



Regional Competitiveness: A Structural-Based Topic Analysis on Recent Literature

Maria Gabriella Grassia¹ · Marina Marino¹ · Rocco Mazza¹ · Michelangelo Misuraca²  · Emma Zavarrone³ · Martha Friel³

Accepted: 6 May 2022
© The Author(s) 2022

Abstract

The concept of Regional Competitiveness (RC), in the last years, had a considerable influence on the development of strategies and policies focused on sub-national areas from an economic and social perspective. In particular, RC attracted the substantial attention of policymakers due to the possibility of evaluating and comparing different territories without any established political or conceptual framework. Conversely, scholars have long debated the meaningfulness of territorial competitiveness itself and, in case, which dimensions have to be included. This study reviewed the most recent scientific literature concerning RC with a bibliometric approach based on topic modelling. The use of a textual-based statistical approach offered an interesting insight into the RC research domain. We highlighted the topics discussed by scholars, showing the patterns emerging across years and from different publication types, and marked the differences between the vocabulary used by authors coming from EU and non-EU countries. Finally, a comparison between the RC issues coming from the literature analysis with the RC issues defined by policymakers tried to link the two standpoints followed by the institutions and the academia in a more comprehensive conceptual framework.

✉ Michelangelo Misuraca
michelangelo.misuraca@unical.it

Maria Gabriella Grassia
mariagabriella.grassia@unina.it

Marina Marino
marina.marino@unina.it

Rocco Mazza
rocco.mazza@unina.it

Emma Zavarrone
emma.zavarrone@iulm.it

Martha Friel
martha.friel@iulm.it

¹ DSS, University of Naples Federico II, Naples, Italy

² DiScAG, University of Calabria, Rende, Italy

³ DSU, IULM University, Milan, Italy

Keywords Competitiveness theory · Science mapping · Topic modelling

1 Introduction

The concept of competitiveness can be considered from different standpoints, depending on the domain and the nature of the involved entities (Aiginger, 2006; Thyroff & Kilbourne, 2018). As a general matter, all the possible definitions encompass the idea of successfully keeping up and then prevailing over other entities in a given context. In this sense, it is possible to regard competitiveness at an individual or organisation level, considering several case-specific determinants. During the years, competitiveness has attracted particular attention of policymakers and scholars, particularly in its economic facet. In this acceptance, competitiveness has been studied both from a micro and macro perspective, focusing on increasingly wider layers based on firms, industries or the whole nations (Shvindina, 2020).

Porter (1990) originally referred to the competitiveness at a national level to the achievement and the maintenance of advantageous positions in several key industries, specifically looking at productivity. Policymakers (and politicians) seemed to be very interested in this view of competitiveness because it offers the possibility of measuring differences among countries—discussing the relative performance of economies in a benchmarking sense—without employing any other political or conceptual framework. Conversely, despite the relevance of Porter’s studies in the broader competitiveness theory, this extension to territories has been intensely debated by scholars (e.g., Krugman, 1994; Moon & Peery, 1995; Lall, 2001) and still leading the discussion on the reference literature. According to Esser et al. (1996), the competitiveness of a territory can be considered as a meta-level and expressed as the ability of a place to generate high and rising incomes, and improve the livelihoods of residents. This interpretation in terms of wealth, also stated by Aiginger (2006), led to an emphasis on ‘soft’ or less tangible assets as sources of competitiveness, including together with the productive capacity also factors like the human capital, the innovative capacity and the sustainability.

Since the beginning of the 2000s, policymakers and then scholars, mainly focused on a sub-national layer of competitiveness, bringing attention to a regional dimension. In a globalised economy, regions became an essential source of development and organisation (Malecki, 2004; Werker & Athreye, 2007). Regional Competitiveness (RC) offers a general fuzzy concept that covers aspects concerning firms as well as residents of a region. Overcoming the idea that attracting international investments is the only way to make regions more competitive, the regional policy started to focus also on the development of the critical aspects of the domestic environment. Beyond the practical implications of such a novel perspective, scholars have few points of agreement in defining the key aspects underpinning RC. Nevertheless, creating a taxonomy is still challenging because of the coexistence of different domains and positions in the same framework, frequently counterposed due to the different doctrines they refer to. At the same time, there are cultural differences among the scholars—depending on their geographical origin or localisation—that can influence their view about competitiveness and its drivers. In addition, the classificatory schemes proposed in the past are affected by the mutation induced by the last industrial revolutions—still in progress—and by the propulsive effects deriving from the activation of virtuous circles between academia, industry and government (Etzkowitz, 2003; Leydesdorff & Meyer, 2012).

A systematic literature review of the different scientific contributions concerning RC can help highlight the various conceptualisations. The analysis of scientific contributions based on their content and the developed themes can be faced considering a semantic and a statistical perspective. In this sense, the definition of a taxonomy can take place on a qualitative or a quantitative level. To the qualitative level belong all those methods based on the manual attribution of documents to groups, as the content analysis (e.g., Berelson, 1952; Krippendorff, 1980; Mayring, 2004). This approach relies on a series of steps that aim to identify the most prominent topics, starting from identifying the meaning units in a single sentence through successive aggregations (condensed meaning units, codes, categories and themes). Although the steps have been often represented as a hierarchy, for the sake of convenience, the classification process offered by content analysis is a continuous skip between coding and categorisation, returning to the raw data to reflect on the initial choices (Kolbe & Burnett, 1991). The inherent subjectivity of this approach represents a weakness (Hsieh & Shannon, 2005), since one topic or one point of view may prevail over another due to the researcher's background and sensitivity. Therefore, creating a taxonomy requires a deep knowledge of the domain under examination and a significant amount of time. The quantitative level can be seen in the more general framework of the so-called textual statistics, which allows transforming a textual body into a numerical data structure and performing a synthesis to identify the essential topics characterising the text (Lebart, 2020). In this way, it is possible to automatically (or at least semi-automatically) create a taxonomy and visualise the latent lexical structures embodied into the scientific contributions, leaving the interpretation at the end of the process.

Trying to systematise the knowledge emerging from the RC literature and identify the main characterising topics, in this paper a quantitative study is proposed and implemented, in the framework of the science mapping procedures commonly used in bibliometrics (e.g., He, 1999; Börner et al., 2003; Cuccurullo et al., 2016; Aria et al., 2020). This approach, to the best of our knowledge, was few explored in competitiveness studies and specifically focused on some theme of interest (Teixeira & De Matos Ferreira, 2018; Teixeira & Pocinho, 2020). The strategy here followed belongs to the sphere of text-based analyses. It stands as a novel approach compared to the previous attempts of classification made so far with other methods. We propose to analyse the RC literature with the structural topic model (STM) introduced by Roberts et al. (2014). This approach allows extracting a prior-established number of topics by including in the model a set of covariates associated with the surveyed documents. One of the strength points of STM is the possibility of knowing which documents contributed more in characterising the different topics, incorporating context information in the identification process.

Three research objectives can mainly express the challenges posed by this study. The first research objective (RO1) involves the identification of the prominent topics embodied in the most recent RC literature, aiming at bridging the gaps caused by the multi-disciplinary of the domain. The second research objective (RO2) refers to the identity of RC topics between EU and non-EU authors. This question tries to investigate the invariance of the RC topics, focusing in particular on the country affiliation of the authors. The interest is verifying whether there is an exogenous effect of the localisation of the scholars among the selected topics. Implicitly, it aims at evaluating how the different geographic locations can influence the attention to the different topics in connection with the socio-economic development of the reference territory. A third research objective (RO3) comes from the necessity of comparing the issues emerging from the literature and the dimensions empirically measured in RC. Even knowing that there are some differences between the theoretical backgrounds underlying the themes developed in the literature and the pillars of RC, the

effort is to connect the different aspects characterising the RC both from a theoretical and an empirical standpoint.

The paper is structured as follows. Section 2 briefly reviews the RC literature, underlying the different views at the basis of the research area, and the conceptual evolution across years. In Sect. 3, we introduce the main features of the STM here implemented. Section 4 describes the data used in the analysis and shows the main descriptive statistics. After presenting the study's findings with respect to the different research objectives in Sect. 5, we conclude the work with Sect. 6, underlying the significant theoretical and empirical implications as well as the limits of the study.

2 Literature Review on Regional Competitiveness

Before identifying the prominent topics embodied in the most recent RC research frontiers, an overview of the huge reference literature is necessary. Defining RC is not an easy task since this concept encompasses different traits at an economic, managerial, political, social and, last but not least, geographical level (OECD, 2012). This multi-disciplinary nature makes complex referring to RC with a generic and simplistic vision. Earlier studies approached the competitiveness of territories from two perspectives: as a set of determinants influencing the level of country productivity and a factor of the sustained improvement in population's well-being. The first view of competitiveness can be found, for example, in Schwab and Porter (2008) and it is implicitly encompassed in the Global Competitiveness Index developed by the World Economic Forum. The second view of competitiveness can be found instead—among others—in Meyer-Stamer (2008), in the framework of the so-called systemic competitiveness. The concepts at the basis of the latter perspective partially overlap with the rationale of the Human Development Index developed by the United Nations, in which the development (and the competitiveness) of a country is not declined as economic growth alone. A rich literature flourished starting from the 2000s (e.g., Porter, 1998, 2001; Camagni, 2002; Cellini & Soci, 2002; Porter & Ketels, 2003), aiming at advancing in the competitiveness definition and its quantitative measurement.

The theme of RC rose to prominence in the EU with the so-called Lisbon strategy, launched in March 2000. The primary intention of this initiative was to make the economic processes capable of a sustainable growth—given a pace-stated and dynamic world such as the one we live in—with an increase in the employment and better quality jobs as well as a greater social cohesion. In this standpoint, the two perspectives were combined, leading to a definition of RC based on the management of resources and capacities to obtain a sustained increase in business productivity as well as in the well-being of the region's population (Dijkstra et al., 2011). These ambitious goals have strengthened the interest of international agencies (e.g., OECD, 2005, 2009) and attracted the attention of those scholars already studying the national economic growth paradigms related to the models of competitiveness at an international level. RC attracted so much attention because, differently from the traditional themes debated in the economic doctrine, it is tangential to several disciplines and has practical implications in the development of an area both from an economic and socio-cultural viewpoint (Boschma, 2004; Kitson et al., 2004; Komarova et al., 2014). In 2004, the Regional Studies journal dedicated a special issue to the theme of RC (e.g., Budd & Hirmis, 2004; Polenske, 2004). Since then, the academic interest in this issue has grown more and more, summing a large number of publications but without a unified framework, a shared definition nor an agreement upon how the measurement of this

concept should occur (Huggins & Williams, 2011; Cristelli et al., 2015; Annoni & Weziak-Bialowolska, 2016).

