Technological diversity to address complex challenges: the contribution of American universities to sdgs

Ascione, Grazia Sveva

Università degli Studi di Torino

1 October 2023

Technological diversity to address complex challenges: the contribution of American universities to SDGs

Grazia Sveva Ascione[1,2]

[1] *Bordeaux School of Economics (BSE), Université de Bordeaux, UMR 6060, F-33600 Pessac, France*

[2] *Centre national de la recherche scientifique (CNRS)*

[*] *Corresponding author: Grazia Sveva Ascione, grazia.ascione@u-bordeaux.fr*

December 12, 2023

Abstract

Much of the literature emphasizes the relationship between interdisciplinarity and the Sustainable Development Goals (SDGs), which are seen as closely linked and highly interdisciplinary. Therefore, innovation related to the SDGs is expected to be technologically diverse, especially when it emanates from academia, where teams of researchers collaborate to create innovation for the benefit of society. However, research on innovation for the SDGs is still in its infancy due to a lack of comprehensive quantitative analysis about its characteristics and a lack of consideration of potentially relevant actors, such as universities. This paper aims to make a threefold contribution to the existing literature by analyzing USPTO patent data from 2006 to 2020. First, we develop a novel method for tagging SDGs-related patents using an unsupervised natural language processing (NLP) approach. Starting from an initial list of keywords, we build an extended dictionary of keywords for each SDG based on the patent text by combining the TF-IDF method with a vector representation of the patent text and SDGs keywords. Second, we analyze innovation related to the SDGs, focusing particularly on the contribution of universities. Third, we compare the diversity of SDGs and non-SDGs patents using the Rao-Stirling index. Our results show that patents related to the SDGs are on the rise, but the trend is more pronounced for universities, where the majority of innovation production revolves around SDG 3 (good health and well-being). Moreover, the rise in SDGs patents seems to be led not only by green technologies, but mainly by high technologies. Eventually, the empirical results point in two directions. On the one hand, SDGs related patents are more diverse than their counterparts across almost all technology sectors. However, if we consider university patents only, there is a diversity premium only for a few SDGs, namely SDG 2, SDG 3, and SDG 15.

Keywords: Technological diversity; University patents; Interdisciplinarity; SDGs; United States

JEL codes: C25; O3; Q56

1 Introduction

The Sustainable Development Goals (SDGs) aim to address the complex challenges of our century through 167 interlinked and interdisciplinary targets; progress on one goal depends on and influences other goals. For example, developments in agriculture towards zero hunger (SDG 2) depend on affordable and clean energy (SDG 7), while also reducing inequalities (SDG 5 and SDG 10) and protecting life on land (SDG 15) and in water (SDG 14). The use of technology and innovation is crucial to achieving the goals of the SDGs, as it fosters the knowledge economy and the resulting creation of inventions that can help decouple economic growth from the risk of environmental and social crises and improve living conditions in areas such as the environment, energy, medicine and transport (Blohmke, 2014; Walz et al., 2017). In this sense, technological development can unleash its potential for systemic change while having a positive impact on the environment and society (Deuten, 2003). The importance of science, technology and innovation to reach the SDGs has been stressed in the UN official documents where they are intended to boost the capacity of countries to change the current trajectory and accelerate progress toward a sustainable future (UN, 2020). However, there is no direct reference to IP in the goals and targets of the 2030 Agenda, with the exception of paragraph 3.b of Goal 3, nor any IP-related indicators in the current Global Indicator Framework. Therefore, in this research we contribute to reduce this gap, proposing a novel methodologies to map patents to the SDGs, adding evidence about the relationship between IP and SDGs.

Although the literature linking interdisciplinarity to the SDGs is extensive, less attention has been dedicated to the characteristics of innovation in the context of the SDGs (van der Waal et al., 2021; Hajikhani and Suominen, 2021) and recent studies do not consider potentially relevant stakeholders such as universities, which are expected to contribute to the achievement of the SDGs through a mix of education, research and innovation (Owens, 2017; Sánchez-Carracedo et al., 2021; Kopnina, 2020). Further, a popular view is that exploiting a single domain promotes "one way of thinking, damping creativity, while combining knowledge from diverse and distant domain leads to more breakthrough innovation." (Hargadon and Sutton, 1997; Ahuja and Morris Lampert, 2001). This has been proven to be the case for green technologies, which are characterized by intrinsic complexity and therefore result from the integration of different and heterogeneous technologies and knowledge sources (Quatraro and Scandura, 2019; Barbieri et al., 2020; Fusillo et al., 2020). Therefore, this research builds on the green innovation literature and explores the characteristics of SDGs-related innovation, which includes not only green innovation but also the so-called "blue" innovation which relates to unmet sustainable development needs, such as reducing poverty and hunger, promoting health and well-being, education, biodiversity, water and sanitation (van der Waal et al., 2021). In particular, we investigate if SDGs-related innovation is more diverse than its counterpart, as this information might add on the debate about the best policy intervention to foster the development this kind of technologies.

In addition, considering that universities and research centers around the world have made significant progress in establishing collaborative, interdisciplinary initiatives in sustainability science thanks to their more diverse knowledge and skills base (De Marchi, 2012; De Marchi and Grandinetti, 2013) , this research also examines whether universities are able to exploit their favorable position to produce more diverse innovation when it is linked to the SDGs.

Thus, the contribution of this paper to the literature is threefold. First, exploiting the textual part of patents (title, abstract and claims), we develop a novel methodology for tagging SDGs-related patents through an unsupervised natural language processing (NLP) approach; starting from a pre-validated list of keywords, we create a keywords' dictionary for each SDG based on patent text. To do that, we combine the TF-IDF (Term Frequency-Inverse Document Frequency) method with a vectorial representation of patent text. Thanks to the enriched vocabulary, we manage to better identify those patents that were missed in the initial matching due to the peculiarities of the legal jargon characterizing patents in general, but especially claims (Bonino et al., 2010; Tseng et al., 2007). This is, to our knowledge, one of the first attempts to create a proxy measure to analyse the progress towards the SDGs in the innovation system.

Second, we are among the first to analyse innovation related to each SDG, providing original descriptive evidence about that; then we empirically compare the diversity of American SDGs and non-SDGs patents across the main technological classes.

Third, we provide evidence about American universities SDGs patent portfolios composition and investigate whether and for which specific SDGs, there is a diversity premium.

Our results show that, overall, patents related to the SDGs are on the rise, but the trend is more pronounced for patents owned by universities, highlighting that universities are increasingly aware of their role in the sustainability journey; however, most of the production of university patents related to the SDGs seems to revolve around SDG 3 (good health and well-being). Overall, the rise of SDGs patents seem to be led not only by green technologies, but mostly by high technologies, a relationship which has been underestimated by the literature (Vinuesa et al., 2020; Kostoska and Kocarev, 2019). Furthermore, we prove that SDGs related patents are more diverse than non-SDGs patents across most of the main technological fields. Although the role of universities in providing an interdisciplinary perspective to the SDGs is highlighted in the literature, university patents only have a diversity premium for few of the SDGs (namely SDG 2, SDG 3, and SDG 15). This may suggest that interdisciplinarity is seen as a valuable resource in universities, but paradoxically it is more difficult to achieve in sustainable innovation.

The rest of the paper is organized as follows: Section 2 presents the theoretical background and the research hypotheses, Section 3 presents the Research Design, Section 4 shows the main results and Section 5 concludes.

2 Theoretical Background and hypotheses development

2.1 Intellectual property and the SDGs

Intellectual property (IP) is a critical incentive for innovation and creativity, which are key to achieving the Sustainable Development Goals (Walz et al., 2017; Cordova and Celone, 2019). Only through human ingenuity it is possible to develop new solutions not only to promote economic growth, but also to eradicate poverty, increase agricultural sustainability and ensure food security, combat disease, improve education and equality, protect the environment, and accelerate the transition to a low-carbon economy to combat climate change and preserve biodiversity (Rimmer, 2018). From this perspective, innovation and creativity are not goals in themselves; they are methods and tools for innovative solutions to development problems and, because they are at the core of the system, have an impact on a number of SDGs. As such, technologies are deemed to directly impact SDG 2 (zero hunger) (Blakeney, 2009; Oguamanam, 2006), SDG 3 (good health and well-being) (Abbott, 2002), SDG 6 (clean water and sanitation), SDG 8 (decent work and economic growth) and SDG 9 (industry, innovation and infrastructure) (WIPO, 2019). In addition, increasing the share of environmental oriented technologies is essential to achieve SDG 7 (affordable and clean energy), SDG 11 (sustainable cities and communities), SDG 12 (responsible production and consumption), SDG 13 (climate action), SDG 14 (life under below water) and SDG 15 (life on land) (Henry and Stiglitz, 2010; Rimmer, 2014). Further, the so-called "blue" technologies, namely those aiming at "improving conditions" (van der Waal et al., 2021) might help in achieving SDG 1 (no poverty) (Idris, 2003), SDG 4 (quality education), SDG 5 (gender equality) and SDG 10 (reduce inequalities) (WIPO, 2019). In a perfect world, a sustainable development agenda for IP would include a "universal call to action" to ensure that the IP system helps address the sustainability related issues (Bannerman, 2020). However, it should be noted that there is no direct reference to IP in the goals and targets of the 2030 Agenda, with the exception of paragraph 3.b of Goal 3, which mentions IP rights in relation to flexibilities to protect public health. In addition, there are no IP-related indicators in the current Global Indicator Framework adopted by the UN Statistical Commission, the UN Economic and Social Council, and the UN General Assembly in 2017. On the one hand, the World Intellectual Property Organization (WIPO) has considered relatively few of its activities to contribute directly to the SDGs, limiting its contribution to explicitly acknowledge the role of IP for SDG 9 and proposing an accurate classification of green technologies (the WIPO *Green Inventory*) that can be used to spot environmental related patents which might be consistent with some SDGs objectives such as those of SDG 6, SDG 7, SDG 13, SDG 14, and SDG 15 (Walz et al., 2017; Guo et al., 2020). Consistently, several scholars emphasize the role of green technologies in fulfilling the 2030 Agenda. For example, the study by Walz et al. (2017) examines the dynamics of green energy and resource efficiency innovation, looking at the position of northern and emerging economies. Instead, Guo et al. (2020) examine the characteristics of sustainable development in the context of green technology, using the indicators of the Sustainable Development Goals Index (SGDI) in its environmental component. On the other hand, researchers urge a more in-depth examination of innovation and the SDGs, particularly in relation to social issues. This call has been echoed by scholars, such as van der Waal et al. (2021) and Hajikhani and Suominen (2021), and practitioners; for instance, IP specialist consultancy *Lex Machina* recently published a study describing an additional feature of the Lexis Nexis database to map patents connected to the SDGs. They emphasize that mapping patents to the Sustainable Development Goals allows companies to objectively measure progress, understand their portfolio and that of their competitors, identify licensing and M&A targets, and assess risks and opportunities in the context of sustainable development. From a policy perspective, SDGs patent mapping could support strategic decision-making for sustainable investment, as well as identify gaps in sustainable technology development and the most effective innovations. This can lead to improved sustainability-related decision making and reporting, as well as influence R&D investments and support investment plans. Therefore, our first research contribution consists in an original attempt of mapping of patents related to each SDG, aiming to identify the key enabling technologies, therefore expanding the literature with new evidence about the relationship IP and the SDGs. The methodological section (Section 3.2) provides details about the methodology through which we accomplish the task.

