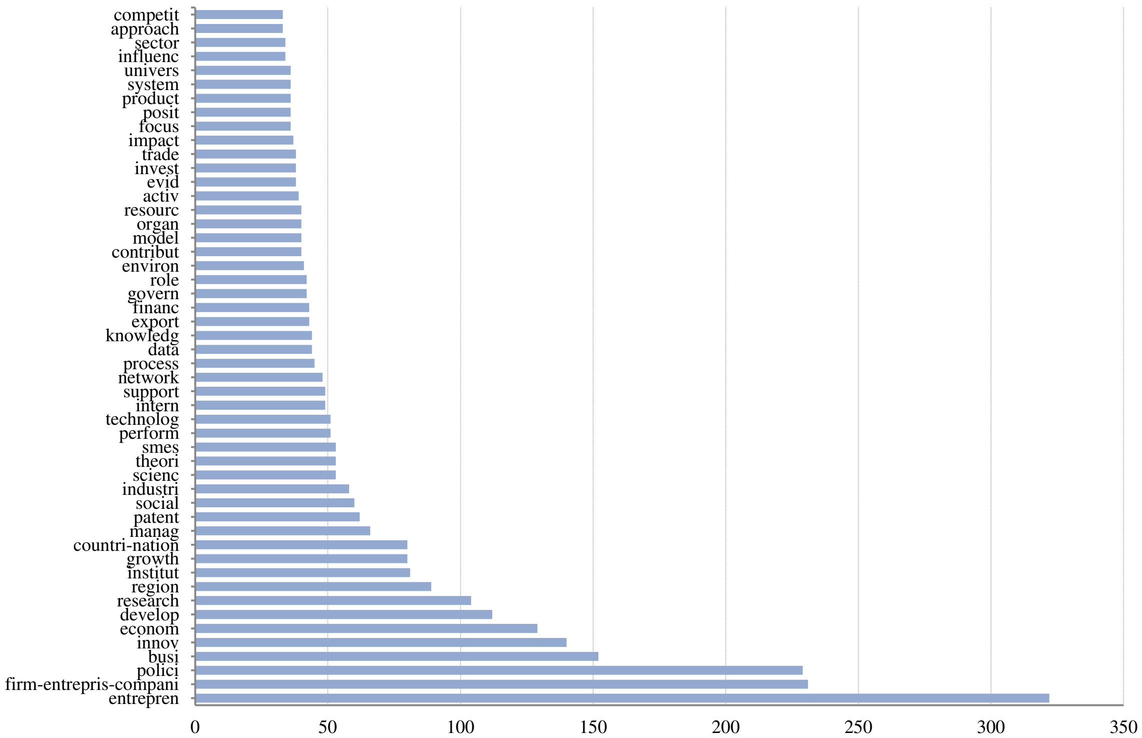


Figure s13 - 50 most frequent word stems for the corpus Academia in 2015-16

A2 -Cluster analysis - Corpus Academia

See Table ss 1 –9

Table s2 Top 20 more relevant word stems for four cluster analysis in the corpus Academia (1990-2016)

Order	C1	C2	C3	C4
1	entrepren	technolog	foreign	network
2	growth	innov	polit	cluster
3	firm	industri	organiz	region
4	employ	transfer	invest	knowledg
5	busi	univers	manag	innov
6	ventur	system	corpor	industri
7	smes	patent	subsidiari	entrepren
8	capit	firm	strategi	firm
9	export	research	organ	social
10	tax	product	intern	district
11	start	incub	academi	develop
12	region	develop	firm	local
13	enterpris	govern	institut	spillov
14	econom	knowledg	cultur	global
15	size	technov	environment	learn
16	market	nation	strateg	institut
17	credit	biotechnolog	respons	growth
18	self	process	countri	system
19	develop	scienc	perform	framework
20	perform	countri	china	analysi

Table s3 Top 20 more relevant word stems for five cluster analysis in the corpus Academia (1990-2016)