The multifaceted nature of RC, as stated above, makes challenging to provide a definition. The debate concerning themes like industrial organisation, economic retardation or the 'new competition' influences the diverse strategies and actions that can be carried out for improving the socio-economic conditions of a given territory (Budd & Hirmis, 2004). The formulation of RC has also been affected by the previous debate between economists regarding a change in the concept of advantage, from the Ricardian perspective of comparative advantage to the more recent definition of Porter of competitive advantage. This transition, based on a view of regional productivity as the primary engine and measure of competitiveness, was not completely uncritical (Krugman, 1996; Boltho et al., 1997), coming up with an initial balance between the work of Kitson et al. and Budd and Hirmis. Kitson et al., in particular, proposed a wide-ranging definition combining the three approaches debated in the reference literature, i.e. the neo-classical theory, the increasing returns theory, and endogenous growth theory. At the same time, these authors indirectly included also the Porter's view, merging three different conceptions of RC: a first standpoint considering regions as sites of specialisation, a second one looking at regions as a source of increasing returns, and a third one considering regions as hubs of knowledge and economic trade (Garcia-Alvarez-Coque et al., 2020). In a further attempt of unifying some RC key elements, Martin (2005) elaborated the conceptual model of regional competitiveness that built by considering several layers: regional outcomes, regional outputs, regional through-puts and RC determinants. Among the RC determinants, Martin included the production factors (labour, capital and land), the production environment, the infrastructures and the human resources primarily, together with some secondary drivers ranging from internationalisation and technological development to the environment and the demographic aspects. Budd and Hirmis made some additional reflections by putting into a single framework the comparative and competitive advantage with different geographical levels (regional and national). Despite this formidable effort, many scholars questioned these forms of transition from a comparative to a competitive advantage, along with the shift from an international to a regional view. Foray (2015) noted that the literature increasingly emphasised the importance of considering competitiveness at a regional and a national level. In this complex overview of RC, together with socio-cultural and economic aspects, an adequate level of education, clean energy availability, and quality of life find a suitable place. Unfortunately, these latter aspects are not directly measurable. For this reason, in many contributions, they are viewed in a qualitative perspective, not sufficient to realistically capture the different levels of regional development because of the regions' heterogeneity (Cristelli et al., 2015).

Heterogeneity is a widely discussed theme in the RC literature and can be triggered by several factors, from a geographical (Diamond, 1997), cultural (McCloskey, 2010), and biological (Ashraf & Galor, 2013) viewpoint. This source of differentiation, regarded as the intersection of socio-economic and cultural components, can be the key to success for many regions (Capello et al., 2009; Lavrinenko et al., 2019; Pietrzak et al., 2017; Sagiyeva et al., 2018; Zeibote et al., 2019). At the same time, human capital factors were considered of primary importance for RC. The body of aspects concerning human capital transferred to RC studies the Economics branch known as the neo-classical Economics (see Becker, 1993), stating that the availability and the quality of human capital are relevant drivers of the economic growth (Barro, 1989; Lucas Jr, 1988; Mankiw et al., 1992; Solow, 1956). A study conducted at a regional level in OECD countries, dating back to 2007, highlighted the determinant role of human capital in RC, claiming that higher shares of poorly educated

people are more detrimental than lower shares of highly educated people (OECD, 2007). Human capital is likely to influence economic growth through higher labour productivity and technological progress, increasing the overall progress of countries and regions (Annoni & Weziak-Bialowolska, 2016).

Taking into account the different positions that emerged from the literature review, we considered reasonable to define RC as in the work of Annoni and Dijkstra (2013): RC is the ability of a region to offer an attractive and sustainable environment for firms and residents to live and work. Consistently with this vision, RC determinants can be found in the 11 dimensions used to build the EU Regional Competitiveness Index (RCI), the so-called pillars:

- P01: Institutions
- P02: Macroeconomic stability
- P03: Infrastructure
- P04: Health
- P05: Quality of Primary and Secondary Education
- P06: Higher education/training and lifelong learning
- P07: Labour market efficiency
- P08: Market Size
- P09: Technological Readiness
- P10: Business Sophistication
- P11: Innovation

The 11 pillars are derived from the sequential aggregation of 74 indices observed for each EU country at a regional level. They can be grouped in basic (P01–P05), efficiency (P06–P08) and innovation (P09–P11) sub-indexes (D’Urso et al., 2019). The resulting composite indicator has been of extreme importance for studying RC in the EU, since it allows tracking competitiveness of 268 regions at the NUTS-2 level across all the EU Member States over time¹. Nevertheless, the conceptual framework offered by RCI can be extended trying to define competitiveness at a regional level also for non-EU areas (e.g., González Catalán, 2021).

To analyse the most recent literature concerning RC, we decided to perform an automatic topic extraction from the scientific publications that appeared in 2016–2020. In particular, we considered an unsupervised approach able to highlight the per-document topic distributions and the per-topic word distribution simultaneously, known as topic modelling. Because of the evolution of the RC over time, and because of the influence that other factors—like the geographical localisation of scholars—may have on the RC research, we carried on a structural topic model (STM), in which it is possible to incorporate some external information on the analysed textual body.

3 Structural Topic Model: A Gentle Introduction

STM is a particular extension of topic models allowing researchers to find the main themes within a set of documents automatically and estimate the relationships between topics and some text metadata of interest. Topic models are usually defined as unsupervised contents

¹ See <https://ec.europa.eu/eurostat/web/nuts/background>.

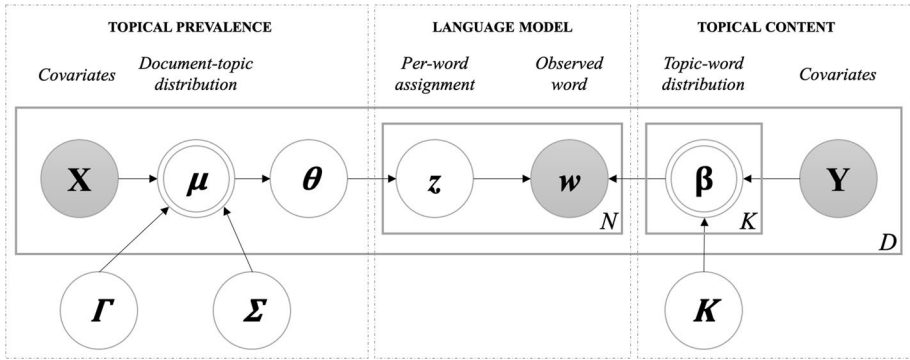


Fig. 1 Plate diagram for the structural topic model

extraction methods (see Misuraca & Spano, 2020), able to infer the content of textual collections in terms of latent topics (Blei et al., 2003; Grimmer, 2010; Wang & Blei, 2011). The rationale is defining topics as distributions over a set of words (namely the vocabulary) that semantically represent interpretable discussion themes. Topic models can be included in the more general class of mixed-membership models (Erosheva et al., 2004), since each document is characterised by a mixture of topics and at the same time, each word within a document characterises topics. STM (Roberts et al., 2014) encompasses document metadata as covariates into the prior distributions for document-topic proportions and topic-word distributions, including additional information in the statistical inference procedure. The model can be decomposed into three sub-models: (a) a topical prevalence model, which controls how documents are allocated to topics as a function of covariates; (b) a topical content model, which controls words frequency in each topic as a function of covariates; (c) a core language model, which combines the two sources of variation to produce the actual words in each document.

Formally, documents are indexed as $d \in \{1, \dots, D\}$, while words are indexed as $n \in \{1, \dots, N_d\}$. The observations $w_{d,n}$ are occurrences of words from a vocabulary indexed by $v \in \{1 \dots V\}$. The number of topics is set by the researcher and indexed by $k \in \{1, \dots, K\}$. Two matrices represent the document-level additional information. \mathbf{X} is the $D \times P$ matrix with topical prevalence covariates, while \mathbf{Y} is $D \times A$ matrix with topical content covariates. The rows of these matrices—each representing a vector of covariates for a given document—are denoted by \mathbf{x}_d and \mathbf{y}_d , respectively. The model can be depicted in plate notation as in Fig. 1, where each box is a replicate over the enclosed nodes (coloured in grey if observed or in white if latent).

The data generating process for a document d , given the K topics, the observed words $w_{d,n}$, and the design matrices \mathbf{X} and \mathbf{Y} for topical prevalence and topical content, respectively, can be summarised as in the following:

$$\gamma_k \sim \text{Normal}(0, \sigma_k^2 I_p) \tag{1}$$

$$\theta_d | \mathbf{x}_d \Gamma, \Sigma \sim \text{LogisticNormal}(\mu = \mathbf{x}_d \Gamma, \Sigma) \tag{2}$$

$$\beta_{d,k} \propto \exp(\mathbf{m} + \kappa_k^{(t)} + \kappa_{y_d}^{(c)} + \kappa_{y_{d,k}}^{(i)}) \tag{3}$$

$$\mathbf{z}_{d,n} | \boldsymbol{\theta}_d \sim \text{Multinomial}(\boldsymbol{\theta}_d) \quad (4)$$

$$\mathbf{w}_{d,n} | \mathbf{z}_{d,n}, \boldsymbol{\beta}_{d,k=z_{d,n}} \sim \text{Multinomial}(\boldsymbol{\beta}_{d,k=z_{d,n}}) \quad (5)$$

The process starts by drawing the document-level attention to each topic from a logistic-normal generalised linear model (Eqs. 1 and 2), based on a vector of covariates \mathbf{x}_d and considering a $P \times (K - 1)$ matrix of coefficients $\boldsymbol{\Gamma}$ for the topical prevalence and a $(K - 1) \times (K - 1)$ covariance matrix $\boldsymbol{\Sigma}$. On the other side, given a document-level content covariate \mathbf{y}_d , the topic-specific distribution over words is formed by representing each topic k with the V -dimensional baseline word distribution \mathbf{m} (see Airolidi et al., 2004), the topic-specific deviation $\boldsymbol{\kappa}_k^{(t)}$, the covariate group deviation $\boldsymbol{\kappa}_{y_d}^{(c)}$ and the interaction between the two $\boldsymbol{\kappa}_{y_d,k}^{(i)}$ (Eq. 3). Finally, for each word in a document, topic assignment based on the document-specific distribution over topics and word assignment to a topic are drawn from multinomial models (Eqs. 4 and 5). The latter represents the core language model. Differently from the basic version of topic model defined by Blei and Lafferty (2007), in which $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ are global parameters shared by all documents, they are specified as a function of the document-level covariates in the STM.