2.2 Technological diversity and the SDGs

Many authors believe that inventions result from the combination of existing ideas and devices (Weitzman, 1998; Arthur, 2007). Recombinant inventions are frequently referred to be breakthrough because, in contrast to incremental innovation, merging information from various and distant disciplines nurtures creativity and promotes innovative ideas that are more likely to result in valuable inventions (Ahuja and Morris Lampert, 2001; Audia and Goncalo, 2007; Zhu et al., 2022). This type of innovation is consistent with that required to address sustainability-related challenges (Lam et al., 2014; Jones et al., 2010), where an interdisciplinary approach is required to adequately map multi-layered, complex issues, and without such an approach, necessary solutions risk not be identified. Complex systems, such as acid rain or rapid population expansion, are multifaceted, and standard disciplinary techniques are severely limited in their ability to provide a full view of such phenomena by examining them from the perspective of a single discipline. Newell et al. (2001) suggests that a complete understanding of the interrelationships and dynamics between the various components of these complex events can be achieved by drawing on and integrating multiple perspectives. To this end, it is necessary to combine the efforts of experts from multiple disciplines to address the complex socio-ecological problems of our time (Morse et al., 2007).

Interdisciplinarity has already been proven as an effective means of addressing complex challenges in the innovation system, such as those posed by climate change and environmental degradation. In particular, the link between diversified knowledge sources and green innovation has been analyzed through the framework of recombinant technologies and recombinant competences (Fusillo et al., 2020; Orsatti et al., 2020; Zeppini and van Den Bergh, 2011). For example, preliminary evidence from Fusillo et al. (2020) showed that green technologies have a higher degree of diversity than their non-green counterparts, reinforcing the idea that green technologies should be considered complex due to the different bodies of technologies they combine. Further, Popp and Newell (2012) find that patents in sustainable energy domains are cited by a variety of other technological domains.

Consistently, innovation related to the SDGs is expected to be interdisciplinary and overcome the lack of holistic vision that often characterizes individual disciplines (Annan-Diab and Molinari, 2017). Although van der Waal et al. (2021) confirms that green technologies are relevant for many SDGs, no study attempts to quantitatively measure the interdisciplinarity of SDGs-related innovations, neither those more environmentally or more socially related.

To this end, in this study we follow the framework proposed by Rafols and Meyer (2010) who define diversity as the differences in the body of integrated knowledge that can be summarized by three attributes:

1. Variety: the number of different categories in which an element can be classified;
2. Balance: the evenness of the distribution of elements among categories;
3. Disparity: the degree of diversity between these categories.

The Rao-Stirling (RS) diversity index was originally proposed by Rao (1982) and then revised by Stirling (2007) to consider all the above three elements simultaneously, and it is often considered in the analysis of research interdisciplinarity, although its validity has been discussed in recent studies such as Leydesdorff et al. (2019).

In view of this, we put forward the first hypothesis:

Hypothesis 1 (H1): *SDGs related patents are more diverse than non-SDGs related ones across the different technological fields.*

2.3 The role of American universities for SDGs

Universities are fundamental actors in the innovation ecosystem, which are expected to have a "public mission" which consists of providing knowledge, critical thinking, and technological advances to tackle society's fundamental problems (Winickoff, 2013; Papadimitriou, 2020). In particular, since the Bayh-Dole Act, American universities have been encouraged to pursue the so-called "third mission", maximizing the societal benefit of technology they produce (Lemley, 2007). Therefore, the third mission of contemporary universities includes fostering the achievement of the SDGs (Lozano et al., 2013; Ceulemans et al., 2015; Blasco et al., 2020). Education, research and innovation are the three areas where universities play a vital role in putting society on the path of sustainable development (Körfgen et al., 2018; Leal Filho et al., 2019). While there are extensive studies in the literature about implementation of SDGs into university curricula (Albareda-Tiana et al., 2018; Álvarez et al., 2021; Sánchez-Carracedo et al., 2021; Thomas, 2016), academic research about the SDGs is expected to use a transformative approach which brings together different fields of study and uses interdisciplinarity as a "competitive advantage" to address complex societal challenges (Leal Filho et al., 2019). To this end, universities and research centers around the world have made significant progress towards establishing collaborative, interdisciplinary initiatives in sustainability science (Hernandez-Aguilera et al., 2021). This idea is consistent with the momentum that interdisciplinary research (IDR) is having in universities, producing wide-ranging scientific advances and leading to the establishment of interdisciplinary research centers (Biancani et al., 2018). Although interdisciplinarity research of SDGs is a trending topic (El-Jardali et al., 2018; Kestin et al., 2017), quantitative contributions on this topic are still scarce.

Further, the role of scientific research by universities to achieve the SDGs is not limited to a "knowledge phase", where research is conducted to answer the question "what is at present", but it includes a "technological phase" as well, where universities develop technological solutions to solve specific SDGs challenges (Kestin et al., 2017). Therefore, the research function of universities is closely linked to the production of innovation. Recently, universities have been shown to be particularly important in the development of environmental innovations because they accumulate a wide range of expertise and competencies that are distributed across different organisations (De Marchi, 2012; De Marchi and Grandinetti, 2013). Because of their particular educational endowments, inventors within universities are thought to have diverse knowledge bases and skills that enable them to successfully recombine bits of knowledge from different technological fields domains (Quatraro and Scandura, 2019).

Based on this conceptual background, we intend to analyze the production of innovation of American universities related to the SDGs, to check whether the interdisciplinary environment that contemporary universities are deemed to foster has an impact on the technological diversity of university produced innovation; thus, we put forward our second hypothesis:

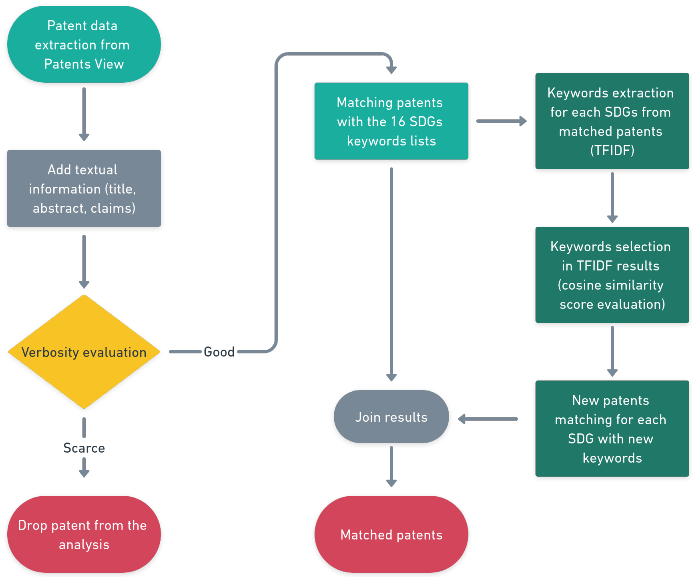
Hypothesis 2 (H2): *University patents related to each of the SDGs are more diverse compared to other university patents.*

3 Research Design

3.1 Patent data collection

We use as a source of data for this research the information in patents granted at USPTO, considering them as a viable proxy to study the domain of technological domains in the knowledge economy (Jaffe and Trajtenberg, 2002). In order to use patents filed at USPTO, we relied on Patents View (Version 2021) where we collected all patents granted from 2006 to 2020. The Patents View platform is built on a

Figure 1: Flowchart of SDGs related patent identification



Notes: The flowchart represents the research design of this work, emphasizing the four steps that compose the SDG tagging process: i) the first round of patent tagging using the original keywords lists, ii) the tf-idf keywords expansion, iii)the cosine-similarity based keywords selection iiii) the final round of patent tagging

regularly updated database that longitudinally links inventors, their organizations, locations, and overall patenting activity and reports the name(s) of first assignee(s), while the US Patent Assignment Database contains detailed information on patent assignments and other transactions recorded at the USPTO. The data collected from Patents View include general information about the patents (patent identification code, grant date), patent text (title, abstract and claims)[1], the IPC and WIPO classes associated with each patent. Information about patent quality is retrieved from OECD 2021 patent quality dataset.[2] Our analysis includes the specific subset of American universities' patents, defined as those having at least one university among the assignees in the patent history. After merging together the general information about patent text, the dataset has 85'169 patents including US universities among their applicants granted in the time range from 2006 to 2020.

3.2 Tagging SDGs related patents

We approach the research questions exploring SDGs-relevant innovations using patent data and a textual analysis of their content. To identify patents related to interdisciplinary or integrated technologies or

[1]Two tables were used to collect the textual part of the patents. The first table is "patent.tsv" where it is possible to find information about abstract and title of patents; then, the claims information was retrieved joining each year's claim table from Patents View to the patents.tsv table. The data is free and accessible at https://patentsview.org/download/data-download-tables

[2]The OECD patent quality dataset (version 2021) is available upon request at https://www.oecd.org/sti/inno/intellectual-property-statistics-and-analysis.htm

emerging products, keyword search can be used as an effective method (Xie and Miyazaki, 2013). Patent documents are divided into several elements, including title, abstract, claims, and description, and since their purposes differ, their sentence structures and vocabulary also differ from each other. First, the title and abstract use distinctive and significantly differentiated words to properly express the relevant technologies, but are short and lack specific details about them. On the other hand, the claims are more complete and explicitly describe the related technical features to ensure complete legal protection, which is essential for patents (Noh et al., 2015). Previous work by van der Waal et al. (2021) and Hajikhani and Suominen (2021) also used patent text to map patents to the SDGs. In particular, the former use direct keyword matching from an initial keyword list and then label the green and "blue" (socially oriented) patents relevant to the SDGs; the latter propose a supervised machine learning approach instead: first, they create a dictionary of SDG-relevant keywords from UN SDGs documents, then they use it to identify relevant publications in the SCOPUS database. Second, they preprocess the text and convert it into a TF-IDF matrix, to create word embeddings to be used for classification. After validating the model, their classifier is trained on labeled publication data to predict the vector of probabilities that a patent belongs to each of the SDGs. However, they are not able to evaluate the quality of their patent results because they lack a *ground truth* on patent data (they only have pre-labeled data on publications and not on patents). In line with the previously mentioned literature, we perform a text analysis of the title, abstract and claims of patents, which are considered by the authors to be the most appropriate parts of the text for performing a quantitative analysis and a keyword search to avoid type I errors (missing patents that should be identified) and type II errors (retrieving irrelevant patents) (Xie and Miyazaki, 2013).

3.2.1 First round of matching and TF-IDF

As reported in Figure 1, the first step for tagging SDGs related patents consists of a direct matching between a *corpus* generated from joining each patent title, abstract and claims and 16 lists of keywords related to the SDGs (one for each SDG expect for SDG 17 which focuses on strengthening the partnerships to reach the other goals, thus is omitted from this analysis) developed by the University of Auckland SDGs keywords mapping research project [3]. The choice of this keyword list is due to its completeness, considering it combines Elsevier's queries, a subset of Sustainable Development Solutions Network (SDSN) and UN queries. Moreover, the list was generated using a text mining approach on academic publications, hence we selected it also for its consistency with the methodology proposed in this research.[4] Through this first round of matching, we are able to tag 426'863 patents related to at least one SDG as those that have at least one keyword from the corresponding list in their text.[5] However, considering that the

[3] More information on the project is available at https://www.sdgmapping.auckland.ac.nz/

[4] In particular, the list was made applying an n-gram model to mine the abstracts of academic publications, in order to identify relevant sequences of words. Afterwards, n-gram tokens were then scored by a range of factors, including counts and measures of frequency, and were then ranked by those scores. Keywords with a high rank were then evaluated in more detail and manually reviewed to confirm that they were relevant to the Goal in question.