Order	C1	C2	C3	C4	C5
1	entrepren	smes	cluster	organiz	technolog
2	region	export	network	manag	innov
3	employ	innov	region	polit	univers
4	ventur	firm	innov	organ	transfer
5	growth	invest	knowledg	corpor	industri
6	start	foreign	industri	academi	patent
7	busi	size	firm	environment	system
8	capit	market	global	famili	knowledg
9	firm	enterpris	entrepren	perform	research
10	econom	credit	agglomer	theori	incub
11	self	product	develop	respons	biotechnolog
12	social	medium	analysi	strategi	learn
13	develop	growth	framework	firm	park
14	rural	countri	local	cultur	develop
15	educ	subsidiari	structur	institut	govern
16	institut	guarante	type	work	project
17	activ	sme	learn	employe	firm
18	job	subsidi	evolut	chang	manag

19	women	industri	competit	owner	nation
20	tax	financ	system	resourc	scienc

Table s4 Top 20 more relevant word stems for six cluster analysis in the corpus Academia (1990-2016)

Order	C1	C2	C3	C4	C5	C6
1	entrepren	network	technolog	foreign	organiz	smes
2	ventur	cluster	innov	invest	manag	export
3	capit	region	transfer	subsidiari	patent	firm
4	busi	biotechnolog	industri	domest	polit	growth
5	start	innov	system	direct	univers	employ
6	region	knowledg	incub	china	organ	tax
7	growth	industri	firm	multin	institut	credit
8	educ	local	park	countri	research	size
9	women	develop	develop	intern	strategi	subsidi
10	rural	firm	learn	spillov	corpor	medium
11	social	entrepren	process	host	firm	busi
12	activ	social	product	knowledg	perform	enterpris
13	econom	global	nation	japanes	spin	guarante
14	owner	district	scienc	outward	govern	market
15	firm	learn	manag	local	strateg	job
16	decis	agglomer	research	firm	theori	bank
17	cultur	econom	countri	locat	intellectu	self
18	develop	structur	govern	economi	cultur	product
19	immigr	institut	competit	corpor	journal	rate
20	invest	process	knowledg	industri	knowledg	loan

Table s5 Top 20 more relevant word stems for seven cluster analysis in the corpus Academia (1990-2016)

Order	C1	C2	C3	C4	C5	C6	C7
1	smes	export	innov	entrepren	technolog	organiz	network
2	employ	invest	energi	busi	transfer	cultur	cluster
3	growth	foreign	system	manag	univers	logic	region
4	tax	ventur	product	firm	industri	workplac	knowledg
5	firm	capit	region	perform	innov	intent	innov
6	credit	trade	smes	social	patent	manag	industri
7	size	subsidiari	firm	institut	incub	respons	social
8	subsidi	firm	technolog	environment	research	strategi	entrepren
9	self	countri	process	econom	biotechnolog	chang	firm
10	medium	investor	design	famili	park	integr	spillov
11	busi	intern	enterpris	region	govern	polit	learn
12	guarante	market	market	rural	scienc	employ	local
13	job	direct	nation	ventur	develop	ident	global
14	start	financ	instrument	women	spin	job	develop
15	loan	china	manufactur	activ	project	structur	district
16	enterpris	decis	develop	educ	firm	typolog	competit
17	bank	japanes	industri	develop	commerci	break	agglomer
18	region	capitalist	approach	polit	manag	understand	endogen

19	market	domest	competit	chang	base	corpor	peripher
20	financi	multin	increment	corpor	knowledg	behavior	growth

Table s6 Top 20 more relevant word stems for eight cluster analysis in the corpus Academia (1990-2016)