In implementing STM, researchers have to make two critical choices concerning the covariates to include for the topical prevalence and the topical content, and the number of topics to consider in the model.

As stated above, it is possible to use covariates both on the topical prevalence and topical content. The aim is to evaluate the impact of covariates to study how or by whom a particular topic is discussed and how covariates can affect the words representing a topic. According to Roberts et al. (2019), to estimate the effect of covariates on topics, it is possible to carry on a regression where topic-proportions are the outcome variable, obtaining the conditional expectation of topic prevalence given document characteristics. Nevertheless, depending on the investigated phenomenon and the research goals, it is also possible to use covariates only on one of the two sides or to not use covariates at all. In the latter case, the modelling approach can be seen as a (fast) implementation of the correlated topic model (CTM: Blei & Lafferty, 2007).

As in other topic models, STM requires setting a fixed number of topics. In the framework of topic modelling, a variety of methods and algorithms for setting the optimal number of topics have been proposed (e.g., Buntine, 2009; Chang & Blei, 2009). However, there is no shared opinion about which is the best strategy. Many approaches are based on some distributive analysis, as for example in the proposals of Griffiths and Steyvers (2004), Cao et al. (2009), Arun et al. (2010) or Deveaud et al. (2014). Here after, we followed a data-driven approach commonly adopted for STM, calculating the held-out likelihood (Wallach et al., 2009) and applying a residual analysis (Taddy, 2012). The held-out likelihood represents how well a model predicts words within a document, computing the probability of a word to appear within a document when it has been removed from the document itself in the estimation step (Asuncion et al., 2009; Hoffman et al., 2013). The residual analysis tests the variance overdispersion of the multinomial distribution within the data generating process performed by STM. The results of these strategies can be read concurrently to find the best approximation to the topic number, selecting the highest held-out likelihood value that minimises the residuals' check.

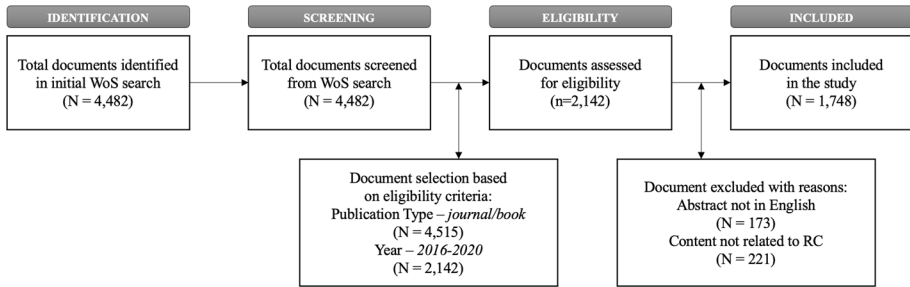


Fig. 2 PRISMA flow diagram used for the present study

4 Data Description and Preparation

To determine which topics have been the most discussed in the most recent RC literature, we accessed on January 2021 the Web of Science (WoS) database to obtain a bibliographic dataset. WoS—initially developed by the Institute for Scientific Information and at present maintained by Clarivate Analytics—is one of the primary sources to explore the publications of a research field. It includes different citation databases focused on specific fields (e.g., the Social Science Citation Index for Social Science), covering more than 20,000 journals, conference proceedings and books. Alternative databases can be either considered to retrieve bibliometric data, like Scopus or Google Scholar, and there is an intense debate concerning which data source is better (see de Winter et al., 2014; Harzing & Alakangas, 2016). Nevertheless, the quality of WoS information is often considered the highest. For this reason, we decided to consider only this database. We used the query [regional AND competitiveness] to retrieve the abstracts of publications related to this research area. The number of texts downloaded initially was 4482. Subsequently, we selected only documents published in journals (articles and reviews²) and books (books, chapters in edited books and conference proceedings), and we focused on the last 5 years publications (from 2016 to 2020) to consider the most recent literature, reducing the dataset to 2142 publications. A careful review process led to the exclusion of documents with an abstract not in English and some publications not related to RC, obtaining at the end of the process 1748 texts.

Figure 2 shows the information flow through the different searching steps, mapping out the number of identified publications, the included and excluded ones, and the reasons for exclusions (see Moher et al., 2009).

Before performing the STM³, as in other text analyses, we had to pre-process the abstract of the publications. Texts can be seen as a sequence of characters, therefore a multistage process has to be carried out to transform unstructured data into structured data. The 1748 abstracts have been parsed and tokenised to obtain a set of distinct strings (namely, the tokens) separated by blanks and punctuation marks. These tokens correspond to the words used in the documents. The particular scheme obtained with tokenisation is commonly

² In the Clarivate's Journal Citation Report system, any article containing the words 'review'/'overview' in the title or published in the review sections of journals are coded as reviews. Moreover, articles with more than 100 references are automatically coded in the same way (source: <https://clarivate.com/webofsciencereport/essays/impact-factor>).

³ The analysis presented in the paper has been carried out using the R package *stm*, developed by Roberts et al. (2019).

known as bag-of-words. Each document is seen as a multi-set of its tokens in this scheme, disregarding grammatical and syntactical roles but keeping multiplicity. Typically, words are arranged in a set of unique entries (namely, the types) together with the corresponding number of occurrences in the collection, forming the vocabulary. After atomising the documents into their basic components, we reduced language variability to avoid possible noise sources and improve the effectiveness of the next analytical steps. In particular, we normalised the spelling of the different tokens (e.g., multiwords with and without the hyphen) and brought back each inflected word to its canonical form (e.g., nouns and adjectives from plural to singular, verbs to the infinitive form). Furthermore, we lexicalised the collocations by considering the couple of words with the highest number of co-occurrences (e.g., 'labour' + 'market' → 'labour_market'). Finally, we pruned non-informative words (e.g., pronouns, prepositions, conjunctions) and non-alphabetic characters (e.g., numbers) from the documents, obtaining a vocabulary of 3984 types.

At the end of pre-process, each text has been represented as a document-vector and included in a document \times word matrix, with 1748 rows (documents) and 3984 columns (words). In order to reduce the dimensionality of this matrix, we filtered sparse words (with a sparsity threshold of 2%) and excluded empty documents (documents without retained tokens because of the pre-processing). On the resulting 1743×3902 matrix, we performed the STM. For each document included in the quantitative study, we considered as metadata a set of additional variables:

1. *Publication year (PY)* referring to the year of publication of the documents, a period between 2016 and 2020 was set;
2. *Publication type (PT)* to describe the type of publication, a variable with two categories (journal and book) was coded. The category book encompasses both authored books and edited books;
3. *Principal author country (PAC)* to identify the geographical localisation of publications, a dichotomous variable with EU and non-EU categories were coded, using the country of the corresponding author of each publication;
4. *Total citations (TC)* to evaluate the impact of publications, a continuous variable was considered for counting the number of associated citations.

In Fig. 3, the distributions of metadata are reported. In each year belonging to the considered period, the number of publications was over 300, with a maximum in 2016 (371 documents) and a minimum in 2020 (307 documents), showing a slight decrease of scientific production on RC⁴. Concerning the publication type, we found that 70% of documents were articles while 30% were books. A notable aspect pertains the country affiliation variable, since it presents a higher value for the non-EU category (60.47%) in comparison with the EU category (39.53%). Moreover, in the year-wise scientific production, the shares of EU and non-EU authors did not show outliers (Fig. 4).

⁴ It is important to remember that bibliographic databases as WoS have some known update lags. This means that the actual number of publications, especially for the last year, could be greater.