[5] Considering the overlapping among SDGs (Nilsson et al., 2016), some keywords might be repeated in different lists.

keywords lists are based on academic publication text as well the peculiarities of patent texts and their specific legal jargon (especially in the claims) and the technical words that are not common in everyday language (Bonino et al., 2010; Tseng et al., 2007), we decided to perform a keyword extraction procedure to avoid type I errors as much as possible. The criteria for selecting keywords from a document may also vary. For example, words that occur most frequently in certain documents may be considered critical, or words that fit well with the main topics of the document are often assumed to be important. In general, while words that occur frequently in patent documents are likely to be representative keywords, those that occur too frequently in such documents are also likely to be general words that occur in all documents rather than representative words that allow specific patents to be identified. Noh et al. (2015) sought to identify the most effective keyword strategy for text mining-based patent analysis by evaluating and verifying the most commonly used keyword selection and processing methods in existing studies. Their results highlight TF-IDF (Term Frequency-Inverse Document Frequency) as the best performer because it has the lowest entropy values. The term frequency (TF) is the number of times a term appears in a document and is calculated as follows:

$$tf_{ij} = \frac{n_{ij}}{|d_j|} \tag{1}$$

tf_{ij} stands for the number of occurrences of word i in document j and $|d_j|$ is the dimension, expressed by the number of words of j. Inverse Document Frequency (IDF) measures the rarity of a term in the whole corpus. It is calculated as follows:

$$idf_i = log_{10} \frac{|D|}{|d : i \epsilon d|} \tag{2}$$

where the denominator is the number of documents containing i. The concepts of term frequency and inverse document frequency are combined, to produce a composite weight for each term in each document, with the following formula:

$$tf - idf = tf_{ij} * idf_i \tag{3}$$

In this way, the TF-IDF method can retrieve important keywords that are closely related to a representative technology while avoiding general terms in the corpus (Usui et al., 2007). Thus, TF-IDF allows us to score each word in each patent document according to its weight both in the single patent text and in the whole collection of texts. For this analysis, we performed TF-IDF separately for each SDG, considering each SDG as a separate collection of texts.[6]

3.2.2 Keyword selection through cosine similarity and second round of matching

To evaluate the effectiveness of TF-IDF, we confronted the resulting extra keywords with the original lists. To make the comparison, we selected the 10 most relevant keywords for each patent according to

[6]In this research, Python *scikit-learn* library is used to carry out the tf-idf because it considers also bi-grams and tri-grams that are more common than monograms in the original SDGs lists.

their score and we compared them with the keywords in the original lists that had at least one match in any patent text. The results confirm the validity of TF-IDF: more than half of the keywords scored among the top 10 for each patent are also in the original SDGs lists, hence confirming the validity of this unsupervised method. Moreover, a further step of selection was needed to choose among the top 10 scored keywords per patent the most relevant ones to expand the original SDGs dictionaries. To do that, we took advantage of the vectorial representation of words elaborated by a pre-trained transformer based neural network (SentenceTransformers)[7], a kind of technique which has already demonstrated huge potentialities for patent analysis (Li et al., 2018; Chen et al., 2020; Roudsari et al., 2021). For each SDG, we compared the n-dimensional vector representing each keyword in the original list to the n-dimensional vector representing each word in the top 10 words per patent according to TF-IDF score. Among them, for each keyword in the original lists we selected the closest 3 in terms of cosine similarity.[8] Therefore, exploiting the vectorial representations of patents and their spatial and semantic closeness, as measured by cosine similarity, we are able to enlarge the original dictionaries with two different types of keywords:

1. Keywords which enrich concepts already present in the original list with semantically relevant keywords/synonyms (e.g., in the original list of SDG 9-Industry, Innovation and Infrastructure- there is the keyword "Sustainable Industrialization" whose closest keywords are "industrial waste" and "renewable", which are semantically close ideas; or, in SDG 3- Good Health and Wellbeing- the keyword "Sexual Health" has as closest keywords "sexual disorders" and "reproductive health").

2. Keywords which specify concepts and ideas already present in the original list with more technical wording (e.g., "Photochemistry" is a keyword present in the original list of SDG 7-Sustainable Development- and one of the closest identified keywords is "photocatalysis"; or in SDG 13-Climate Action- the bigram "greenhouse gas" has as closest keyword "CO2" which is the primary greenhouse gas emitted through human activities).[9]

Table 8 in the Appendix gives further detail about the results of the two rounds of matching. In particular, in the second round of matching, using the keywords derived from TF-IDF and selected through the cosine similarity method, we are able to add 1’784 keywords that produce a match in our patent record.[10] The total number of SDGs related patents identified is 693’571.[11]

[7]SentenceTransformers (`https://www.sbert.net/`) is a modification of the pretrained BERT network that use siamese and triplet network structures to derive semantically meaningful sentence embeddings that can be compared using cosine-similarity. The main advantage of the model is that it is optimized to calculate cosine similarity, while maintaining the accuracy from BERT. Further information on the architecture is available in the work of Reimers and Gurevych (2019).

[8]Cosine similarity measures the cosine of the angle between two vectors, whose values range between -1 and 1. It is widely used as technique to assess semantic similarity between two documents, including patents' text. For instance see recent work by Rogers (2020) and Feng (2020)

[9]For each SDG, the enlarged list of keywords is available upon request.

[10]Based on the frequencies of the matched terms we excluded 34 noisy keywords, derived from the TF-IDF. We decided, for each SDG, to exclude from the keywords those whose matching frequency was higher than the average frequency of all terms.

[11]As robustness check for this results, we compared the number of SDGs related patents to the number of green patents (as defined by two established international classifications, both based on the International Patent Classification

Table 1: Number of SDGs related patents and % by assignee type

SDG	all-assignees	% on tot.patents	%_all_SDG	univ-assignee	%_tot_univ	%_SDG univ
SDG1	66'998	1,8	9,7	1'093	1,3	4,3
SDG2	39'367	1,1	5,7	2'193	2,6	8,7
SDG3	101'129	2,8	14,6	13'486	15,8	53,4
SDG4	46'255	1,3	6,7	1'533	1,8	6,1
SDG5	135'821	3,7	19,6	3'712	4,4	14,7
SDG6	18'712	0,5	2,7	663	0,8	2,6
SDG7	85'761	2,4	12,4	2'801	3,3	11,1
SDG8	23'294	0,6	3,4	520	0,6	2,1
SDG9	26'271	0,7	3,8	552	0,6	2,2
SDG10	8'156	0,2	1,2	77	0,1	0,3
SDG11	235'204	6,5	33,9	2'387	2,8	9,5
SDG12	31'708	0,9	4,6	819	1	3,2
SDG13	12'426	0,3	1,8	255	0,3	1
SDG14	8'949	0,2	1,3	249	0,3	1
SDG15	10'211	0,3	1,5	358	0,4	1,4
SDG16	11'306	0,3	1,5	187	0,2	0,7

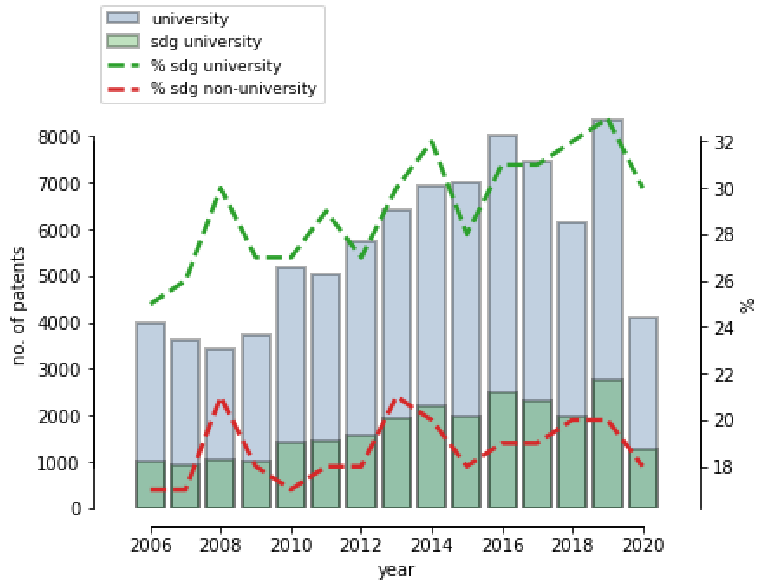
Notes: The table represents the number and percentages of patents for each SDG. In particular, the second column shows the total number of patents for each SDG considering all the patents in our sample. The third column shows the percentages of each SDG patents on the total number of patents. The fourth column instead shows the percentage of each SDG on the total of all SDGs patents. The last three columns respectively represents the total number of each SDG related patent in university patents, the percentage on the total number of university patents and eventually the percentage of each SDG on the total number of university SDG patents. The sum of the percentages are greater than the unity because each patent can be assigned to more than one SDG at the same time.

3.3 SDGs related patents

Through the two rounds of matching, we are able to identify 693'571 SDGs related patents. Figure 2 depicts the trends in SDGs related patents granted from 2006 to 2020. The green line refers to the share of SDGs related university patents on the total of university patents for each year of the range, while the red line depicts the share of non-university SDGs related patents on the total of non-university patents. The share of university patents related to the SDGs is greater than the non-university counterpart, peaking at 33% in 2019, while patents related to the SDGs filed by other actors rise up to around 21%. Further, Table 1 presents the total number of SDGs related patents and the percentage of each SDG over the total of patents (Column 3) and the total of university patents (Column 6). Column 4 and Column 7 respectively report the composition of SDGs patents with respect to each SDG in the total number of patents and of university patents. The results presented in the Table 1 highlight the differences between production of SDGs related innovation in American universities and considering all US assignees. The biggest difference is represented by SDG 3 which is prominent in the total SDG landscape with a share of almost 15 %, but predominant in university innovation production, representing around 53 % of all university patents related to the SDGs. Universities also seem to focus more on SDG 2 (almost 9 % vs 6 % in all patents). This figure is to some extent consistent with the results of van der Waal et al. (2021) who find, using EPO patent data, that the most important contribution of multinational corporations is related to SDG 3. Figure 3 represents the percentage of patents, for each SDG, belonging to the top 10 most common

(IPC): WIPO IPC Green Inventory (WIPO, 2012) and OECD ENV-TECH (Haščič and Migotto, 2015) and to the number of circular economy patents, as defined in Fusillo et al. (2021). In our data, almost 34% of green patents and 63% of circular economy patents are identified also as SDGs-related patents.

Figure 2: Trends in SDGs related patents



Notes: This distribution is related to all university owned patents in the time range 2006-2020.