Order	C1	C2	C3	C4	C5	C6	C7	C8
1	entrepren	biotechnolog	technolog	subsidiari	patent	smes	export	network
2	ventur	firm	innov	multin	organiz	tax	foreign	cluster
3	capit	polit	transfer	knowledg	univers	employ	invest	region
4	busi	market	industri	japanes	spin	growth	domest	knowledg
5	start	industri	system	foreign	research	firm	direct	innov
6	region	competit	park	tourism	knowledg	credit	intern	entrepren
7	famili	telecommun	incub	local	intellectu	subsidi	spillov	industri
8	women	environment	firm	invest	institut	size	firm	social
9	rural	govern	learn	intern	properti	job	outward	firm
10	educ	strateg	develop	top	innov	guarante	china	develop
11	growth	manag	process	strateg	project	medium	countri	global
12	invest	trade	product	manag	manag	busi	perform	endogen
13	social	servic	scienc	agenc	corpor	self	chines	growth
14	firm	strategi	govern	polit	collabor	enterpris	review	agglomer
15	owner	countri	nation	theori	cultur	bank	young	local
16	econom	product	manag	parent	academ	loan	trade	district
17	activ	electron	research	strategi	interact	financ	host	spillov
18	decis	compani	model	host	work	rate	impact	learn
19	inform	develop	countri	cultur	privat	effect	manufactur	competit
20	develop	local	energi	risk	technolog	econom	price	geograph

Table s7 Top 20 more relevant word stems for nine cluster analysis in the corpus Academia (1990-2016)

Order	C1	C2	C3	C4	C5	C6	C7	C8	C9
1	entrepren	univers	subsidiari	cycl	network	export	incub	technolog	employ
2	venture	organiz	multin	life	cluster	foreign	park	innov	smes
3	region	patent	knowledg	scrap	biotechnolog	invest	technolog	transfer	growth
4	capit	manag	foreign	technolog	region	domest	scienc	industri	firm
5	busi	polit	japanes	metal	innov	china	univers	system	tax
6	rural	research	invest	product	knowledg	intern	innov	region	credit
7	women	corpor	local	cluster	firm	firm	tenant	firm	size
8	start	institut	host	recycl	industri	spillov	locat	product	busi
9	develop	strategi	expatri	local	learn	enterpris	firm	develop	subsidi
10	econom	academi	risk	stage	social	direct	multimedia	process	self
11	educ	spin	polit	curv	global	countri	servic	nation	job
12	social	knowledg	divest	nanotechnolog	entrepren	outward	custom	model	guarante
13	activ	perform	oversea	market	technolog	trade	base	govern	market
14	enterpris	privat	acquisit	energi	develop	market	manag	competit	financ

15	cultur	intellectu	intern	constraint	collabor	perform	link	research	medium
16	growth	govern	typolog	simul	district	review	promot	countri	effect
17	institut	respons	strateg	cast	structur	economi	linkag	learn	bank
18	invest	strateg	embedded	accumul	local	intens	virtual	smes	loan
19	immigr	project	review	innov	competit	busi	analog	manag	start
20	decis	firm	frontier	industri	analysi	young	ventur	technic	cost

Table s8 Top 20 more relevant word stems for ten cluster analysis in the corpus Academia (1990-2016)

Order	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
1	entrepreneur	cluster	technolog	subsidiari	univers	smes	export	network	polit	employ
2	ventur	biotechnolog	innov	multin	patent	firm	foreign	region	academi	self
3	capit	industri	industri	knowledg	spin	growth	invest	knowledg	environment	incub
4	busi	innov	transfer	japanes	knowledg	tax	domest	innov	corpor	work
5	start	region	system	foreign	intellectu	credit	china	district	decoupl	organiz
6	region	firm	firm	parent	properti	size	direct	entrepren	manag	theori
7	rural	network	develop	local	research	subsidi	spillov	develop	institut	manag
8	women	framework	product	invest	technolog	employ	intern	social	logic	organ
9	famili	global	govern	acquisit	innov	busi	outward	industri	stakehold	cultur
10	educ	agglomer	process	strateg	academ	guarante	firm	endogen	govern	job
11	econom	learn	nation	typolog	collabor	product	countri	firm	organiz	human
12	growth	resourc	park	embedded	transfer	medium	perform	growth	respons	individu
13	activ	research	competit	risk	organiz	market	young	territori	bank	resourc
14	institut	knowledg	model	host	off	enterpris	host	local	strategi	servic
15	firm	peripher	technov	intern	project	financi	review	spillov	journal	firm
16	privat	nano	scienc	top	invent	rate	trade	institut	organ	respons
17	social	case	learn	polit	privat	bank	impact	cluster	pressur	team
18	develop	technolog	research	review	commerci	loan	enterpris	creativ	field	worker
19	decis	geograph	countri	outflow	interact	cost	manufactur	econom	regul	pay
20	cultur	nation	manag	strategi	ip	job	industri	global	conting	strategi