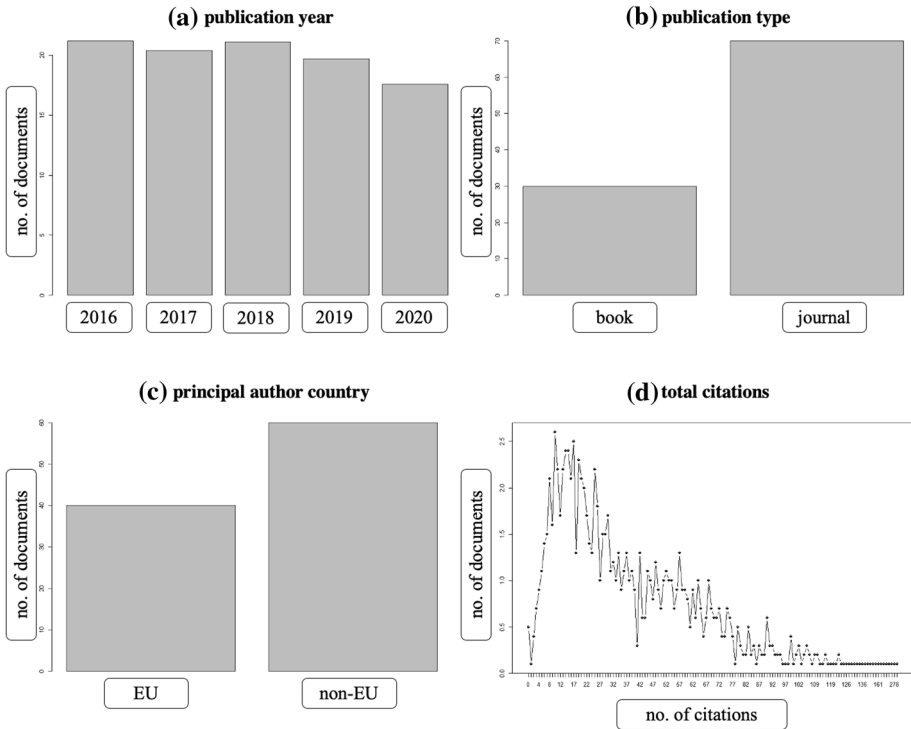


Fig. 3 Metadata distributions over the document collection

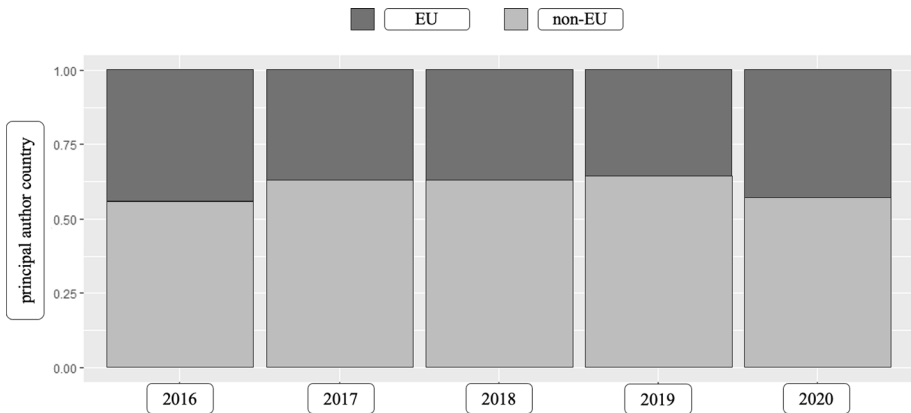


Fig. 4 Year-wise proportion of European and non-European principal authors

5 Topic Model Setup and Main Findings of the Analysis

Before fitting the model, we defined which variables had to be included to evaluate their effect on the topic extraction process. In particular, we considered PY, and PT as covariates of the topical prevalence, whereas we considered PAC for the topical content. The first input that the

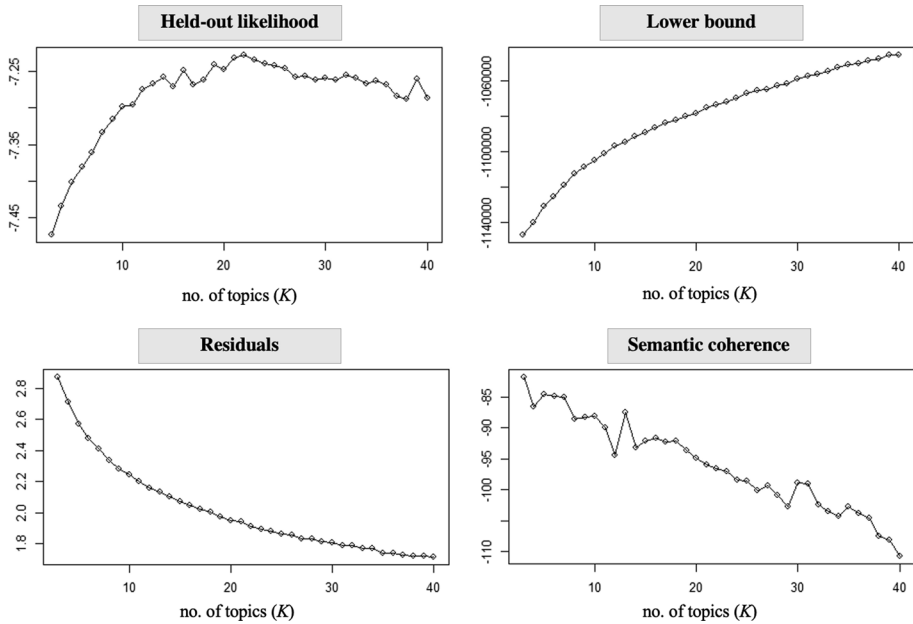


Fig. 5 Optimal topic number diagnostics

model needs is the parameter K , which represents the optimal number of topics to extract. The choice is guided by considering specific metrics, as explained in the following. The estimation of topic proportions is performed using a logistic normal generalised linear model with covariates. Topic proportions are the outcome variable, conditionally predicting the prevalence of each topic as a function of document characteristics. We evaluated vocabulary differences between EU and non-EU authors in discussing the RC themes. For this scope, the model uses a multinomial logit with covariates. The logic is to parameterise the distribution of word occurrences as log-transformed rate deviations from a collection background distribution rates.

In the following, the main findings of STM are reported for the research objectives of the study.

5.1 RO1: Automatic Identification of RC Topics

We estimated the optimal number of topics K to set for the model. Some diagnostics can be used to understand how the models perform at various numbers of topics, considering as metrics the held-out likelihood, its lower bound and residuals and the semantic coherence (Lee & Mimno, 2014). Plots in Fig. 5 show the results obtained from this analysis.

The optimal value for the held-out likelihood is between 15 and 20 topics. We can see an increasing lower bound (i.e., the lower bound of the marginal log-likelihood) and decreasing residuals (i.e., the difference between expected and predicted topics) for these values. In order to choose a value of K , it is necessary to consider also the semantic coherence (Mimno et al., 2011). Semantic coherence considers the conditional

Table 1 Topics label and rank

Rank	Topic	Label	Words
1	T08	Macroeconomics	<i>provision, assessing_competitiveness, comprehensive_assessment, innovative_activity, industrialization</i>
2	T07	Institutions	<i>development_eu, european_union, cohesion_policy, funds, programming_period</i>
3	T12	Technological development	<i>science_technology, competitiveness_innovation, innovation_system, scoreboard, knowledge_creation</i>
4	T04	Business	<i>international_business, intensive_business, territorial_servitization, servitization, family_business</i>
5	T14	Urban development	<i>local_competitiveness, city_regional, currency, city, municipality</i>
6	T01	Growth determinants	<i>share_analysis, employment_growth, growth_region, shift_share, regional_productivity</i>
7	T03	Literature review	<i>article, identity, science, bibliometric, interdisciplinarity</i>
8	T13	China development	<i>polycentricity, subnational, pearl, polycentric, regional_urban</i>
9	T15	International trade	<i>southeast_asian, international_trade, southeast_asia, regional_trade, foreign_economy</i>
10	T02	Education	<i>foreign_student, higher_education, graduate, student, teaching</i>
11	T11	Industry	<i>industrial_structure, footwear, manufacturing_industry, car, regional_industry</i>
12	T09	Tourism	<i>destination, tourist, tourism, tourism_product, attraction</i>
13	T10	Infrastructure	<i>port, rail, route, hinterland, renewable_energy</i>
14	T05	Agriculture	<i>agriculture, place_marketing, farmer, regional_agricultural, agricultural_production</i>
15	T06	Labour market	<i>labour_market, transparency, estate, real_estate, compensation</i>
16	T16	Corporate Social Responsibility	<i>csr, corporate_social, corporate, commitment, social_responsibility</i>

likelihood of the co-occurrence of words in a topic, and it is maximised when the most probable words in a given topic frequently co-occur together. The trade-off between coherence and likelihood level led us to consider a model with 16 topics reasonably appropriate to represent the different research themes developed in the RC literature. According to the themes developed in RC studies, the extracted topics were reasonably interpretable. Table 1 lists the topics extracted from the document collection with respect to their rank. Rank expresses the prevalence of the topics (i.e., the expected topic proportions) for the covariates included in the model. Each topic has been manually labelled using expert knowledge, looking at the word distributions in the topics

(please note: only the top five most probable words per topic are reported below in the table).

Looking at the set of topics, the institutional aspects are declined at different levels, from a supra-national to an urban dimension (e.g., T08, T07, T14). Moreover, T14 confirms the important presence in the literature of studies concerning the relationship between urban dynamics and RC (Budd & Hirmis, 2004; Martin et al., 2012; Nijkamp, 2017). Economic sectors emerged in specific topics as *Business* (T04), *Agriculture* (T05), *Industry* (T11) and *Tourism* (T09). As regards topic T06—labelled as *Labour market*—we noticed that there were both references to employment and housing, highlighting the fundamental relation between regional development and quality of life in some RC studies (see Hämäläinen & Böckerman, 2004; Head & Lloyd-Ellis, 2012). We also found reference to two relevant aspects for regional development as *Education* (T02) and *Technology* (T12). Intellectual capital, in fact, has increasingly been considered as the base for competitive growth and more recently for sustainable regional competitiveness (Audretsch et al., 2012; Januškaite & Užiene, 2018), with a growing number of studies focusing on the role of universities and of university's entrepreneurial activity for regional competitiveness. Particularly noteworthy, a specific topic emerged for China (T13) due to the attention in RC literature to the social and economic changes induced in recent years by the development policies of the Chinese government (e.g., Yeh & Xu, 2008; Butollo, 2015; Wang & Shen, 2017). A comparable interest in RC studies devoted to Southeast Asia (e.g., Cui et al., 2020) also emerged from the topic *International trade* (T15), due to the importance of export in the development of several sites located in this area (McDonald et al., 2008). Finally, highly interesting is the topic T16 referred to the issue of *Corporate Social Responsibility* (Andreoni & Miola, 2016; Porter & Kramer, 2006). The pressure for adopting or improving the use of ethically-oriented practices has influenced the competitiveness paradigm in recent years, both at a private and a public level (Aiginger & Vogel, 2015; Boulouta & Pitelis, 2014).

The topic T03 (*Literature review*, rank 7) represents a meta-discourse on the analysed documents related to the discussion on the RC literature itself in the different contributions concerning this research domain. Due to the lack of centrality with respect to our research objectives, more detailed comments about this topic are omitted in the following of the present study.

The 16 topics extracted through the STM explained what the literature of this research area dealt with within the last five years. It is interesting to note that, although the analysis revealed almost all the major topics traditionally present in previous RC literature, there are some important absences and some issues that do not emerge autonomously but more transversely, characterising the theme of competitiveness in an endogenous way. For example, the regulation issue did not emerge as a defined topic, even if some reference appeared between the terms of several other topics (e.g., 'transparency' in T06). Similarly, the innovation issue emerged across other topics like *Macroeconomic* (T08) and *Technological development* (T12).