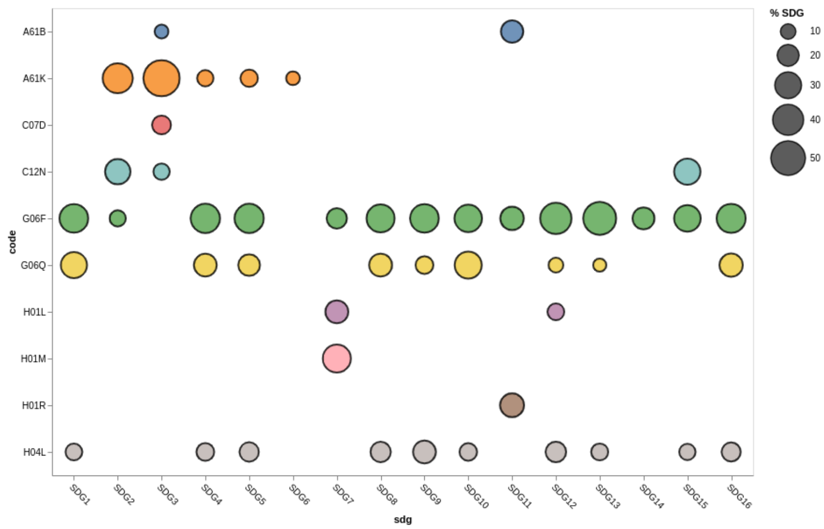
IPC codes. We observe that, on the one hand, out of the 10 codes, only 4 are identify green patents, namely H01L, H01M, C12N and G06Q according to the classification of the WIPO *Green Inventory*. It is to be noted that green IPC codes stand out more in SDGs related to environmental innovation. On the other hand, most of the SDGs seem to be associated with the G06F class which is related to computer systems based on specific computational models and with the H04L class which covers the transmission of digital information. Both of these classes belong to the high-tech IPC classification[12]The definition of high-technology patents proposed by Eurostat uses specific subclasses of the International Patent Classification (IPC) as defined in the trilateral statistical report of the EPO, JPO and USPTO. The list is accessible at https://ec.europa.eu/eurostat/cache/metadata/Annexes/pat_esms_an2.pdf Another relevant technological class for some of the SDGs, especially for SDG 3, is A61K, which refers to medical preparations and pharmaceutical products. These results are especially interesting considering that currently the IPC classes related to sustainable development are limited to green technologies and do not cover "improved conditions" and social sustainability, which is currently possible to identify through semantic search.

3.3.1 Universities' SDGs related patents

This research has a special focus on universities' SDGs innovation. To this end, we identify 25'247 SDGs-related university owned patents. Figure 4 is a network where each of the 25'247 university SDG-related

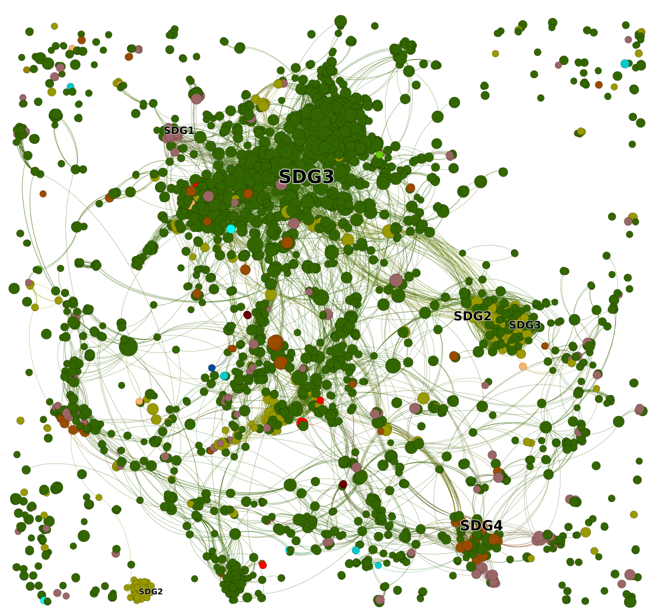
[12].

Figure 3: Top 10 most comment 4-digit IPC codes for SDGs related patents



Notes: The figure represents the percentage of patents, for each SDG, belonging to the top 10 most frequent 4-digit IPC codes in the total distribution. The bubbles are proportional to the percentage and each color is linked to a specific IPC code.

Figure 4: SDGs cosine similarity network (university patents only)



Notes: The figure represents a section of a network where each patent of the 25'247 identified SDGs university patents is a node and the edges' length is proportional to the cosine distance between each couple of patents. For the purpose of clarity, we plot only the edges where the cosine similarity is above a threshold of 0.6.

Table 2: Top 15 universities for share of SDGs related patents

University	No. patents SDGs	No. patents	% SDG
Johns Hopkins University	620	1'691	37
University of South Florida	427	1'145	37
New York University	339	925	37
Duke University	314	897	35
University of North Texas	562	1'661	34
University of Florida	323	1'038	31
Ohio university	842	2'787	30
Cornell University	359	1'185	30
University of Illinois	290	967	30
University of California	2'136	7'500	28
Michigan State University	345	1'269	27
Stanford University	677	2'597	26
University of Wisconsin Madison	375	1'530	25
Massachusetts Institute of Technology	733	3512	21
California Institute of Technology	337	1'945	17

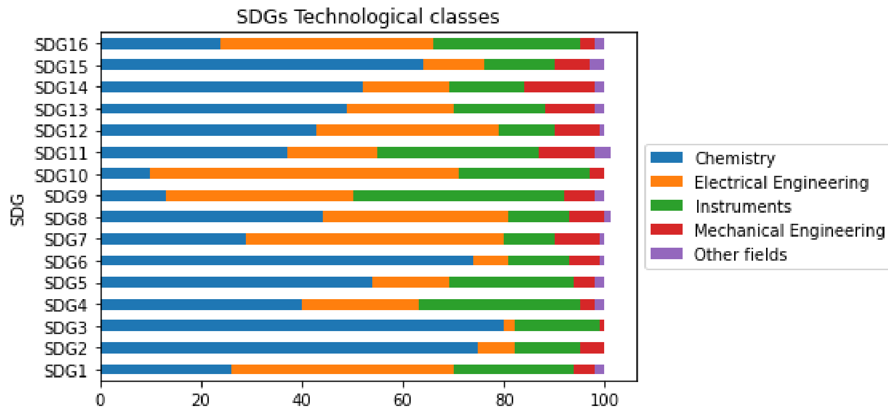
Notes: The ranking is based on the number of USPTO patents held by American universities in the 2006-2020 time span.

patents is a node and the edges are weighted according to the cosine similarity among the vectorial representation of each patent. From the graph it is clear that university SDGs innovation production mostly revolves around SDG 3 and SDG 2 and, interestingly, the patents related to both are semantically closer one another. This might be a hint about the diversity of these patents, which will be verified through the econometric estimations in Section 3.6. Table 2 reports the top 15 universities in terms of SDGs related patents in the time span 2006-2020. In absolute numbers, the best performer is University of California with 2'136 SDGs related patents. However, looking for universities with the highest share of SDGs related patents the best performer are the Johns Hopkins University, the University of South Florida and the New York University with a share of around 37% SDGs related patents. Figure 5 shows the percentage of the 5 WIPO classes composing each of the 16 SDGs. Chemistry seems to be the prevailing class in almost all the SDGs, especially for SDG 2 (Zero Hunger), SDG 3 (Good Health and Wellbeing) and SDG 6 (Clean water and Sanitation) where the Chemistry share is around 75%. This seems reasonable, considering health and sanitation are both fields where chemical technologies are fundamental. Furthermore, SDG 7 (Affordable and Clean Energy) and SDG 9 (Industry, Innovation and Infrastructure) have the greatest share of Electrical Engineering patents, around 50%. This is consistent with the fact that, being focused on innovation, SDG 9 includes many ICT related technologies. At the same time, ICT technologies seem to be relevant also for more socially oriented goals, such as SDG 16 (Peace, Justice and Strong Institutions) and SDG 8 (Decent work and Economic Growth).

3.4 Technological diversity of patents through Rao-Stirling index

As mentioned before, interdisciplinarity can be decomposed in three main components: variety, balance and disparity (Leydesdorff, 2018; Wang et al., 2015), which are accounted for in the Rao-Stirling index.

Figure 5: SDGs technological class composition



Notes: The graph represents the technological composition in percentage of each SDG (SDG 1 to SDG 16) according to the five WIPO technological classes (Chemistry, Electrical Engineering, Instruments, Mechanical Engineering and Other fields).

The Rao-Stirling index has been widely used to measure diversity and in general interdisciplinarity (Porter et al., 2007; Porter and Rafols, 2009). The indicator is defined as follows:

$$\Delta = \sum_{ij} p_i p_j d_{ij} \tag{4}$$

where d_{ij} is a measure of cognitive distance between classes i and j and p_i and p_j are the proportions of elements assigned respectively to class i and j. Considering patents as our work unit, the classes are, in this case, 4-digits-IPC codes to which patents are co-assigned. Thus, the diversity of each patent is calculated as the proportion of IPC4 codes weighted by their cognitive distance.

The first step is calculating the cognitive distance. This research uses co-classification, the frequency with which two classes are assigned to a patent, as a measure of cognitive distance (defined also as 'proximity'). Considering a patent usually belongs to multiple classes, scholars have often used co-classification of patents to develop indicators of distance among technological fields (Engelsman and van Raan, 1994). The underlying assumption is that if the frequency with which two classes are jointly assigned is high, these two classes are proximate (Yan and Luo, 2017). Thus, we use co-occurrences between IPC4 codes as a measure of cognitive distance, creating a symmetric co-occurrence matrix C in which each term C_{ij} represents the number of patents linked to IPC class i and j.

However, the recent work by Alstott et al. (2017) highlighted that all the measures of technology proximity might be affected by factors other than the technologies themselves. In the case of measures of co-occurrence, the authors claim that the simple number of occurrences can be considered as an impinging factor. In particular, the probability that two classes co-occur in the same patent depends on the number of classes that are associated with a patent and the number of patents that are associated with a certain technological class. Bottazzi and Pirino (2010) propose to overcome this issue comparing the observed co-occurrence against the null hypothesis in which the co-occurrences of classes are randomly distributed, preserving, at the same time, both the number of occurrences of a class and the number of classes that are associated with the selected patent. Alstott et al. (2017) propose to verify this null hypothesis creating 1000 randomized control matrices in which the number of occurrences of each class and the number of patents per class are preserved. However, the two constraints proposed by Bottazzi and Pirino (2010) are consistent with the moments (such as mean and standard deviation) of a hypergeometric distribution as explained in the seminal work of Teece et al. (1994). Hence, under the assumption of joint random occurrences and hypergeometric distribution, the mean and the standard deviation are calculated as following:

$$\mu = \frac{C_i C_j}{N} \tag{5}$$

$$\sigma_{ij}^2 = \mu_{ij}(\frac{N - C_i}{N})(\frac{N - C_j}{N - 1}) \tag{6}$$

Where N is the number of patents, C_i and C_j the numbers of patents respectively linked to class i and class j. The relation between class i and j can be express through a z-score, where C_{ij} is the empirical co-occurrence value between class i and j:

$$r_{ij} = \frac{C_{ij} - \mu_{ij}}{\sigma_{ij}} \tag{7}$$

Calculating term r for every couple of IPC classes we obtain a matrix R_{ij} and using a cosine normalization, we obtain a S_{ij} matrix whose values range from 0 to 1. At this point, it is possible to calculate the cognitive distance for each patent, through the following:

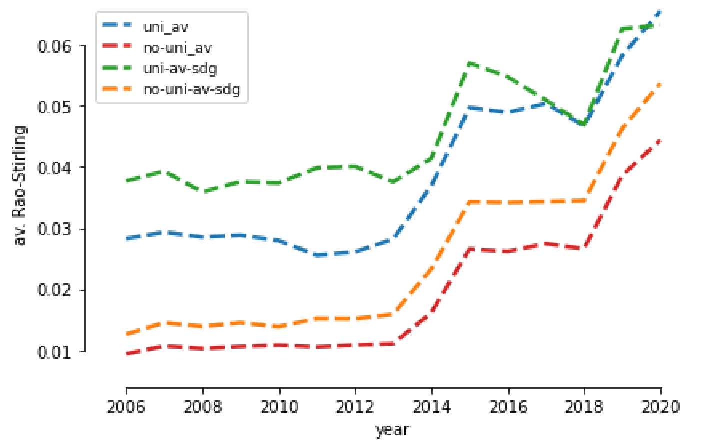
$$d_{ij} = 1 - s_{ij} \tag{8}$$

whose result is used to weight the proportion of technologies in Rao-Stirling index.