Table s9 Top 20 more relevant word stems for eleven cluster analysis in the corpus Academia (1990-2016)

Order	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
1	smes	polit	cycl	entrepren	technolog	entrepren	ventur	foreign	export	network	innov
2	employ	organiz	life	famili	univers	intellectu	capit	invest	price	region	cluster
3	growth	manag	product	women	transfer	social	invest	subsidiari	trade	tie	region
4	firm	environment	scrap	start	industri	minor	capitalist	rural	smes	knowledg	system
5	tax	corpor	recycl	busi	innov	gem	investor	knowledg	firm	innov	knowledg
6	region	institut	cluster	ventur	patent	busi	equiti	spillov	size	social	firm
7	credit	strategi	technolog	owner	incub	educ	fund	direct	manufactur	firm	smes
8	size	theori	behaviour	femal	knowledg	properti	vc	china	nonexport	entrepren	product
9	enterpris	academi	market	job	research	internationalis	bootstrap	locat	servic	bridg	process
10	busi	organ	stage	immigr	biotechnolog	enterpris	financ	intern	orient	organ	industri
11	medium	perform	local	gender	project	econom	inform	host	perform	learn	nation
12	self	firm	model	educ	spin	stage	market	firm	young	collabor	framework
13	subsidi	respons	curv	cultur	learn	institut	decis	countri	intern	communiti	technolog
14	entrepren	chang	biotechnolog	firm	park	definit	criteria	local	medium	world	competit
15	sector	social	firm	individu	govern	issu	entrepren	region	intens	technolog	model
16	innov	regul	simul	work	scienc	research	vcs	domest	india	theori	enterpris
17	econom	journal	process	human	commerci	rule	privat	outward	product	district	develop
18	market	cultur	accumul	nascent	develop	ethnic	human	multin	distributor	structur	research
19	guarante	govern	restructur	male	technov	theori	australian	review	immigr	divers	network
20	industri	bank	develop	founder	manag	innov	govern	industri	promot	inter	local

Table s10 Top 20 more relevant word stems for thirteen cluster analysis in the corpus Academia (1990-2016)

Order	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
1	smes	export	cycl	environment	patent	organiz	cluster	polit	employ	ventur	entrepren	technolog	foreign
2	credit	price	life	manag	innov	intent	network	firm	self	capit	region	innov	subsidiari
3	guarante	trade	scrap	academi	project	workplac	region	corpor	job	entrepren	enterpris	transfer	invest
4	innov	smes	recycl	perform	collabor	organ	innov	market	growth	invest	rural	univers	direct
5	tax	firm	energi	field	cooper	manageri	knowledg	biotechnolog	women	decis	innov	industri	china
6	medium	size	metal	institut	technolog	cultur	tie	govern	tax	vc	develop	incub	host
7	loan	manufactur	technolog	theori	japan	logic	firm	trade	firm	start	econom	park	spillov
8	size	nonexport	product	telecommun	univers	ident	industri	manag	start	fund	local	system	domest
9	bank	servic	market	adopt	invent	employe	learn	countri	busi	financ	growth	develop	intern
10	enterpris	orient	sustain	pay	govern	break	social	effect	entrepren	equiti	busi	spin	knowledg
11	firm	perform	curv	strategi	research	chang	entrepren	invest	unemploy	bootstrap	district	learn	outward
12	financ	young	plant	energi	system	integr	technolog	capit	founder	portfolio	social	firm	multin
13	growth	intern	behaviour	decoupl	nation	rule	global	busi	creation	serial	sector	research	countri
14	debt	medium	simul	firm	privat	alloc	analysi	strateg	work	human	knowledg	scienc	review
15	lend	intens	innov	journal	knowledg	strategi	structur	cost	state	inform	industri	knowledg	firm
16	region	india	substitut	resourc	industri	typolog	type	risk	effect	nascent	institut	manag	locat
17	subsidi	product	cast	pressur	competit	behavior	local	intern	famili	busi	firm	commerci	impact
18	busi	distributor	constraint	relationship	instrument	norm	framework	subsidi	rate	market	privat	base	acquisit
19	support	immigr	section	social	firm	lever	agglomer	uncertainti	up	criteria	educ	process	local
20	sba	promot	stage	complex	orient	crime	develop	product	size	risk	servic	competit	industri