5.2 RO2: Influence of Covariates on Topic Prevalence and Content

Concerning the role of metadata in the chosen topic model, in Table 2 we jointly mapped the topics extracted through the STM with the publication years and publication types,

Table 2 Topic prevalence by publication year (PY) and by publication type (PT)

Topic	Publication year					Publication type	
	2016	2017	2018	2019	2020	book (B)	journal (J)
Macroeconomics	0.073	0.121	0.131	0.174	0.081	0.186	0.100
Institutions	0.130	0.076	0.093	0.070	0.033	0.205	0.054
Technological development	0.097	0.085	0.082	0.090	0.088	0.105	0.091
Business	0.057	0.062	0.087	0.081	0.124	0.009	0.108
Urban development	0.086	0.065	0.060	0.076	0.062	0.082	0.071
Growth determinants	0.065	0.054	0.060	0.041	0.059	0.086	0.044
Literature Review	0.054	0.073	0.057	0.052	0.059	0.036	0.065
China development	0.062	0.042	0.054	0.058	0.085	0.032	0.069
International trade	0.059	0.076	0.076	0.058	0.081	0.036	0.081
Education	0.057	0.070	0.063	0.076	0.059	0.045	0.052
Industry	0.078	0.037	0.041	0.041	0.029	0.045	0.040
Tourism	0.073	0.079	0.052	0.073	0.068	0.050	0.076
Infrastructure	0.062	0.051	0.057	0.049	0.091	0.000	0.077
Agriculture	0.024	0.068	0.052	0.041	0.036	0.027	0.042
Labour market	0.016	0.028	0.019	0.017	0.033	0.041	0.020
Corporate social responsibility	0.005	0.014	0.016	0.003	0.013	0.014	0.011

respectively, to highlight the specific proportion of topics with respect to the considered categories.

A comparison among years revealed different patterns of topic proportion, with overall increasing attention on *Macroeconomics* (but with a reduction in the last year), *Business* and *China development* topics, and overall decreasing attention on topics related to *Institutions* and primary and secondary economic sectors such as *Agriculture* and *Industry*, respectively. Topics related to *Labour market* and *Corporate Social Responsibility* seemed to be still marginal in the debate concerning competitiveness at a regional level, with a minor share across the different years. Looking at the publication types, we observed that publications appeared as book chapters or contributions in conference proceedings (both included in book category) treated more conceptual aspects concerning *Macroeconomics*, *Institutions*, *Technological development* and *Urban development*, *Growth determinants*, and less specific sectoral aspects. On the contrary, publications that appeared as journal papers seemed to be focused more equally on the different aspects, with a major presence of topics concerning the sectors involved in the RC discourse.

The localisation of scholars (PAC) was introduced as a covariate for the topical content to evaluate possible differences in their language.

Figures 6, 7 and 8 show which terms within the topics are more associated with EU and non-EU authors. The label size of each term considers the probability of appearing in a topic conditional to covariates' categories instead. The terms are distributed on the x-axis, taking into account the significance with respect to the two categories.

The terms used by scholars to define the different themes suggested some interesting observations on the RC conceptualisation:

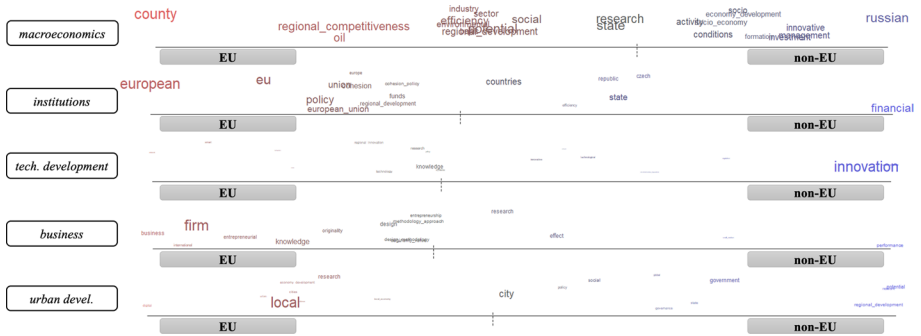


Fig. 6 Topical content representation (topic rank: 1–5)

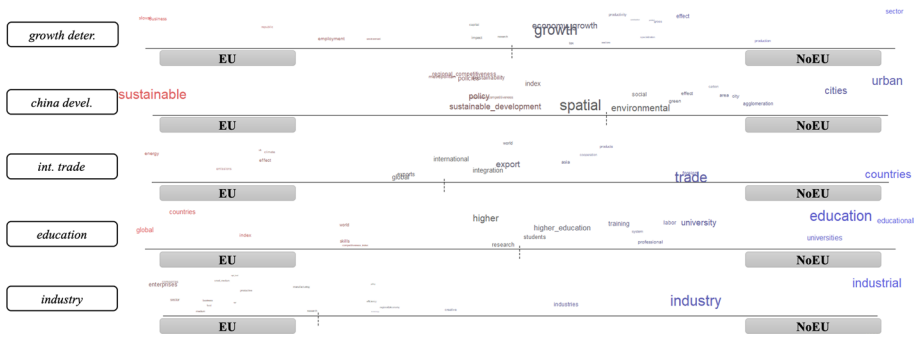


Fig. 7 Topical content representation (topic rank: 6–11)

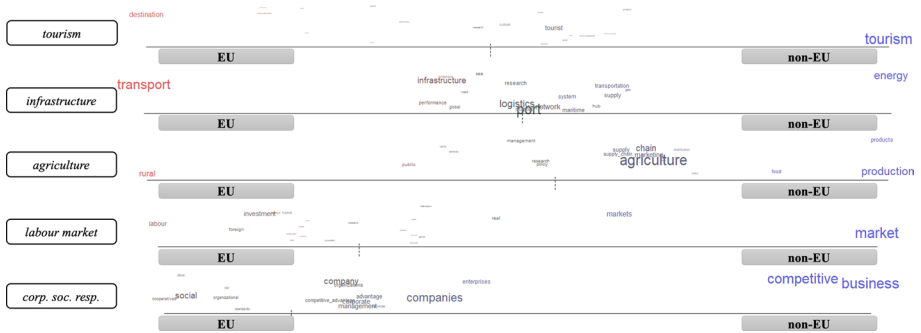


Fig. 8 Topical content representation (topic rank: 12–16)

- *Macroeconomics* there is a dichotomy between a territorial development with reference to the industrial sector (EU) and an economic development based on innovation and training (non-EU);
- *Institutions* authors from EU focused on European institutions, using terms referred to cohesion policies, whereas non-EU authors focused on financial institutions, speaking about efficiency and trends;

- *Technological development* even if it did not show a great impact of the covariate—highlighted by the label sizes—the vocabulary moved between research (EU) and technological innovation (non-EU);
- *Business* there is a differentiation between a business focused on the competitiveness of brands and products (EU) and competitiveness based on performances (Non-EU);
- *Urban development* the topic vocabulary opposed references to local governance (EU) and centralised governance (non-EU);
- *Growth determinant* the topic vocabulary opposed employment and investment on the EU side and growth and productivity on the non-EU side;
- *China development* the attention of authors was focused on sustainable development (EU) and urban development (non-EU);
- *International trade* the topic vocabulary opposed references to the energy transition and the related emissions on a global scale (EU) and trade and exports in an international context (non-EU);
- *Education* there was a variation between the internationalisation of the educational offer (EU) and the professionalisation (non-EU).
- *Industry* there was a focus on the EU side towards small-medium manufacturing, opposed to a more general discourse on the non-EU side;
- *Tourism* the topic did not show great lexical variability, identifying an urban tourism on the EU side and a rural tourism on the non-EU side;
- *Infrastructure* the topic focused for both categories on a logistic dimension, positioned at the centre of the graph, with an opposition between transport (EU) and energy (non-EU);
- *Agriculture* European studies still considered a rural dimension, with emerging references to management and services, while non-European studies focused on production, supply chain and marketing;
- *Labour Market* studies focusing on labour market issue opposed job conditions and investments (EU) and a general discourse on markets (non-EU);
- *Corporate Social Responsibility* for this topic, there were references to the social responsibilities of organisations (EU) versus references to the competitiveness of companies and businesses (non-EU).

5.3 RO3: Convergence Between Topics and Pillars

As regards the third research objective, we evaluated the connections (and the gaps) between the theoretical background (expressed by the different issues discussed in the literature) and the practical aspects (referred to as the conceptual framework underlying RC). As stated above, we considered as a reference the European RCI, consisting of 11 pillars. The diverse determinants of competitiveness covered by these pillars substantially overlap the determinants of competitiveness considered in other indexes as, for example, the Global Competitiveness Index (GCI) developed by the World Economic Forum (Schwab, 2019). Since GCI is used to measure competitiveness at a national level, the use of RCI pillars—conceived properly to measure competitiveness at a regional level—seemed more coherent with the objectives of this study.

Following the textual-based standpoint followed in the previous analysis, we decided to compare the lexicon containing the per-topic top words derived from the STM with a lexicon built from the pillar descriptions derived from the EU official documents and technical

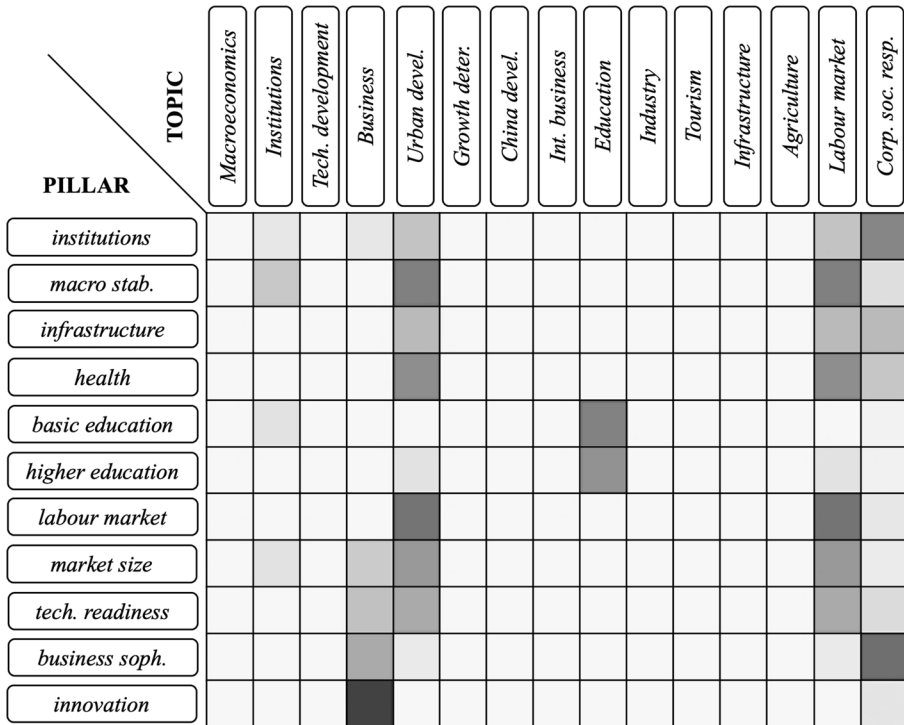


Fig. 9 Heat matrix of topic-pillar similarities

reports on RC⁵. These texts have been pre-treated with the same procedure described in Sect. 4.