3.5 Technological diversity assessment

Figure 6 represents the average of Rao-Stirling index, calculated as previously mentioned and plotted for the time range 2006-2020. The green line represents the trend of university patents related to SDGs, while the blue line represents university patents not related to SDGs. University patents (blue and green lines) are, on average, more diverse compared to non university patents (red and orange lines). This might be explained considering that diverse innovation, such as the one required by SDGs, might be better tackled by collective efforts of universities and research institutions where teams of inventors collaborate to generate innovation rather than individual inventors (Quatraro and Scandura, 2019; Orsatti et al., 2020). Furthermore, patents related to SDGs seem to be slightly more diverse compared to non-SDG related

Figure 6: Rao-Stirling index



Notes: The graph represents the yearly average of Rao-Stirling index for patents granted between 2006 and 2020.

patents for both university and non-university patents. In particular, university patents perform the best in terms of diversity from 2015 (the year of the adoption of the Agenda for Sustainable Development and of the SDGs). These findings will be further explored through the econometric models presented in the following section.

3.6 Empirical Strategy

3.6.1 Diversity of SDGs vs non-SDGs related patents across different technological fields

In order to empirically test Hypothesis (1) we estimate the following model:

$$\begin{aligned} \Delta_i = \alpha + \beta_1 SDG_i + \beta_2 inventors_i + \beta_3 familySize_i \\ + \beta_4 backCits_i + \beta_5 claims_i + \beta_6 univ_i + IPC.3digit_i + t_i + w_i + \epsilon_i \end{aligned} \tag{9}$$

The dependent variable is Δ_i, the Rao-Stirling diversity index and our focal explanatory variable is SDG_i, taking value 1 if the patent is related to SDGs and 0 otherwise. The model includes controls for the number of claims ($claims_i$) and the number of backwards citations ($backCits_i$), as well as for the number of distinct inventors ($inventors_i$), family size ($familySize_i$), which can be linked to the degree of diversity, and for the fact of being owned by a university ($univ_i$). To account for time varying effects we include a set of 15 dummies for the 15 years span considered (t_i) and a set of state controls to account from geographic heterogeneity within the US (w_i); to account for technological heterogeneity instead, we add narrow technological controls ($IPC.3digit_i$) which is a set of IPC 3-digit dummy variables that capture the specific features of each technological domain. We include the latter control following the suggestion of Barbieri et al. (2020) to increase the robustness of the analysis by eliminating the risk of the coefficient of SDG variable being driven by effects that are related to the different technological fields. The estimation could be biased by ignoring the peculiarities of technological areas, such as the availability of consolidated previous work and the inclination to rely on a broader knowledge base. However, it should be noted that adding these dummies, limits the analysis to those IPC 3-digit codes that include at least one SDG and one non-SDG patent.

Furthermore, all reported standard errors are heteroskedastic robust. Considering our dependent variables ranges between 0 and 1 and there is no consensus about the econometric specification for this kind of dependent variable, the best choice seems to carry out the estimation through OLS regression (Fusillo et al., 2020). Furthermore, we take into account that the sample of considered patents might be highly heterogeneous in terms of patent characteristics and to partially tackle this issue, the analysis is conducted considering separately the 5 WIPO technological macro field (Chemistry, Mechanical Engineering, Instruments, Electrical Engineering and Other).

3.6.2 Diversity of university patents related to the SDGs

After testing for the technological diversity of SDGs related patents, we specifically focus on American universities patent portfolios, to check whether university patents related to each SDG have a diversity premium. Thus, in order to empirically test Hypothesis (2) we estimate the following model using university owned patents only:

$$\begin{aligned}\Delta_i = \alpha + \beta_1 SDG1_i + \beta_2 SDG2_i + \beta_3 SDG3_i + \beta_4 SDG4_i + \beta_5 SDG5_i + \beta_6 SDG6_i + \beta_7 SDG7_i \\ +\beta_8 SDG8_i + \beta_9 SDG9_i + \beta_{10} SDG10_i + \beta_{11} SDG11_i + \beta_{12} SDG12_i + \beta_{13} SDG13_i + \beta_{14} SDG14_i \\ +\beta_{15} SDG15_i + \beta_{16} SDG16_i + \beta_{17} inventors_i + \beta_{18} familySize_i \\ +\beta_{19} backCits_i + \beta_{20} claims_i + \beta_{21} renewal_i + t_i + w_i + \epsilon_i\end{aligned} \tag{10}$$

The dependent variable is Δ_i, the Rao-Stirling diversity index and each of the focal explanatory variables ($SDG1_i$ to $SDG16_i$) is a dummy taking the value 1 if the patent is related to corresponding SDG. The controls are the same as presented in Section 3.6.1.

4 Results

4.1 Technological diversity of SDGs-related patents

Section 2.2 and Section 2.3 put respectively forward the two hypotheses of the present research. The first is that SDGs related patents are more diverse than non-SDGs related ones across the different technological fields. The second hypothesis is that university patents related to the SDGs are more diverse compared to other university patents. Table 3 presents the descriptive statistics. Further, Table 9 in Appendix presents

Table 3: Descriptive statistics (2006-2020)

	Obs	**Mean**	**SD**	**Min**	**Max**
Rao Stirling	3’640’513	.022	0.059	0	.386
SDG	3’640’513	.191	0.393	0	1
SDG 1	3’640’513	.018	0.134	0	1
SDG 2	3’640’513	.011	0.103	0	1
SDG 3	3’640’513	.028	0.164	0	1
SDG 4	3’640’513	.013	0.112	0	1
SDG 5	3’640’513	.037	0.190	0	1
SDG 6	3’640’513	.005	0.072	0	1
SDG 7	3’640’513	.024	0.152	0	1
SDG 8	3’640’513	.006	0.080	0	1
SDG 9	3’640’513	.007	0.085	0	1
SDG 10	3’640’513	.002	0.047	0	1
SDG 11	3’640’513	.065	0.246	0	1
SDG 12	3’640’513	.009	0.093	0	1
SDG 13	3’640’513	.003	0.058	0	1
SDG 14	3’640’513	.002	0.050	0	1
SDG 15	3’640’513	.003	0.053	0	1
SDG 16	3’640’513	.003	0.056	0	1
Family Size	3’303’762	3.747	3.788	1	57
Inventors	3’640’513	2.759	1.944	1	133
Backward cits.	3’303’762	27.603	75.379	0	858
Claims	3’303’728	16.708	10.580	1	803

Notes: Unit of observation: patent. Grant years: 2006-2020

the pairwise correlations.

Table 4 presents the results related to the first hypothesis. Columns 1, 2, 3, 4 and 5 report the estimates for the 5 WIPO technological sectors. The SDG coefficient is positive and signficant in all the cases expect for the Chemistry field where it is significant and negative. Thus, this first set of results mostly confirm our first hypothesis, showing that SDGs related technologies are more diverse as compared to non-SDGs related inventions.

Table 5 presents the results related to the second hypothesis, shedding light on the interplay between university patents, technological diversity and SDGs. Considering SDGs cover wide and distinct areas of knowledge and technologies, it is worth observing them separately. The results show that, when looking at university patents only, the diversity effect is heterogeneous across different SDGs. On the one hand, SDG 2, SDG 3, SDG 4 and SDG 15 have a positive and significant coefficient; on the other hand, SDG 1, SDG 6, SDG 7, SDG 9, SDG 10 and SDG 14 have a significant and negative coefficient, while SDG 5, SDG 8, SDG 11, SDG 12, SDG 13 and SDG 16 have a non significant coefficient.

Overall, these results might be aligned with the literature which defines technologies aiming to tackle the multifaceted issue of our century as more diversified (Fusillo et al., 2020) and the idea that the promotion of interdisciplinarity within higher education institutions has become widespread over the last few decades (Tarrant and Thiele, 2017). However, these results only partially confirm our second

Table 4: OLS regression results of SDG on Rao-Stirling with time state and IPC controls

	(1) Chemistry	(2) Elec.Eng	(3) Instr.	(4) Mech.Eng	(5) Other
SDG	-0.0024***	0.0003***	0.0040***	0.0027***	0.0020***
	(0.0003)	(0.0001)	(0.0003)	(0.0003)	(0.0004)
family_size	-0.0004***	0.0002***	0.0004***	0.0005***	0.0001*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
bwd_cits	0.0000***	0.0000***	-0.0000***	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
claims	0.0001***	0.0000***	0.0000***	-0.0000	0.0000*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
no_inv	-0.0000	-0.0001***	0.0002***	0.0004***	-0.0001
	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
university	0.0010**	0.0018***	0.0035***	0.0048***	0.0067***
	(0.0004)	(0.0003)	(0.0005)	(0.0012)	(0.0022)
_cons	0.1251	0.0386***	0.0460***	0.0122***	0.0203***
	(0.0927)	(0.0041)	(0.0030)	(0.0019)	(0.0038)
Year Dummies	YES	YES	YES	YES	YES
State Dummies	YES	YES	YES	YES	YES
IPC.3digit	YES	YES	YES	YES	YES
Observations	264726	872576	303690	248920	104621
R^2	0.182	0.156	0.133	0.121	0.140

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: The dependent variable is the Rao-Stirling index as defined by Rao (1982). Unit of observation: patent. Grant years: 2006-2020. Heteroskedastic-Robust standard errors in parentheses; ***p<0.01,**p<0.05, *p<0.1

hypothesis, considering there is a diversity premium only for university patents linked to SDG 2, SDG 3, SDG 4 and SDG 15.

4.2 Robustness checks

To check the robustness of our empirical study, we present some additional estimates in this subsection. Nevertheless, we assume that the direction of the previously identified relationships should be robust to the choice of diversity measure, as long as these such measures are intended to capture the diversity construct. Therefore, we further check the robustness of our analysis by using alternative measures of technological diversity. To this end, we choose the index of technological diversity developed by Blau (1977) and recently adopted by Zhu et al. (2022) to measure the degree of knowledge recombination in a patent application. The indicator is defined as follows:

$$\text{Technological Diversity} = 1 - \sum_i \left(\frac{\text{no. of IPC}_i \text{ codes}}{\text{tot IPC codes}} \right)^2 \tag{11}$$

where IPC_i represents each unique 4-digit IPC code assigned to a patent. Further, considering our dependant variable is a continuous variable between 0 and 1, we use a Fractional response model which can be used to model a variable that takes values within a bounded range; the dependent variable may be

Table 5: OLS regression results of SDG on Rao-Stirling with time and state controls (university patents only)

	(1) Rao-stirling	(2) Rai-stirling	(3) Rao-stirling
SDG 1	-0.0095***	-0.0107***	-0.0103***
	(0.0021)	(0.0021)	(0.0021)
SDG 2	0.0055***	0.0063***	0.0066***
	(0.0017)	(0.0018)	(0.0018)
SDG 3	0.0140***	0.0134***	0.0133***
	(0.0007)	(0.0008)	(0.0008)
SDG 4	0.0068***	0.0086***	0.0088***
	(0.0021)	(0.0022)	(0.0022)
SDG 5	-0.0004	-0.0004	-0.0003
	(0.0013)	(0.0013)	(0.0013)
SDG 6	-0.0145***	-0.0136***	-0.0136***
	(0.0026)	(0.0027)	(0.0027)
SDG 7	-0.0116***	-0.0128***	-0.0126***
	(0.0013)	(0.0013)	(0.0013)
SDG 8	0.0003	-0.0002	0.0000
	(0.0032)	(0.0033)	(0.0033)
SDG 9	-0.0157***	-0.0162***	-0.0155***
	(0.0025)	(0.0025)	(0.0025)
SDG 10	-0.0183**	-0.0153*	-0.0148*
	(0.0072)	(0.0080)	(0.0080)
SDG 11	-0.0008	-0.0011	-0.0010
	(0.0016)	(0.0017)	(0.0017)
SDG 12	0.0006	0.0004	0.0005
	(0.0026)	(0.0026)	(0.0026)
SDG 13	0.0039	0.0011	0.0012
	(0.0049)	(0.0049)	(0.0049)
SDG 14	-0.0076*	-0.0091**	-0.0090**
	(0.0045)	(0.0046)	(0.0046)
SDG 15	0.0265***	0.0301***	0.0305***
	(0.0050)	(0.0052)	(0.0053)
SDG 16	-0.0025	-0.0008	-0.0006
	(0.0054)	(0.0056)	(0.0056)
family_size		0.0008***	0.0007***
		(0.0001)	(0.0001)
no_inv		0.0004**	0.0004**
		(0.0002)	(0.0002)
bwd_cits			0.0000***
			(0.0000)
claims			-0.0002***
			(0.0000)
_cons	0.0259***	0.0227***	0.0264***
	(0.0028)	(0.0029)	(0.0030)
Year Dummies	YES	YES	YES
State Dummies	YES	YES	YES
Observations	84488	76957	76957
R^2	0.034	0.032	0.033