References

Acs, Zoltan, Thomas Åstebro, David Audretsch, and David T. Robinson. 2016. "Public Policy to Promote Entrepreneurship: A Call to Arms." *Small Business Economics* 47(1): 35–51.

Acs, Zoltan J., David B. Audretsch, Pontus Braunerhjelm, and Bo Carlsson. 2012. "Growth and Entrepreneurship." *Small Business Economics* 39(2): 289–300.

Ananiadou, Sophia et al. 2009. "Supporting Systematic Reviews Using Text Mining." *Social Science Computer Review*.

Arbelaitz, Olatz et al. 2013. "An Extensive Comparative Study of Cluster Validity Indices." *Pattern Recognition* 46(1): 243–56.

Asheim, Bjørn T., and Arne Isaksen. 2002. "Regional Innovation Systems: The Integration of Local 'Sticky' and Global 'Ubiquitous' Knowledge." *Journal of Technology Transfer* 27(1): 77–86.

Audretsch, David B., Werner Bönte, and Max Keilbach. 2008. "Entrepreneurship Capital and Its Impact on Knowledge Diffusion and Economic Performance." *Journal of Business Venturing* 23(6): 687–98.

Audretsch, David B, Isabel Grilo, and a Roy Thurik. 2007. "Explaining Entrepreneurship and the Role of Policy: A Framework." *Handbook of research on entrepreneurship policy*: 1–17.

Autio, Erkko, and Heikki Rannikko. 2016. "Retaining Winners: Can Policy Boost High-Growth Entrepreneurship?" *Research Policy* 45(1): 42–55.

Baptista, Rui. 1998. "Clusters, Innovation, and Growth: A Survey of the Literature." In *The Dynamics of Industrial Clustering International Comparisons in Computing and Biotechnology*, eds. Peter G M Swann, Martha Prevezer, and David Stout. Oxford University Press, 13–51.

Bennett, Robert J. 2014. Entrepreneurship, Small Business and Public Policy: Evolution and Revolution *Entrepreneurship, Small Business and Public Policy: Evolution and Revolution*.

Bloom, Paul N, and Gregory Dees. 2008. "Cultivate Your Ecosystem." *Stanford Social Innovation Review* Winter: 47–53.

Bragge, Johanna, and Jan Storgårds. 2007. "Utilizing Text-Mining Tools to Enrich Traditional Literature Reviews . Case : Digital Games." In , 1–24.

Busenitz, Lowell W. et al. 2014. "Entrepreneurship Research (1985-2009) and the Emergence of Opportunities." *Entrepreneurship: Theory and Practice* (September 2015).

Cohen, Aaron M., and William R. Hersh. 2005. "A Survey of Current Work in Biomedical Text Mining." *Briefings in Bioinformatics*.

Cohen, Boyd. 2006. "Sustainable Valley Entrepreneurial Ecosystems." *Business Strategy and the Environment*.

Cooke, Philip, Mikel Gomez Uranga, and Goio Etxebarria. 1997. "Regional Innovation Systems: Institutional and Organizational Dimensions." *Research Policy* 26: 475–91.