We built a 667×15 binary matrix for the emergent topics (not considering the topic *Literature review*) and a 232×11 binary matrix for the pillars, then we computed the so-called cosine similarity (e.g., Iezzi, 2012; Misuraca et al., 2019) for each couple of vectors, to compare topics τ_i with pillars π_j :

$$\cos(\tau_i, \pi_j) = \frac{\tau_i \cdot \pi_j}{\|\tau_i\| \cdot \|\pi_j\|} \quad (6)$$

An important property of the cosine is its independence of text length. Values are bounded in a $[0, 1]$ range, where 1 means two texts use the same terms and 0 means two texts have no terms in common. The cosine values obtained from the analysis of topics and pillars are represented as a heat matrix in Fig. 9: a darker colour denotes a higher overlap between topics and pillars while a lighter colour denotes a lower overlap. The search for similarity between the extracted topics and the RC pillars, functional for deriving a convergence classification that can integrate the dimensions of competitiveness with the issues discussed in the literature, showed only a partial correspondence.

We observed that some topics are not covered by the pillars, whereas some pillars seems to have a cross-cutting lexicon that is reflected throughout different extracted topics. As an

⁵ See https://ec.europa.eu/regional_policy/en/.

example, the topic *Tourism* did not show similarity with any of the pillars since possible related terms were not enough expressed in the official documents. Similarly, there were no terms linked explicitly to *Industry* or *Agriculture*, and to the *Technological development*, which refers to the use of ICT technologies as a strategic lever for competitiveness. This aspect is interesting because it points out how the digital transformation could be considered from a diverse viewpoint between policymakers and scholars. The most evident overlap appeared for the *Education* topic, which was transverse to the two pillars related to primary and secondary education (*basic education* in the table, for the sake of simplicity) and the *higher education* (and lifelong learning). In a similar fashion, the *Business* topic is strictly connected with the pillars related to the *market size* and to the *technological readiness*, the *business sophistication* and the *innovation* (i.e., the ‘innovation’ sub-index defined by D’Urso et al. 2019). On the other hand, we observed some topics linked to several pillars. The topics *Urban development* and *Labour market* seemed to cover different issues encompassed in the pillars, revealing a central role in the debate about RC. Similarly, we observed a link between the topic *Corporate Social Responsibility* and different pillars, particularly with the pillars related to *institutions* and *business sophistication*, showing the importance of sustainability both in a public and in a private perspective.

6 Conclusion and Final Remarks

Over the last few years, the extension of the concept of competitiveness to territories, and regions more particularly, has driven the debate of scholars interested in this research area, with different opposing positions concerning the possibility of focusing on the meta-level offered by RC rather than the more traditional layers investigated in competitiveness studies, focusing on a whole economy (macro-level), industry (meso-level) or firm level (micro-level). Moreover, scholars persuaded of the validity of RC concept have deepened its determinants, taking into account both an economic and a social perspective. RC deals with a broad set of factors challenging to classify because of the multifaceted nature of the concept and the different standpoints induced by the diverse scientific areas active in this research domain. In the literature produced in the last years, the role of some RC determinants have been gradually contained (e.g. productivity), while other determinants have become more critical (e.g., education, social change, technology).

The growing attention to regions as primary sites of economic growth and wealth creation pushed RC to become a primary focal point for developing public policies. Policymakers have been less sensitive to the theoretical disagreements of scholars, taking advantage of RC as a new ideal ground of comparison among territories on a narrower scale with respect to the whole country. Some composite indicators—like the Global Competitiveness Index developed by the World Economic Forum—were introduced and increased their popularity, considering a national or a supra-national scale. According to Kitson et al. (2004), it is hazardous to transfer a concept of competitiveness initially defined for a national level to a sub-national level. On the other hand, RC cannot be seen as a spatial disaggregation of national competitiveness (Aiginger & Firgo, 2015). To measure RC directly, the EU developed a composite indicator known as the Regional Competitiveness Index, able to express synthetically the different dimensions of competitiveness at a sub-national scale and establish a cross-sectional and longitudinal perspective.

In this paper, we aimed to explore the most recent scientific production concerning RC, referring to 2016–2020, to highlight the topics that mainly dominate this research area,

taking into account the influence of some covariates. In particular, we analysed how the publication year and the publication type impacted the topic distribution. At the same time, we analysed how the geographical localisation of scholars (using the corresponding author as a proxy) impacted the topic content, looking at the differences in the vocabulary. Finally, we jointly mapped the topics which scholars dealt with and the topics covered by the pillars used to construct the RCI, to evaluate the convergence between scholars and policymakers.

The implemented topic extraction strategy made possible to identify the topics that characterised the scientific production about RC in the period under examination. The topic model inferred a set of linguistic patterns that would be hard to identify through a qualitative study given the size of the analysed collection. Other competing approaches could be considered in an unsupervised perspective (Misuraca & Spano, 2020), but the topic model is becoming more and more popular also in a bibliometric framework (e.g., Asmusen & Møller, 2019; Bohr & Dunlap, 2018; Chen & Xie, 2020), showing that the use of such an approach will increase in the future. The main advantage is that the topic model allows considering a mixed-membership, producing overlapping word clusters and estimating topic proportion in a document rather than attributing a single label. Furthermore, the topic model used in the study—known as structural topic model—allows estimating topical prevalence and topical content with respect to document metadata, studying how covariates impacted the presence of themes and the corresponding vocabulary.

RC literature rapidly changed during the years. The knowledge base resulting from the analysis highlighted three main sets of topics discussed by the scholars. These topics are related to the environment of competitiveness at a regional level (*Macroeconomics, Institutions, Infrastructure, Urban development*), to the economic ecosystem of regions (*Business, Industry, Tourism, Agriculture, Labour market, Technological development, Corporate Social Responsibility*), to the human capital (*Education*). This set of topics seems to confirm the framework defined by Martin (2005). Besides these themes, the topic concerning the RC in South-East Asia also emerged from the analysis (*China development, International trade*), manifesting the attention devoted to this concept also in economic systems very different from the EU countries, with particular regards to the export of goods produced in the local territories as a key asset of development. Particularly noteworthy is the absence of a specific topic concerning health. This asset is considered a major one for regional development and competitiveness, able to reduce economic and social disparities and enhance the wealth of people living in the different territories. On the other hand, the presence of CSR confirms the results discussed by Boulouta and Pitelis (2014): social responsibility can be a powerful driver of competitiveness, especially in a context characterised by weak innovations.

A focus on the language used by EU and non-EU scholars highlighted the conceptual divergences underpinning the RC paradigm and the different cultural backgrounds of the scholars. The geographical dichotomy characterising each topic suggested two parallel developments of RC issues in the literature. For the non-European authors, RC seems more related to investments, innovation and performances and less to a social dimension. Non-EU authors, therefore, seemed to be less focused on the sustainability concept and more focused on adopting an economic approach. However, this vision is surely affected by the lack of a supra-national layer like the EU. For the EU authors, the developed RC discourse recalls the evolution of the triple helix approach (government-industry-academia) addressed by several authors (e.g., Leydesdorff & Etzkowitz, 2003). Etzkowitz and Zhou (2006) proposed a new vision of the classic triple helix model pairing innovation with sustainability. In general, EU authors seemed to be more oriented towards a social and cohesive conception of RC, with the emergence of themes such as knowledge, culture

Table 3 Proposed topic classification

Geographical level	Sustainability		
	Government	Industry	Academia
Central	Institutions, Infrastructure, Macroeconomics	Growth determinants, Labour market Technological development, CSR	Education
Local	Urban development	Agriculture, Industry, Tourism, Business, CSR	Education

and sustainability. This is quite consistent with the idea of Europe as a knowledge-based society and a knowledge-based economy (e.g., Archibugi & Coco, 2005; Carayannis et al., 2012). This new perspective considers in which way the civic engagement affects the model, introducing a sustainable (entrepreneurial) university, a sustainable industry and a sustainable government (Cai & Etzkowitz, 2020; Cai & Lattu, 2019). These considerations led us to propose an organisation of the emerging knowledge base with respect to the different geographical levels and the different actors involved in the territorial competitiveness (Table 3).

The proposed systematisation takes into account the dichotomy between regional and supra-regional areas, highlighting the role of local districts as drivers for competitiveness. (Camagni & Capello, 2010) conceptualised this aspect through the introduction of a territorial capital in which the use of digital resources can act as a further key element of development.

The comparison between the topics obtained through the STM and the pillars used to build the RCI showed a limited overlap. The dissimilarities between the issues debated by scholars at a theoretical level and the factors used to operatively measure the RC may reflect a diverse perspective towards RC. Scholars have often pointed out differing aspects of competitiveness at a regional level, favouring an economic interpretation or rather a socio-economic interpretation. Moreover, as stated above, geographical localisation may induce scholars' vocabulary in portraying a given issue because of the context and the cultural background. On the other hand, the operationalisation of RC through the pillars, overcoming the contrasting positions of scholars on the validity of an RC concept, relies on the necessity of considering assets that can be directed observed on the regions, putting aside other significant aspects that can enhance the competitiveness of some places with respect to others, like the sustainability and the social responsibility. This attempt of connecting two perspectives, more academic or more governmental, deserves a more profound reflection that will be one of the future development of this study.