$^{*}\ p < .10$, $^{**}\ p < .05$, $^{***}\ p < .01$

Notes: The dependent variable is the Rao-Stirling index as defined by Rao (1982). Unit of observation: patent. Grant years: 2006-2020. Heteroskedastic-Robust standard errors in parentheses; ***p<0.01,**p<0.05, *p<0.1

Table 6: Fractional regression results of SDG on Technological Diversity with time and state and IPC controls

	(1) Chemistry	(2) Elec.Eng	(3) Instr.	(4) Mech.Eng	(5) Other
SDG	0.0075	0.0187***	0.0194***	0.0693***	0.0775***
	(0.0055)	(0.0040)	(0.0069)	(0.0072)	(0.0132)
family_size	0.0033***	0.0252***	0.0122***	0.0137***	0.0061***
	(0.0004)	(0.0005)	(0.0008)	(0.0010)	(0.0017)
bwd_cits	-0.0003***	0.0002***	-0.0003***	0.0002***	0.0001
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0001)
claims	0.0017***	0.0007***	0.0019***	0.0011***	0.0033***
	(0.0002)	(0.0002)	(0.0003)	(0.0004)	(0.0006)
no_inv	0.0317***	-0.0013	0.0111***	0.0190***	0.0210***
	(0.0010)	(0.0008)	(0.0014)	(0.0017)	(0.0031)
university	0.0706***	0.0993***	0.1712***	0.2188***	0.2516***
	(0.0072)	(0.0112)	(0.0114)	(0.0232)	(0.0507)
_cons	-1.2462	-0.8497***	-0.7092***	-1.8281***	-1.6391***
	(0.9112)	(0.0664)	(0.0549)	(0.0411)	(0.1037)
Year Dummies	YES	YES	YES	YES	YES
State Dummies	YES	YES	YES	YES	YES
IPC3.digit	YES	YES	YES	YES	YES
Observations	264726	872576	303690	248920	104621
pseudo R^2	0.062	0.121	0.120	0.133	0.153

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: The dependent variable is the Technological Diversity as defined by Blau (1977). Unit of observation: patent. Grant years: 2006-2020. Robust standard errors in parentheses; ***p<0.01,**p<0.05, *p<0.1

any continuous variable bounded between 0 and 1, so that: $0 \leq y_i \leq 1$ (Papke and Wooldridge, 1996). The results of the robustness checks are presented in Table 6 and in Table 7.[13] In the former, we observe that the previous results hold even if we use as dependent variable the technological diversity (*tech_div*) as defined by Blau (1977). Further, the latter confirms our results for SDG 1, SDG 2, SDG 3, SDG 8, SDG 12, SDG 13 and SDG 16; SDG 5, SDG 11 and SDG 15 become significant keeping the same sign while SDG 6, SDG 7 and SDG 14 take the opposite sign; finally, SDG 4, SDG 9 and SDG 10 lose their significant effect.

Further, considering literature suggests that green technologies are more diverse (Quatraro and Scandura, 2019; Fusillo et al., 2020), we check whether the premium diversity we observe is only due to environmental related SDGs technologies. To this end, in the Appendix we present the results of econometric specification where the independent variable SDG_i is split in the three components: environmental, social and development related. As shown in Table 10, the results confirm the diversity premium for all the three subgroups.

[13] However, the coefficients of Fractional response model cannot be easily interpreted by themselves and to ease interpretation, elasticities should be calculated.

Table 7: Fractional regression results of SDG on Technological Diversity with time and state controls (university patents only)

	(1) Tech_div	(2) Tech_div	(3) Tech_div
SDG 1	-0.1315***	-0.1432***	-0.1411***
	(0.0411)	(0.0439)	(0.0439)
SDG 2	0.1349***	0.1304***	0.1311***
	(0.0263)	(0.0286)	(0.0286)
SDG 3	0.1016***	0.0973***	0.0910***
	(0.0121)	(0.0131)	(0.0132)
SDG 4	-0.0260	-0.0094	-0.0082
	(0.0342)	(0.0364)	(0.0365)
SDG 5	-0.1773***	-0.1848***	-0.1840***
	(0.0228)	(0.0246)	(0.0246)
SDG 6	0.1417***	0.1647***	0.1626***
	(0.0506)	(0.0556)	(0.0556)
SDG 7	0.1057***	0.1159***	0.1162***
	(0.0238)	(0.0255)	(0.0255)
SDG 8	0.0007	-0.0067	-0.0072
	(0.0548)	(0.0574)	(0.0574)
SDG 9	-0.0213	-0.0309	-0.0305
	(0.0536)	(0.0581)	(0.0582)
SDG 10	-0.0262	-0.0790	-0.0724
	(0.1391)	(0.1620)	(0.1622)
SDG 11	-0.1053***	-0.1033***	-0.0988***
	(0.0277)	(0.0300)	(0.0300)
SDG 12	0.0443	0.0313	0.0319
	(0.0449)	(0.0477)	(0.0478)
SDG 13	0.0414	0.0498	0.0480
	(0.0725)	(0.0773)	(0.0773)
SDG 14	0.2247***	0.2104**	0.2114**
	(0.0804)	(0.0861)	(0.0859)
SDG 15	0.1035	0.1276*	0.1252*
	(0.0704)	(0.0758)	(0.0759)
SDG 16	-0.0090	-0.0249	-0.0224
	(0.0919)	(0.1008)	(0.1007)
family_size		0.0126***	0.0142***
		(0.0011)	(0.0011)
no_inv		0.0194***	0.0202***
		(0.0026)	(0.0026)
bwd_cits			-0.0006***
			(0.0001)
claims			-0.0010**
			(0.0004)
_cons	-1.3492***	-1.4682***	-1.4494***
	(0.0509)	(0.0535)	(0.0541)
Year dummies	YES	YES	YES
State dummies	YES	YES	YES
Observations	84488	76957	76957
pseudo R^2	0.049	0.051	0.051

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: The dependent variable is the Technological Diversity as defined by Blau (1977). Unit of observation: patent. Grant years: 2006-2020. Robust standard errors in parentheses; ***p<0.01,**p<0.05, *p<0.1

5 Discussion and Conclusion

Innovation for sustainable development plays a fundamental role in achieving the SDGs by fostering the creation of inventions that can help address the complex challenges of this century, such as environmental and social crises, and improve people's lives through advances in relevant sectors such as energy, medicine, and transportation (Blohmke, 2014; Bannerman, 2020; Rimmer, 2018).
However, the potential contribution of intellectual property to advancing the SDGs does not appear to have been explored in depth by scholars and practitioners. In particular, the World Intellectual Property Organization (WIPO) has only acknowledged the link between SDG 9 and IP (WIPO, 2018), while making a stronger contribution to the mapping of green technologies through the IPC codes based *Green Inventory*. Nevertheless, as we show in this research, green technologies are only a partial response to the challenges of SDGs (van der Waal et al., 2021).

In this work, starting from an initial NLP-derived keywords list, we used a robust, unsupervised methodology, to create a patent-related enriched dictionary that references 16 of the SDGs, allowing us to quantify interest in patents related to the SDGs and identify the most represented technology areas. This is a first contribution of this research, as no such patent-related dictionary has been proposed so far. These dictionaries might serve as a starting point for extracting other sustainability-related information from patent texts, thus improving the use of this type of data, which is not normally intended to facilitate communication about sustainability among stakeholders (Abrahamson and Baumard, 2008).

In addition, semantic-based patent analysis allows us to evaluate and compare which technical areas (i.e., IPC codes) contribute more to achieving the SDGs. This information could not be retrieved if considering only the technological classes of patents, since specific classes related to the Sustainable Development Goals have not yet been proposed (van der Waal et al., 2021). Our results shed light on the fact that green technologies only partially contribute to the achievement of the SDGs, while the predominant role is played by high technologies, whose contribution is in fact hardly recognized in the literature (Kostoska and Kocarev, 2019; Vinuesa et al., 2020). Therefore, this finding calls for a more careful consideration and understanding of the role of digital technologies in achieving the Sustainable Development Goals.
Second, this research also provides information on the role of US universities in producing innovation related to the SDGs, a role that is consistent with their mission to maximize societal benefits from the innovations they produce (Papadimitriou, 2020). In this context, we note that the filing of SDG-related patents by universities is increasing at a faster rate compared to other actors. This could be interpreted as universities becoming more aware that part of their public mission is to support the realization of the SDGs (Owens, 2017; Nilsson et al., 2016). However, university-generated innovations related to the SDGs are not evenly distributed among them: SDG 3 related patents account for more than 50% of the total number of SDGs-related patents from universities. This finding is consistent with van der Waal et al.

(2021) results and it is explicable considering the increasing importance of disease control in our society, especially after COVID19. At the same time, it could give rise to further debate and analysis, because patenting in the pharmaceutical field is not without negative consequences and criticism, especially when the university is the owner of the patent (Sampat, 2020, 2021). With this in mind, universities that are heavily involved in patenting for SDG 3 might consider adopting specific licensing strategies to maximize the associated social benefits, such as requiring licensing companies to distribute the product in developing countries beforehand or setting a fair price for the product at the time of technology transfer (Nelsen, 2002).

Third, this research has shown that, consistently with the literature on green innovation (Fusillo et al., 2020; Quatraro and Scandura, 2019), most techonologies related to the SDGs are more technologically diverse, confirming the first research hypothesis. This finding calls for a consistent policy intervention to better stimulate technological diversity to increase technological progress related to the SDGs and reduce environmental and social pressures.

This is especially true for universities, which are uniquely positioned to lead cross-sectoral implementation of the SDGs and provide an invaluable source of expertise in research and education on all areas of the SDGs (Owens, 2017; Nilsson et al., 2016). Moreover, it is increasingly important for universities to demonstrate not only to their financiers but to all stakeholders their ability to generate positive impacts on the territory through their coupling strategies. In this sense, it is critical to understand how the university generates societal benefits and how research activities impact societal benefits. Thus, the results of this study are intended to provide additional data that can help inform how university research projects generate societal impact. Indeed, interpolation between universities and SDGs does not always lead to a diversity premium for university patents. From this study, universities patents show a diversity premium only for SDG 2, SDG 3, which is the most prevalent in university SDG innovation production, and SDG 15. Thus, most university patents related to the SDGs do not show a diversity premium, even those related to greener SDGs, in contrast with what one would expect from the literature (Barbieri et al., 2020; Quatraro and Scandura, 2019). To improve this situation, we believe that specific investments in research and development that foster interactions between different disciplines and abet the creation of new knowledge-based networks are necessary to increase the diversity of SDG-related patents. In addition, to encourage the recombination of knowledge, it should be easier to obtain funding for interdisciplinary research than for mainstream activities (Rylance, 2015).
However, incentivizing interdisciplinarity does not come without any *caveat*. For instance, recent work by Zhu et al. (2022), showed that patent filings related to high diverse innovation run the risk of having a delayed patent grant, due to the increased complexity and ambiguity for patent examiners. For these reasons, the process to create highly diverse technologies should be carefully monitored, because delays in this kind of patents would entail delay in solving the *grand challenges* they are supposed to tackle.