Davies, David L., and Donald W. Bouldin. 1979. "A Cluster Separation Measure." *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-1(2): 224–27. http://ieeexplore.ieee.org/document/4766909/.

Delen, Dursun, and Martin D. Crossland. 2008. "Seeding the Survey and Analysis of Research Literature with Text Mining." *Expert Systems with Applications*.

Dennis Jr., William J. 2011. "Entrepreneurship, Small Business and Public Policy Levers." *Journal of Small Business Management*.

Fan, Weiguo, Linda Wallace, Stephanie Rich, and Zhongju Zhang. 2006. "Tapping the Power of Text Mining." *Communications of the ACM*.

Felizardo, Katia R. et al. 2011. "Using Visual Text Mining to Support the Study Selection Activity in Systematic Literature Reviews." *2011 International Symposium on Empirical Software Engineering and Measurement*: 77–86.

Ferreira, Manuel P., Nuno R. Reis, and Rui Miranda. 2015. "Thirty Years of Entrepreneurship Research Published in Top Journals: Analysis of Citations, Co-Citations and Themes." *Journal of Global Entrepreneurship Research* 5(1): 17. http://www.journal-jger.com/content/5/1/17.

Figueroa-Armijos, Maria, and Thomas G. Johnson. 2016. "Entrepreneurship Policy and Economic Growth: Solution or Delusion? Evidence from a State Initiative." *Small Business Economics* 47(4): 1–15.

Fluck, Juliane, and Martin Hofmann-Apitius. 2014. "Text Mining for Systems Biology." *Drug Discovery Today*.

Freeman, Chris. 1995. "The ' National System of Innovation ' in Historical Perspective." *Cambridge Journal of Economics* 19: 5–24.

Gilbert, Brett Anitra, David B. Audretsch, and Patricia P. McDougall. 2004. "The Emergence of Entrepreneurship Policy." *Small Business Economics* 22(3–4): 313–23. http://0-link.springer.com.innopac.up.ac.za/article/10.1023/B:SBEJ.0000022235.10739.a8%5Cnhttp://0-link.springer.com.innopac.up.ac.za/content/pdf/10.1023/B:SBEJ.0000022235.10739.a8.pdf.

Glanzel, W. 2012. "Bibliometric Methods for Detecting and Analysing Emerging Research Topics." *Profesional De La Informacion* 21(2): 194–201.

Gómez-Barroso, José Luis, Claudio Feijóo, Manuel Quiles-Casas, and Erik Bohlin. 2016. "The Evolution of the Telecommunications Policy Agenda: Forty Years of Articles in Telecommunications Policy." *Telecommunications Policy*.

Grimm, Heike. 2006. "Entrepreneurship Policy and Regional Economic Growth BT - Innovative Comparative Methods for Policy Analysis: Beyond the Quantitative-Qualitative Divide." In eds. Benoît Rihoux and Heike Grimm. Boston, MA: Springer US, 123–44. https://doi.org/10.1007/0-387-28829-5_6.

HaCohen-Kerner, Yaakov. 2003. "Automatic Extraction of Keywords from Abstracts." In *Knowledge-Based Intelligent Information and Engineering Systems: 7th INternational Conference, KES 2003, Oxford, UK, September 2003. Proceedings, Part I.*, eds. Vasile Palade, Robert J Howlett, and Lakhmi Jain. Berlin, Heidelberg: Springer Berlin Heidelberg, 843–49. https://doi.org/10.1007/978-3-540-45224-9_112.

Hart, D M. 2003. "The Emergence of Entrepreneurship Policy: Governance, Start-Ups, and Growth in the U.S. Knowledge Economy." In *The Emergence of Entrepreneurship Policy: Governance, Start-Ups, and Growth in the U.S. Knowledge Economy*, , 1–16.

Hotho, Andreas, Andreas Nürnberger, and Gerhard Paaß. 2005. "A Brief Survey of Text Mining." *LDV Forum - GLDV Journal for Computational Linguistics and Language Technology* 20: 19–62. http://www.kde.cs.uni-kassel.de/hotho/pub/2005/hotho05TextMining.pdf.