It is necessary to underline that the present survey has some limitations, mainly related (but not limited) to the bibliometric approach itself. Typically, bibliometric analyses may present false positive and false negative results due to the difficulty of establishing a precise and utterly inclusive query. Moreover, it is impossible to refer to a complete collection, since each of the most common indexing databases has strengths and weaknesses. We extracted publications only from WoS, covering only the portion of literature covered by this database. Many other publications might have been published in not-yet-indexed resources, therefore unable to be retrieved. Additionally, the research only included

publications where the terms in the query were in the title or the abstract or the keywords, but not inside the full text because the complete content of a publication is not available in the form of bibliographic records. Considering all these aspects, the publications analysed in the study might not precisely reflect the entire research activity on RC in the last five years, but the results suggested potentially interesting insights into the topics debated by scholars and highlighted the future frontiers of the domain. Further development of this research will involve both the methodological side and the specific domain under investigation. The use of topic modelling with other approaches for topic extraction will be considered in a joint strategy, leveraging the different viewpoints on the knowledge domain and trying to obtain a consensus between the diverse topic sets emerging from the analyses. At the same time, more in-depth analyses on specific issues of RC will be performed to improve the knowledge on themes currently debated in the reference literature, like for example the role of sustainability and innovation.

Funding Open access funding provided by Università della Calabria within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aiginger, K. (2006). Competitiveness: From a dangerous obsession to a welfare creating ability with positive externalities. *Journal of Industry, Competition and Trade*, 6(2), 161–177.
- Aiginger, K., & Firgo, M. (2015). Regional competitiveness under new perspectives. Policy Paper no. 26, WWWforEurope, <https://doi.org/10.2139/ssrn.2685585>.
- Aiginger, K., & Vogel, J. (2015). Competitiveness: From a misleading concept to a strategy supporting beyond GDP goals. *Competitiveness Review*, 25(5), 497–523.
- Airoldi, E. M., Cohen, W. W., & Fienberg, S. E. (2004). Bayesian models for frequent terms in text. SCS Technical Report CMU-CALD-106, Canegie Mellon University, <http://reports-archive.adm.cs.cmu.edu/anon/cald/CMU-CALD-04-106.pdf>
- Andreoni, V., & Miola, A. (2016). Competitiveness and sustainable development goals. Technical report eur 28136, European Union - Joint Research Centre, <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC103576/lb-na-28316-en-n.pdf>
- Annoni, P., & Dijkstra, L. (2013). EU regional competitiveness index RCI 2013. JRC scientific and policy report, European Union - Joint Research Centre, http://ec.europa.eu/regional_policy/sources/docgen/studies/pdf/6th_report/rci_2013_report_final.pdf
- Annoni, P., & Weziak-Bialowolska, D. (2016). A measure to target antipoverty policies in the European Union regions. *Applied Research in Quality of Life*, 11, 181–207.
- Archibugi, D., & Coco, A. (2005). Is Europe becoming the most dynamic knowledge economy in the world? *Journal of Common Market Studies*, 43(3), 433–459.
- Aria, M., Misuraca, M., & Spano, M. (2020). Mapping the evolution of social research and data science on 30 years of social indicators research. *Social Indicators Research*, 149(3), 803–831.
- Arun, R., Suresh, V., Madhavan, C. V., & Murthy, M. N. (2010). On finding the natural number of topics with latent dirichlet allocation: Some observations. In M. J. Zaki, J. X. Yu, B. Ravindran, V. Pudi (Eds.) *PAKDD '10: 14th pacific-asia conference on knowledge discovery and data mining* (Vol. Part I, pp. 391–402). Springer.

- Ashraf, Q., & Galor, O. (2013). The out of Africa hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, 103(1), 1–46.
- Asmussen, C. B., & Møller, C. (2019). Smart literature review: A practical topic modelling approach to exploratory literature review. *Journal of Big Data*, 6(1), 93. <https://doi.org/10.1186/s40537-019-0255-7>
- Asuncion, A., Welling, M., Smyth, P., & Teh, Y. W. (2009). On smoothing and inference for topic models. In *UAI '09: Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence* (pp. 27–34). AUAI Press.
- Audretsch, D. B., Hülsbeck, M., & Lehmann, E. E. (2012). Regional competitiveness, university spillovers, and entrepreneurial activity. *Small Business Economics*, 39(3), 587–601.
- Barro, R. J. (1989). The neoclassical approach to fiscal policy. In R. J. Barro (Ed.), *Modern business cycle theory*. Harvard University Press.
- Becker, G. S. (1993). Nobel lecture: The economic way of looking at behavior. *Journal of Political Economy*, 101(3), 385–409.
- Berelson, B. (1952). *Content analysis in communication research*. Free Press.
- Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, 1(1), 17–35.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Bohr, J., & Dunlap, R. (2018). Key topics in environmental sociology, 1990–2014: Results from a computational text analysis. *Environmental Sociology*, 4(2), 181–195.
- Boltho, A., Carlin, W., & Scaramozzino, P. (1997). Will East Germany become a new mezzogiorno? *Journal of Comparative Economics*, 24(3), 241–264.
- Börner, K., Chen, C., & Boyack, K. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37, 179–255.
- Boschma, R. (2004). Competitiveness of regions from an evolutionary perspective. *Regional Studies*, 38(9), 1001–1014.
- Boulouta, I., & Pitelis, C. N. (2014). Who needs CSR? The impact of corporate social responsibility on national competitiveness. *Journal of Business Ethics*, 119(3), 349–364.
- Budd, L., & Hirmis, A. (2004). Conceptual framework for regional competitiveness. *Regional Studies*, 38(9), 1015–1028.
- Buntine, W. (2009). Estimating likelihoods for topic models. In Z. H. Zhou & T. Washio (Eds.), *First Asian conference on machine learning (ACML 2009)* (pp. 51–64). Springer.
- Butollo, F. (2015). Growing against the odds: Government agency and strategic recoupling as sources of competitiveness in the garment industry of the Pearl River Delta. *Cambridge Journal of Regions, Economy and Society*, 8(3), 521–536.
- Cai, Y., & Etkowitz, H. (2020). Theorizing the triple helix model: Past, present, and future. *Triple Helix* 1–38. <https://doi.org/10.1163/21971927-bja10003>
- Cai, Y., & Lattu, A. (2019). Civically grounded triple helix: Synergies between triple helix and quadruple helix models of innovation. In *Proceedings of the xvii international triple helix conference* (pp. 70–71).
- Camagni, R. (2002). On the concept of territorial competitiveness: Sound or misleading? *Urban Studies*, 39(13), 2395–2411.
- Camagni, R., & Capello, R. (2010). Macroeconomic and territorial policies for regional competitiveness: an EU perspective. *Regional Science Policy & Practice*, 2(1), 1–19.
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7–9), 1775–1781.
- Capello, R., Caragliu, A., & Nijkamp, P. (2009). *Territorial capital and regional growth: Increasing returns in cognitive knowledge use*. Discussion paper no. 09-059/3, Tinbergen Institute.
- Carayannis, E. G., Barth, T., & Campbell, D. F. J. (2012). The Quintuple Helix innovation model: Global warming as a challenge and driver for innovation. *Journal of Innovation and Entrepreneurship*, 1(2), 1–12. <https://doi.org/10.1186/2192-5372-1-2>
- Cellini, R., & Soci, A. (2002). Pop competitiveness. *PSL Quarterly Review*, 55(220), 71–101.
- Chang, J., & Blei, D. (2009). Relational topic models for document networks. In D. van Dyk, M. Welling (Eds.) *Proceedings of the twelfth international conference on artificial intelligence and statistics, PMLR* (Vol. 5, pp. 81–88).
- Chen, X., & Xie, H. (2020). A structural topic modeling-based bibliometric study of sentiment analysis literature. *Cognitive Computation*, 12, 1097–1129.
- Cristelli, M., Tacchella, A., & Pietronero, L. (2015). The heterogeneous dynamics of economic complexity. *PLoS ONE*, 10(2), e0117174. <https://doi.org/10.1371/journal.pone.0117174>

- Cuccurullo, C., Aria, M., & Sarto, F. (2016). Foundations and trends in performance management. A twenty-five years bibliometric analysis in business and public administration domains. *Scientometrics*, *108*(2), 595–611.
- Cui, L., Fan, D., Li, Y., & Choi, Y. (2020). Regional competitiveness for attracting and retaining foreign direct investment: A configurational analysis of Chinese provinces. *Regional Studies*, *54*(5), 692–703.
- Deveaud, R., Sanjuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document Numérique*, *17*(1), 61–84.
- de Winter, J. C. F., Zadpoor, A. A., & Dodou, D. (2014). The expansion of Google scholar versus web of science: A longitudinal study. *Scientometrics*, *98*(2), 1547–1565.
- Diamond, J. M. (1997). *Guns, germs, and steel: The fates of human societies*. Norton.
- Dijkstra, L., Annoni, P., & Kozovska, K. (2011). *A new regional competitiveness index: Theory, methods and findings*. Working paper n. 2/2011, European commission - directorate-general for regional policy. https://ec.europa.eu/regional_policy/sources/docgener/work/2011_02_competitiveness.pdf
- D'Urso, P., De Giovanni, L., Massari, R., & Sica, F. G. M. (2019). Cross sectional and longitudinal fuzzy clustering of the NUTS and positioning of the Italian regions with respect to the regional competitiveness index (RCI) indicators with contiguity constraints. *Social Indicators Research*, *146*, 609–650.
- Erosheva, E., Fienberg, S., & Lafferty, J. (2004). Mixed-membership models of scientific publications. *Proceedings of the National Academy of Sciences*, *97*(22), 11885–11892.
- Esser, K., Hillebrand, W., Messner, D., & Meyer-Stamer, J. (1996). *Systemic competitiveness: New governance patterns for industrial development*. London: Routledge.
- Etzkowitz, H. (2003). Innovation in innovation: The triple helix of university-industry-government relations. *Social Science Information*, *42*(3), 293–337.
- Etzkowitz, H., & Zhou, C. (2006). Triple Helix twins: Innovation and sustainability. *Science and Public Policy*, *33*(1), 77–83.
- Foray, D. (2015). *Smart specialisation: Opportunities and challenges for regional innovation policy*. Routledge.
- Garcia-Alvarez-Coque, J. M., Roig-Tierno, N., Sanchez-Garcia, M., & Mas-Verdu, F. (2020). Knowledge drivers, business collaboration and competitiveness in rural and urban regions. *Social Indicators Research*. <https://doi.org/10.1007/s11205-020-02478-6>.
- González Catalán, S. (2021). Regional competitiveness in latin America: A comparative study of the key elements for regional performance. *Investigaciones Regionales - Journal of Regional Research*, *2*(50), 125–146.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, *101*(S1), 5228–5235.
- Grimmer, J. (2010). A bayesian hierarchical topic model for political texts: Measuring expressed agendas in Senate press releases. *Political Analysis*, *18*(1), 1–35.
- Hämäläinen, K., & Böckerman, P. (2004). Regional labor market dynamics, housing, and migration. *Journal of Regional Science*, *44*(3), 543–568.
- Harzing, A. W., & Alakangas, S. (2016). Google scholar, scopus and the web of science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*(2), 787–804.
- He, Q. (1999). Knowledge discovery through co-word analysis. *Library Trends*, *48*(1), 133–159.
- Head, A., & Lloyd-Ellis, H. (2012). Housing liquidity, mobility, and the labour market. *The Review of Economic Studies*, *79*(4), 1559–1589.
- Hoffman, M. D., Blei, D. M., Wang, C., & Paisley, J. (2013). Stochastic variational inference. *Journal of Machine Learning Research*, *14*(5), 1303–1347.
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288.
- Huggins, R., & Williams, N. (2011). Entrepreneurship and regional competitiveness: The role and progression of policy. *Entrepreneurship & Regional Development*, *23*(9–10), 907–932.
- Iezzi, D. (2012). Centrality measures for text clustering. *Communications in Statistics - Theory and Methods*, *41*(16–17), 3179–3197.
- Januškaite, V., & Užiene, L. (2018). Intellectual capital as a factor of sustainable regional competitiveness. *Sustainability*, *10*(12), 4848.
- Kitson, M., Martin, R., & Tyler, P. (2004). Regional competitiveness: An elusive yet key concept? *Regional studies*, *38*(9), 991–999.
- Kolbe, R. H., & Burnett, M. S. (1991). Content-analysis research: An examination of applications with directives for improving research reliability and objectivity. *Journal of Consumer Research*, *18*(2), 243–250.
- Komarova, V. N., Zjablova, O. V., & Denmukhametov, R. R. (2014). An infrastructure factor in regional competitiveness. *Mediterranean Journal of Social Sciences*, *5*(18), 355–360. <https://doi.org/10.5901/mjss.2014.v5n18p355>

- Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. SAGE.
- Krugman, P. R. (1994). Competitiveness: A dangerous obsession. *Foreign Affairs*, 73(2), 28–44.
- Krugman, P. R. (1996). Making sense of the competitiveness debate. *Oxford Review of Economic Policy*, 12(3), 17–25.
- Lall, S. (2001). Competitiveness indices and developing countries: An economic evaluation of the global competitiveness report. *World Development*, 29, 1501–1525.
- Lavrinenko, O., Ignatjeva, S., Ohotina, A., Rybalkin, O., & Lazdans, D. (2019). The role of green economy in sustainable development (Case study: the EU states). *Entrepreneurship and Sustainability Issues*, 6(3), 1113–1126.
- Lebart, L. (2020). Looking for topics: A brief review. In D. F. Iezzi, D. Mayaffre, & M. Misuraca (Eds.), *Text analytics* (pp. 215–223). Springer.
- Lee, M., & Mimno, D. (2014). Low-dimensional embeddings for interpretable anchor-based topic inference. In *Proceedings of the 2014 conference on empirical methods in natural language processing, ACL* (pp. 1319–1328).
- Leydesdorff, L., & Etzkowitz, H. (2003). Can the public be considered as a fourth helix in university-industry-government relations? Report on the fourth triple helix conference. *Science and Public Policy*, 30(1), 55–61.
- Leydesdorff, L., & Meyer, M. (2012). The triple helix of university-industry-government relations. *Scientometrics*, 58, 191–203.
- Lucas, R. E., Jr. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1), 3–42.
- Malecki, E. J. (2004). Jockeying for position: What it means and why it matters to regional development policy when places compete. *Regional Studies*, 38(9), 1101–1120.
- Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107(2), 407–437.
- Martin, R. (2005). *Thinking about regional competitiveness: Critical issues*. Tech. rep., East Midlands Regional Development Agency. <https://core.ac.uk/reader/30624618>
- Martin, R., Kitson, M., & Tyler, P. (Eds.). (2012). *Regional competitiveness*. Routledge.
- Mayring, P. (2004). Qualitative content analysis. In U. Flick, E. von Kardoff, & I. Steinke (Eds.), *A companion to qualitative research* (pp. 266–269). SAGE.
- McCluskey, D. N. (2010). *Bourgeois dignity: Why economics can't explain the modern world*. University of Chicago Press.
- McDonald, S., Robinson, S., & Thierfelder, K. (2008). Asian growth and trade poles: India, China, and East and Southeast Asia. *World Development*, 36(2), 210–234.
- Meyer-Stamer, J. (2008). *Systemic competitiveness revisited: Conclusions for technical assistance in private sector development*. Working Paper n. 14, Mesopartner. https://www.mesopartner.com/fileadmin/media_center/Working_papers/mp-wp14_01.pdf
- Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 conference on empirical methods in natural language processing, ACL* (pp. 262–272).
- Misuraca, M., & Spano, M. (2020). Unsupervised analytic strategies to explore large document collections. In D. F. Iezzi, D. Mayaffre, & M. Misuraca (Eds.), *Text analytics* (pp. 17–28). Springer.
- Misuraca, M., Spano, M., & Balbi, S. (2019). BMS: An improved dunn index for document clustering validation. *Communications in statistics - theory and methods*, 48(20), 5036–5049.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>.
- Moon, H. C., & Peery, N. (1995). Competitiveness of product, firm, industry, and nation in a global business. *Competitiveness Review*, 5(1), 37–43.
- Nijkamp, P. (2017). Regional development and urban dynamics. In R. Capello (Ed.), *Seminal studies in regional and urban economics* (pp. 305–312). Springer.
- OECD. (2005). *Building competitive regions: Strategies and governance*. OECD Publishing. <https://doi.org/10.1787/9789264009479-en>.
- OECD. (2007). *Higher education and regions: Globally competitive, locally engaged* (pp. 143–164). USA: OECD Publishing. <https://doi.org/10.1787/9789264034150-8-en>.
- OECD. (2009). *Regions matter: Economic recovery. Innovation and sustainable growth*. OECD Publishing. <https://doi.org/10.1787/9789264076525-en>.
- OECD. (2012). *Promoting growth in all regions*. OECD Publishing. <https://doi.org/10.1787/9789264174634-en>.

- Pietrzak, M. B., Balcerzak, A. P., Gajdos, A., & Arendt, Ł. (2017). Entrepreneurial environment at regional level: The case of Polish path towards sustainable socio-economic development. *Entrepreneurship and Sustainability Issues*, 5(2), 190–203.
- Polenske, K. (2004). Competition, collaboration and cooperation: An uneasy triangle in networks of firms and regions. *Regional studies*, 38(9), 1029–1043.
- Porter, M. E. (1990). *The competitive advantage of nations*. Free Press.
- Porter, M. E. (1998). Clusters and the new economics of competition. *Harvard Business Review*, 76(6), 77–90.
- Porter, M. E. (2001). Strategy and the internet. *Harvard Business Review*, 79(3), 62–78.
- Porter, M. E., & Ketels, C. H. (2003). *UK competitiveness: Moving to the next stage*. DTI economics paper no. 3, Department of Trade & Industry.
- Porter, M. E., & Kramer, M. R. (2006). Strategy & society. The link between competitive advantage and corporate social responsibility. *Harvard Business Review*, 84(12), 78–92.
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064–1082.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R package for structural topic models. *Journal of Statistical Software*, 91(1), 1–40.
- Sagiyeva, R., Zhuparova, A., Ruzanov, R., Doszhan, R., & Askerov, A. (2018). Intellectual input of development by knowledge-based economy: Problems of measuring in countries with developing markets. *Entrepreneurship and Sustainability Issues*, 6(2), 711–728.
- Schwab, K. (2019). *The global competitiveness report 2019*. World Economic Forum.
- Schwab, K., & Porter, M. (2008). *The global competitiveness report 2008–2009*. World Economic Forum.
- Shvindina, H. (2020). Economic competitiveness: An overview of multilevel concept. In T. Wall, W. Leal Filho, A. M. Azul, L. Brandli, P. G. özuyar (Eds.) (pp. 160–172). Springer.
- Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70(1), 65–94.
- Taddy, M. (2012). On estimation and selection for topic models. In: N. D. Lawrence, M. Girolami (Eds.) *Proceedings of the fifteenth international conference on artificial intelligence and statistics*, PMLR (Vol. 22, pp. 1184–1193).
- Teixeira, S. J., & De Matos Ferreira, J. J. (2018). A bibliometric study of regional competitiveness and tourism innovation. *International Journal of Tourism Policy*, 8(3), 214–243. <https://doi.org/10.1504/IJTP.2018.094483>
- Teixeira, S. J., & Pocinho, M. (2020). Hotel industry and regional competitiveness: The bibliometric perspective of web of science. *Journal of Spatial and Organizational Dynamics VII*, 1(2), 129–147.
- Thyroff, A., & Kilbourne, W. E. (2018). Self-enhancement and individual competitiveness as mediators in the materialism/consumer satisfaction relationship. *Journal of Business Research*, 92, 189–196.
- Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. In *ICML '09: Proceedings of the 26th annual international conference on machine learning*, ACM (pp. 1105–1112).
- Wang, C., & Blei, D. M. (2011). Collaborative topic modeling for recommending scientific articles. In: *KDD '11: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining*, ACM (pp. 448–456).
- Wang, L., & Shen, J. (2017). Comparative analysis of urban competitiveness in the Yangtze River Delta and Pearl River Delta regions of China, 2000–2010. *Applied Spatial Analysis and Policy*, 10, 401–419.
- Werker, C., & Athreye, S. (2007). Marshall's disciples: Knowledge and innovation driving regional economic development and growth. *Journal of Evolutionary Economics*, 14(5), 505–523.
- Yeh, A. G. O., & Xu, J. (2008). Regional cooperation in the Pan-Pearl River Delta: A formulaic aspiration or a new imagination? *Built Environment*, 34(4), 408–426.
- Zeibote, Z., Volkova, T., & Todorov, K. (2019). The impact of globalization on regional development and competitiveness: Cases of selected regions. *Insights into Regional Development*, 1(1), 33–47. [https://doi.org/10.9770/IRD.2019.1.1\(3\)](https://doi.org/10.9770/IRD.2019.1.1(3)).