This study is not free from limitations. First, patent data may not be able to fully capture technological innovation in the context of the SDGs, as it only partially represents a broader range of knowledge and technologies needed to promote sustainable development. Second, this research is based on initial keyword lists that influence the overall results and risk being incomplete or unbalanced across the SDGs, which may lead to an overemphasis or overweighting of some SDGs over others. Further, through this methodology is not possible to identify all SDGs-related patents whose text do not hint at SDGs keywords; thus, we fail to identify this kind of patents (false negative cases). This limitation is especially relevant for the identification of university patents related to the SDGs, as they are deemed to be more oriented towards basicness and therefore make less references to practical applications (Trajtenberg et al., 1997). At the same time, the keywords added by the TF-IDF method contain some "noise" that could affect the overall quality of the results obtained.

Future works should consider different approaches to the patent tagging problem, better exploiting the vectorial representations of patent text, in order to understand how patents are written, so to choose the best text representation model. The ideal situation would be to have a *ground truth* about patents related to the SDGs, allowing classification through more sophisticated machine learning methodologies, therefore leading to the creation of a validated dataset. A combination of these techniques, as well as using other patent features, might also allow classification of those patents that do not explicitly mention SDGs related keywords in their text.

Appendix

Additional tables

Table 8: Comparison of results between the two rounds of matching

SDG	**No. keywords**	**No. keywords_machted**	**Intersection_TFIDF**	**First_matching**	**TFIDF_keywords_matched**	**Second_matching**	**Total_matches**
SDG1	63	16	6	617	52	66’381	66’998
SDG2	135	47	28	15’055	107	24’312	39’367
SDG3	214	151	107	91’903	258	9’226	101’129
SDG4	156	43	16	1’560	134	44’695	46’255
SDG5	117	13	4	3’218	63	132’603	135’821
SDG6	163	112	82	15’380	152	3’332	18’712
SDG7	161	120	78	75’282	227	10’479	85’761
SDG8	155	44	13	2’600	82	20’694	23’294
SDG9	79	34	10	12’414	56	13’857	26’271
SDG10	115	27	6	668	44	7’488	8’156
SDG11	171	112	65	220’220	168	14’984	235’204
SDG12	150	76	46	13’013	130	18’695	31’708
SDG13	107	36	15	1’647	45	10’779	12’426
SDG14	127	38	18	5’664	96	3’723	8’949
SDG15	179	66	29	3’337	120	6’874	10’211
SDG16	130	41	11	2’241	50	9’248	11’306
Total	2222	976	534	426’863	1’784	348’354	**693’571**

Notes: The second column of the table (No. keywords) represents the number of keywords in the original SDGs lists. The third (No. keywords_matched) is the number of keywords that produced at least one match in patent corpus. The fourth column (Intersection_TFIDF) represents the number of common keywords, among the top 10 identified by the TF-IDF, with the original lists. The fifth column (First_matching) is the number of patents whose text at least matched with 1 keyword of the original list. The sixth column (TFIDF_keywords_matched) is the number of TF-IDF identified extra keywords that produced at least one match in the patent corpus. The seventh column (Second_matching) is the number of patents whose text at least matched with one of the extra-keywords produced by the TF-IDF. Finally, the last column (Total_matches) represents the number of total matches for each of SDG considering both rounds of matching. Note that the total number of first and second matches as well of the total matches does not double count patents assigned to more than one SDG or patents double labelled in both the matchings.

Table 9: Pairwise correlation

	raostirling	family_size	no_inv	bwd_cits	claims	
raostirling	1.0000					
family_size	0.0781*	1.0000				
no_inv	0.0319*	0.0319*	1.0000			
bwd_cits	0.0188*	0.1070*	0.0617*	1.0000		
claims	-0.0163*	0.0351*	0.0547*	0.0837*	1.0000	

Notes: Significance level 0.05 or more.

Additional econometric specification

In addition to the models provided, we deem important observing the effect of different kinds of SDGs on technological diversity: the environmental related SDGs (including SDG 6, SDG 7, SDG 11, SDG 13, SDG 14 and SDG 15) (Guo et al., 2020), the "social" related SDGs (including SDG 1, SDG 4, SDG 5, SDG 12, SDG 10, SDG 12, SDG 16) (van der Waal et al., 2021) and the "development" related SDGs (including SDG 2, SDG 3, SDG8 and SDG 9) (WIPO, 2019). Table 10 reports the results of the following specifications:

$$\begin{aligned} \Delta_i = \alpha + \beta_1 SDG_i + \beta_2 inventors_i + \beta_3 familySize_i \\ + \beta_4 backCits_i + \beta_5 claims_i + \beta_6 univ_i + IPC.3digit_i + t_i + w_i + \epsilon_i \end{aligned} \tag{12}$$

where the variable SDG_i respectively correspond to environmental SDGs (SDG_env), social SDGs (SDG_social) and development SDGs ($SDG\_develop$).

Table 10: OLS regression results of the 3 kinds of SDGs on Rao-Stirling with IPC, time and state controls

	(1) Environmental	(2) Social	(3) Development
SDG_env	0.0014***		
	(0.0002)		
family_size	0.0001***	0.0001***	0.0001***
	(0.0000)	(0.0000)	(0.0000)
bwd_cits	-0.0000**	-0.0000**	-0.0000*
	(0.0000)	(0.0000)	(0.0000)
claims	0.0001***	0.0001***	0.0001***
	(0.0000)	(0.0000)	(0.0000)
no_inv	0.0001***	0.0001***	0.0001***
	(0.0000)	(0.0000)	(0.0000)
university	0.0049***	0.0049***	0.0047***
	(0.0003)	(0.0003)	(0.0003)
SDG_social		0.0007***	
		(0.0001)	
SDG_develop			0.0019***
			(0.0002)
_cons	0.1252	0.1253	0.1254
	(0.0874)	(0.0874)	(0.0875)
Observations	1794533	1794533	1794533
Year Dummies	YES	YES	YES
State Dummies	YES	YES	YES
IPC.3digit	YES	YES	YES
R^2	0.167	0.167	0.167

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: The dependent variable is the Rao-Stirling index as defined by Rao (1982). Unit of observation: patent. Grant years: 2006-2020. Heteroskedastic-Robust standard errors in parentheses; ***p<0.01,**p<0.05, *p<0.1

References

Abbott, F. M. (2002), 'Wto trips agreement and its implications for access to medicines in developing countries', *Study Paper 2a, United Kingdom Commission on Intellectual Property Rights* .

Abrahamson, E. and Baumard, P. (2008), What lies behind organizational façades and how organizational façades lie: An untold story of organizational decision making, *in* 'The Oxford handbook of organizational decision making'.

Ahuja, G. and Morris Lampert, C. (2001), 'Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions', *Strategic management journal* **22**(6-7), 521–543.

Albareda-Tiana, S., Vidal-Raméntol, S. and Fernández-Morilla, M. (2018), 'Implementing the sustainable development goals at university level', *International Journal of Sustainability in Higher Education* .

Alstott, J., Triulzi, G., Yan, B. and Luo, J. (2017), 'Mapping technology space by normalizing patent networks', *Scientometrics* **110**(1), 443–479.

Álvarez, I., Etxeberria, P., Alberdi, E., Pérez-Acebo, H., Eguia, I. and García, M. J. (2021), 'Sustainable civil engineering: Incorporating sustainable development goals in higher education curricula', *Sustainability* **13**(16), 8967.

Annan-Diab, F. and Molinari, C. (2017), 'Interdisciplinarity: Practical approach to advancing education for sustainability and for the sustainable development goals', *The International Journal of Management Education* **15**(2), 73–83.

Arthur, W. B. (2007), 'The structure of invention', *Research policy* **36**(2), 274–287.

Audia, P. G. and Goncalo, J. A. (2007), 'Past success and creativity over time: A study of inventors in the hard disk drive industry', *Management science* **53**(1), 1–15.

Bannerman, S. (2020), 'The world intellectual property organization and the sustainable development agenda', *Futures* **122**, 102586.

Barbieri, N., Marzucchi, A. and Rizzo, U. (2020), 'Knowledge sources and impacts on subsequent inventions: Do green technologies differ from non-green ones?', *Research Policy* **49**(2), 103901.

Biancani, S., Dahlander, L., McFarland, D. A. and Smith, S. (2018), 'Superstars in the making? the broad effects of interdisciplinary centers', *Research Policy* **47**(3), 543–557.

Blakeney, M. (2009), *Intellectual property rights and food security*, Cabi.

Blasco, N., Brusca, I. and Labrador, M. (2020), 'Drivers for universities' contribution to the sustainable development goals: An analysis of spanish public universities', *Sustainability* **13**(1), 89.

Blau, P. M. (1977), *Inequality and heterogeneity: A primitive theory of social structure*, Vol. 7, Free Press New York.

Blohmke, J. (2014), 'Technology complexity, technology transfer mechanisms and sustainable development', *Energy for Sustainable Development* **23**, 237–246.

Bonino, D., Ciaramella, A. and Corno, F. (2010), 'Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics', *World Patent Information* **32**(1), 30–38.

Bottazzi, G. and Pirino, D. (2010), 'Measuring industry relatedness and corporate coherence', *Available at SSRN 1831479* .

Ceulemans, K., Lozano, R. and Alonso-Almeida, M. D. M. (2015), 'Sustainability reporting in higher education: Interconnecting the reporting process and organisational change management for sustainability', *Sustainability* **7**(7), 8881–8903.

Chen, J., Chen, J., Zhao, S., Zhang, Y. and Tang, J. (2020), 'Exploiting word embedding for heterogeneous topic model towards patent recommendation', *Scientometrics* **125**(3), 2091–2108.

Cordova, M. F. and Celone, A. (2019), 'Sdgs and innovation in the business context literature review', *Sustainability* **11**(24), 7043.

De Marchi, V. (2012), 'Environmental innovation and r&d cooperation: Empirical evidence from spanish manufacturing firms', *Research policy* **41**(3), 614–623.

De Marchi, V. and Grandinetti, R. (2013), 'Knowledge strategies for environmental innovations: the case of italian manufacturing firms', *Journal of knowledge management* .

Deuten, J. (2003), 'Cosmopolitanising technologies. a study of four emerging technological regimes'.

El-Jardali, F., Ataya, N. and Fadlallah, R. (2018), 'Changing roles of universities in the era of sdgs: rising up to the global challenge through institutionalising partnerships with governments and communities', *Health research policy and systems* **16**(1), 1–5.

Engelsman, E. C. and van Raan, A. F. (1994), 'A patent-based cartography of technology', *Research Policy* **23**(1), 1–26.

Feng, S. (2020), 'The proximity of ideas: An analysis of patent text using machine learning', *PloS one* **15**(7), e0234880.

Fusillo, F., Quatraro, F. and Santhià, C. (2021), The geography of circular economy technologies in europe: evolutionary patterns and technological convergence, *in* 'Research Handbook of Innovation for a Circular Economy', Edward Elgar Publishing.