Ishikiriyama, Célia Satiko, Diego Miro, and Carlos Francisco Simões Gomes. 2015. "Text Mining Business Intelligence: A Small Sample of What Words Can Say." In *Procedia Computer Science*,.

Jain, Anil K. 2010. "Data Clustering: 50 Years beyond K-Means." *Pattern Recognition Letters* 31(8): 651–66.

Kontostathis, April, Lynne Edwards, and Amanda Leatherman. 2010. "Text Mining and Cybercrime." In *Text Mining: Applications and Theory*,.

Landström, Hans. 2015. "What Makes Scholarly Works 'Interesting' in Entrepreneurship Research? Learning from the Past." In *Rethinking Entrepreneurship: Debating Research Orientations*, , 147–70.

Lerner, Josh. 2013a. "Entrepreneurship, Public Policy, and Cities." *World Bank Sixth Urban Research and Knowledge Symposium* (November): 1–19.

———. 2013b. "The Boulevard of Broken Dreams: Innovation Policy and Entrepreneurship." *Innovation Policy and the Economy* 13(1): 61–82. http://www.journals.uchicago.edu/doi/10.1086/668239.

Liu, Xinhai et al. 2010. "Weighted Hybrid Clustering by Combining Text Mining and Bibliometrics on a Large-Scale Journal Database." *Journal of the American Society for Information Science and Technology* 61(6): 1105–19.

Lundström, A, and L Stevenson. 2002. 1 of the E *On the Road to Entrepreneurship Policy*.

Lundstrom, A, and L A Stevenson. 2006. *Entrepreneurship Policy: Theory and Practice*. Springer US. https://books.google.es/books?id=7jHQqqGzojAC.

Lundström, Anders, and Lois A. Stevenson. 2005. 9 *Entrepreneurship Policy: Theory and Practice*. Boston: Kluwer Academic Publishers.

Lundvall, Bengt-Åke. 1992. National systems of innovation Towards a theory of innovation and interactive learning *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. http://books.google.com/books?id=iDXGwacw-4oC&pgis=1.

Manning, Christopher D., Prabhakar Ragahvan, and Hinrich Schutze. 2009. "An Introduction to Information Retrieval - Chapter 16: Flat Clustering." *Information Retrieval*.

Mason,Colin; Brown, Ross. 2014. *ENTREPRENEURIAL ECOSYSTEMS AND GROWTH ORIENTED ENTREPRENEURSHIP*.