Fusillo, F. et al. (2020), Are green inventions really more complex? evidence from european patents, Technical report, University of Turin.

Guo, M., Nowakowska-Grunt, J., Gorbanyov, V. and Egorova, M. (2020), 'Green technology and sustainable development: Assessment and green growth frameworks', *Sustainability* **12**(16), 6571.

Hajikhani, A. and Suominen, A. (2021), 'The interrelation of sustainable development goals in publications and patents: A machine learning approach'.

Hargadon, A. and Sutton, R. I. (1997), 'Technology brokering and innovation in a product development firm', *Administrative science quarterly* pp. 716–749.

Haščič, I. and Migotto, M. (2015), 'Measuring environmental innovation using patent data'.

Henry, C. and Stiglitz, J. E. (2010), 'Intellectual property, dissemination of innovation and sustainable development', *Global Policy* **1**(3), 237–251.

Hernandez-Aguilera, J. N., Anderson, W., Bridges, A. L., Fernandez, M. P., Hansen, W. D., Maurer, M. L., Ilboudo Nebie, E. K. and Stock, A. (2021), 'Supporting interdisciplinary careers for sustainability', *Nature Sustainability* **4**(5), 374–375.

Idris, K. (2003), 'A power tool for economic growth', *WIPO publication* **888**.

Jaffe, A. B. and Trajtenberg, M. (2002), *Patents, citations, and innovations: A window on the knowledge economy*, MIT press.

Jones, P., Selby, D. and Sterling, S. (2010), 'More than the sum of their parts? interdisciplinarity and sustainability', *Sustainability education: Perspectives and practice across higher education* pp. 17–38.

Kestin, T., van den Belt, M., Denby, L., Ross, K., Thwaites, J. and Hawkes, M. (2017), 'Getting started with the sdgs in universities: A guide for universities, higher education institutions, and the academic sector'.

Kopnina, H. (2020), 'Education for the future? critical evaluation of education for sustainable development goals', *The Journal of Environmental Education* **51**(4), 280–291.

Körfgen, A., Förster, K., Glatz, I., Maier, S., Becsi, B., Meyer, A., Kromp-Kolb, H. and Stötter, J. (2018), 'It'sa hit! mapping austrian research contributions to the sustainable development goals', *Sustainability* **10**(9), 3295.

Kostoska, O. and Kocarev, L. (2019), 'A novel ict framework for sustainable development goals', *Sustainability* **11**(7), 1961.

Lam, J. C., Walker, R. M. and Hills, P. (2014), 'Interdisciplinarity in sustainability studies: a review', *Sustainable Development* **22**(3), 158–176.

Leal Filho, W., Shiel, C., Paço, A., Mifsud, M., Ávila, L. V., Brandli, L. L., Molthan-Hill, P., Pace, P., Azeiteiro, U. M., Vargas, V. R. et al. (2019), 'Sustainable development goals and sustainability teaching at universities: Falling behind or getting ahead of the pack?', *Journal of Cleaner Production* **232**, 285–294.

Lemley, M. A. (2007), 'Are universities patent trolls', *Fordham Intell. Prop. Media & Ent. LJ* **18**, 611.

Leydesdorff, L. (2018), 'Diversity and interdisciplinarity: how can one distinguish and recombine disparity, variety, and balance?', *Scientometrics* **116**(3), 2113–2121.

Leydesdorff, L., Wagner, C. S. and Bornmann, L. (2019), 'Interdisciplinarity as diversity in citation patterns among journals: Rao-stirling diversity, relative variety, and the gini coefficient', *Journal of Informetrics* **13**(1), 255–269.

Li, S., Hu, J., Cui, Y. and Hu, J. (2018), 'Deeppatent: patent classification with convolutional neural networks and word embedding', *Scientometrics* **117**(2), 721–744.

Lozano, R., Lukman, R., Lozano, F. J., Huisingh, D. and Lambrechts, W. (2013), 'Declarations for sustainability in higher education: becoming better leaders, through addressing the university system', *Journal of Cleaner Production* **48**, 10–19.

Morse, W. C., Nielsen-Pincus, M., Force, J. E. and Wulfhorst, J. (2007), 'Bridges and barriers to developing and conducting interdisciplinary graduate-student team research', *Ecology and Society* **12**(2).

Nelsen, L. (2002), 'The role of university technology transfer operations in assuring access to medicines and vaccines in developing countries', *Yale J. Health Pol'y L. & Ethics* **3**, 301.

Newell, W. H., Wentworth, J. and Sebberson, D. (2001), 'A theory of interdisciplinary studies', *Issues in Interdisciplinary Studies* .

Nilsson, M., Griggs, D. and Visbeck, M. (2016), 'Policy: map the interactions between sustainable development goals', *Nature* **534**(7607), 320–322.

Noh, H., Jo, Y. and Lee, S. (2015), 'Keyword selection and processing strategy for applying text mining to patent analysis', *Expert Systems with Applications* **42**(9), 4348–4360.

Oguamanam, C. (2006), 'Intellectual property rights in plant genetic resources: Farmers' rights and food security of indigenous and local communities', *Drake J. Agric. L.* **11**, 273.

Orsatti, G., Quatraro, F. and Pezzoni, M. (2020), ‘The antecedents of green technologies: The role of team-level recombinant capabilities’, *Research Policy* **49**(3), 103919.

Owens, T. L. (2017), ‘Higher education in the sustainable development goals framework’, *European Journal of Education* **52**(4), 414–420.

Papadimitriou, A. (2020), ‘Beyond rhetoric: Reinventing the public mission of higher education’.

Papke, L. E. and Wooldridge, J. M. (1996), ‘Econometric methods for fractional response variables with an application to 401 (k) plan participation rates’, *Journal of applied econometrics* **11**(6), 619–632.

Popp, D. and Newell, R. (2012), ‘Where does energy r&d come from? examining crowding out from energy r&d’, *Energy economics* **34**(4), 980–991.

Porter, A., Cohen, A., David Roessner, J. and Perreault, M. (2007), ‘Measuring researcher interdisciplinarity’, *Scientometrics* **72**(1), 117–147.

Porter, A. and Rafols, I. (2009), ‘Is science becoming more interdisciplinary? measuring and mapping six research fields over time’, *Scientometrics* **81**(3), 719–745.

Quatraro, F. and Scandura, A. (2019), ‘Academic inventors and the antecedents of green technologies. a regional analysis of italian patent data’, *Ecological Economics* **156**, 247–263.

Rafols, I. and Meyer, M. (2010), ‘Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience’, *Scientometrics* **82**(2), 263–287.

Rao, C. R. (1982), ‘Diversity and dissimilarity coefficients: a unified approach’, *Theoretical population biology* **21**(1), 24–43.

Reimers, N. and Gurevych, I. (2019), ‘Sentence-bert: Sentence embeddings using siamese bert-networks’, *arXiv preprint arXiv:1908.10084* .

Rimmer, M. (2014), ‘Intellectual property and global warming: fossil fuels and climate justice’, *The Sage Handbook of Intellectual Property, London: Sage Publications* pp. 727–753.

Rimmer, M. (2018), ‘A submission on intellectual property and the united nations sustainable development goals’.

Rogers, F. T. (2020), ‘Patent text similarity and cross-cultural venture-backed innovation’, *Journal of Behavioral and Experimental Finance* **26**, 100319.

Roudsari, A. H., Afshar, J., Lee, S. and Lee, W. (2021), Comparison and analysis of embedding methods for patent documents, *in* ‘2021 IEEE International Conference on Big Data and Smart Computing (BigComp)’, IEEE, pp. 152–155.

Rylance, R. (2015), 'Grant giving: Global funders to focus on interdisciplinarity', *Nature* **525**(7569), 313–315.

Sampat, B. N. (2020), 'Whose drugs are these?', *Issues in Science and Technology* **36**(4), 42–48.

Sampat, B. N. (2021), 'The government and pharmaceutical innovation: Looking back and looking ahead', *Journal of Law, Medicine & Ethics* **49**(1), 10–18.

Sánchez-Carracedo, F., Ruiz-Morales, J., Valderrama-Hernández, R., Muñoz-Rodríguez, J. M. and Gomera, A. (2021), 'Analysis of the presence of sustainability in higher education degrees of the spanish university system', *Studies in Higher Education* **46**(2), 300–317.

Stirling, A. (2007), 'A general framework for analysing diversity in science, technology and society', *Journal of the Royal Society Interface* **4**(15), 707–719.

Tarrant, S. P. and Thiele, L. P. (2017), 'Enhancing and promoting interdisciplinarity in higher education', *Journal of Environmental Studies and Sciences* **7**(2), 355–360.

Teece, D. J., Rumelt, R., Dosi, G. and Winter, S. (1994), 'Understanding corporate coherence: Theory and evidence', *Journal of economic behavior & organization* **23**(1), 1–30.

Thomas, I. (2016), Challenges for implementation of education for sustainable development in higher education institutions, *in* 'Routledge Handbook of Higher Education for Sustainable Development', number 4, RMIT University, pp. 56–71.

Trajtenberg, M., Henderson, R. and Jaffe, A. (1997), 'University versus corporate patents: A window on the basicness of invention', *Economics of Innovation and new technology* **5**(1), 19–50.

Tseng, Y.-H., Lin, C.-J. and Lin, Y.-I. (2007), 'Text mining techniques for patent analysis', *Information processing & management* **43**(5), 1216–1247.

UN (2020), Guidebook for the preparation of science, technology and innovation (sti) for sdgs roadmaps, Technical report.

Usui, S., Palmes, P., Nagata, K., Taniguchi, T. and Ueda, N. (2007), 'Keyword extraction, ranking, and organization for the neuroinformatics platform', *Biosystems* **88**(3), 334–342.

van der Waal, J. W., Thijssens, T. and Maas, K. (2021), 'The innovative contribution of multinational enterprises to the sustainable development goals', *Journal of Cleaner Production* **285**, 125319.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M. and Fuso Nerini, F. (2020), 'The role of artificial intelligence in achieving the sustainable development goals', *Nature communications* **11**(1), 1–10.

Walz, R., Pfaff, M., Marscheider-Weidemann, F. and Glöser-Chahoud, S. (2017), 'Innovations for reaching the green sustainable development goals–where will they come from?', *International Economics and Economic Policy* **14**(3), 449–480.

Wang, J., Thijs, B. and Glänzel, W. (2015), 'Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity', *PloS one* **10**(5), e0127298.

Weitzman, M. L. (1998), 'Recombinant growth', *The Quarterly Journal of Economics* **113**(2), 331–360.

Winickoff, D. E. (2013), 'Private assets, public mission: The politics of technology transfer and the new american university', *Jurimetrics* pp. 1–42.

WIPO (2012), 'Ipc green inventory'.

WIPO (2018), 'Report on wipo's contribution to the implementation of the sustainable development goals and its associated targets'.

WIPO (2019), 'Report on wipo's contribution to the implementation of the sustainable development goals and its associated targets'.

Xie, Z. and Miyazaki, K. (2013), 'Evaluating the effectiveness of keyword search strategy for patent identification', *World Patent Information* **35**(1), 20–30.

Yan, B. and Luo, J. (2017), 'Measuring technological distance for patent mapping', *Journal of the Association for Information Science and Technology* **68**(2), 423–437.

Zeppini, P. and van Den Bergh, J. C. (2011), 'Competing recombinant technologies for environmental innovation: extending arthur's model of lock-in', *Industry and Innovation* **18**(03), 317–334.

Zhu, K., Malhotra, S. and Li, Y. (2022), 'Technological diversity of patent applications and decision pendency', *Research Policy* **51**(1), 104364.