Mazzarol, Tim, and Thierry Volery. 2015. “The Evolution of the Small Business and Entrepreneurship Field : A Bibliometric Investigation of the Articles Published ...” *International Small Business Journal* 33 (4)(June): 374–96.
Meyer, M. et al. 2014. “Origin and Emergence of Entrepreneurship as a Research Field.” *Scientometrics* 98(1): 473–85.
Miner, Gary D., John Elder, and Robert A. Nisbet. 2012. Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*.
Minniti, Maria. 2008. “The Role of Government Policy on Entrepreneurial Activity: Productive, Unproductive, or Destructive?” *Entrepreneurship: Theory and Practice* 32(5): 779–90.
Moore, J. F. 1993. “Predators and Prey: A New Ecology of Competition.” *Harvard Business Review* 71(3): 75–86.
Moore, James F. 1996. Leadership *The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems*. http://books.google.com/books?id=At7HQgAACAAJ&pgis=1.
Nelson, R R. 1993. booksgooglecom *National Innovation Systems: A Comparative Analysis*.
Ngai, E. W.T., Li Xiu, and D. C.K. Chau. 2009. “Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification.” *Expert Systems with Applications*.
Ngai, E W T, and P T Y Lee. 2016. “A Review of the Literature on Applications of Text Mining in Policy Making.” *20th Pacific Asia Conference on Information Systems (PACIS 2016)*.
OECD. 2010. “High-Growth Enterprises : What Governments Can Do To Make A Difference.” *OECD Studies on SMEs and Entrepreneurship*.
Porter, Martin. 2001. “Snowball: A Language for Stemming Algorithms.” *Snowball*: 1–15. http://snowball.tartarus.org/texts/introduction.html.
Ramlogan, Ronnie, and John Rigby. 2012. “The Impact and Effectiveness of Entrepreneurship Policy.” *NESTA Compendium of Evidence on Innovation Policy Intervention*.
Rosa, Peter J. 2013. “Recent Trends in Leading Entrepreneurship Research: The Challenge for European Researchers : Entrepreneurship Research Journal.” *Entrepreneurship Research Journal* 3(1): 35–43.
Seo, Joung Hae, and Eun Mi Park. 2018. “A Study on Financing Security for Smartphones Using Text Mining.” *Wireless Personal Communications*.
Shane, Scott. 2009. “Why Encouraging More People to Become Entrepreneurs Is Bad Public Policy.” *Small Business Economics* 33(2): 141–49.
Shane, Scott, and S. Venkataraman. 2000. “The Promise of Entrepreneurship as a Field of Research.” *Academy of Management Review* 25(1): 217–26.
Stam, Erik. 2014. *The Duch Entrepreneurial Ecosystem*. Utrech. ssrn.com/abstract=2473475.
———. 2015. “Entrepreneurial Ecosystems and Regional Policy: A Sympathetic Critique.” *European Planning Studies* 23(9): 1759–69. http://www.tandfonline.com/doi/full/10.1080/09654313.2015.1061484.
Van Stel, André, David J. Storey, and A. Roy Thurik. 2007. “The Effect of Business Regulations on Nascent and Young Business Entrepreneurship.” *Small Business Economics* 28(2–3): 171–86.
Stevenson, L, and A Lundström. 2002. Swedish Fo *Entrepreneurship Policy for the Future*. http://www.donnerenviedentreprendre.com/documentation/IMG/pdf/Entrepreneurship_policy_for_the_future_-_Introduction.pdf.
Storey, D J. 1994. Thomson Learning Emea *Understanding the Small Business Sector*.
Verheul, Ingrid, Sander Wennekers, David Audretsch, and Roy Thurik. 2001. “An Eclectic Theory of Entrepreneurship.” *Erasmus*: 48. http://www.tinbergen.nl.
———. 2002. “An Eclectic Theory of Entrepreneurship: Policies, Institutions and Culture.” In *Entrepreneurship: Determinants and Policy in a European-US Comparison*, , 11–81. http://hdl.handle.net/10419/85867%5Cnhttp://www.tinbergen.nl.
Welter, Friederike. 2011. “Contextualizing Entrepreneurship—Conceptual Challenges and Ways Forward.” *Entrepreneurship: Theory and Practice* 35(1): 165–84.
Welter, Friederike, and Frank Lasch. 2008. “Entrepreneurship Research in Europe: Taking Stock and Looking Forward.” *Entrepreneurship: Theory and Practice* 32(2): 241–48.
Wennekers, Sander; Thurik, Roy. 1999. “Linking Enterpreneurship and Economic Growth.” *Small Business Economics* 13: 27–55. http://link.springer.com/article/10.1023/A:1008063200484.
Wennekers, Sander, and Roy Thurik. 1999. “Linking Entrepreneurship and Economic Growth.” *Small business economics* 13: 27–55. http://link.springer.com/article/10.1023/A:1008063200484.
White, George O. et al. 2016. “Trends in International Strategic Management Research From 2000 to 2013: Text Mining and Bibliometric Analyses.” *Management International Review*.
Zahra, S. A., and M. Wright. 2011. “Entrepreneurship’s Next Act.” *Academy of Management Perspectives* 25(4): 67–83. http://amp.aom.org/cgi/doi/10.5465/amp.2010.0149.
Zanasi, Alessandro. 2009. “Virtual Weapons for Real Wars: Text Mining for National Security.” In *Advances in Soft Computing*,.