

#2002 Bremen Papers on Economics & Innovation

From biotech to bioeconomy

New empirical evidence on the technological transition to plant-based bioeconomy based on patent data

Muhamed Kudic*
Mariia Shkolnykova*

January 2020

Abstract

Nowadays environmental and climate issues have brought the topic of bioeconomy to the political agenda around the world. Plant-based bioeconomy (pBE) has a key role in securing sustainable supplies of energy, food and raw materials for increasingly aging and growing societies. However, the technological roots and development path of pBE are far from being fully understood. Accordingly, we seek to contribute to an in-depth understanding of how biotechnology innovations affected the emergence of bioeconomy by exploring the technological field evolution of plant-based patent applications between 1995 and 2015 in Germany. We employ patent citation data and conduct forward citation analysis to trace technological trajectories within plant-based biotechnology. We extend previous work by combining patent-based citation analysis with text-mining approach. Main path analysis allows the identification of main players within plant-based biotechnology over time. Our explorations reveal dominant and also peripheral technologies within the sphere of plant-based applications and provide us in this way with a more comprehensive understanding of the field's technological evolution. Our findings suggest a transition from basic biotechnological research towards more sustainability- and medicine-related technological orientation in the field.

* University of Bremen, Max-von-Laue-Straße 1,
28359 Bremen, Germany
Corresponding author: Maria Shkolnykova
E-mail: mariia.shkolnykova@uni-bremen.de

Keywords

Plant-based bioeconomy, Technological change, Technological path analysis, Patent analysis, Text mining

JEL Classifications

O30, O33, L65

1. Introduction

Today, the world is facing some fundamental societal challenges such as demographic change, food shortages and further environmental issues. These developments and the urgent search for potential solutions brought the topic of bioeconomy to the political agenda. For example, the National Research Strategy BioEconomy of the German Federal Ministry of Education and Research aims to achieve a more efficient use of biological resources by 2030. EU Research and Innovation Program Horizon 2020 considers bioeconomy to be the answer to various environmental challenges and points its aim as “production of renewable resources (...) and their conversion into (...) bio-based products and bio-energy”.

The breadth of bioeconomy as a scientific field and societal concept is enormous. Bio-based developments enhance commercial applications in chemistry, engineering, pharmaceutical and food industry. When thinking about renewable resources, used in bioeconomy production, plant compounds are usually put forward as its most probable source. One highly influential technological root of this emerging field can be found in biotechnological innovation processes, particularly those dealing with plant usage and breeding. Accordingly, for the purpose of this paper we focus on the plant-based bioeconomy. Plant components here are considered in a broader sense, encompassing different sugars, oils, fats and pigments (Lewandowski, 2017). As such, the case of plant-based bioeconomy provides a rich field to study the transition from biotechnology (especially “green biotechnology”) towards bioeconomy.

The contemporary political and scientific discussion points to the fact that plant-based bioeconomy (pBE) becomes more and more important in securing sustainable supplies of energy, food and raw materials for increasingly aging and growing societies and changing climatic conditions. With the help of the renewable resources bioeconomy may decrease or even eliminate dependence on fossil fuels and thus release the burden on the environment (BMBF, 2014). However, technological roots and the development path of pBE are far from being fully understood. National Research Strategy BioEconomy 2030 of German Federal Government outlines, that “...A combination of approaches in breeding, plant biotechnology, and other disciplines will be necessary to provide technical solutions for these requirements”, thus pointing out the role of biotechnology research in this transformation (BMBF, 2011). To the best of our knowledge, no longitudinal empirical analysis on the transition from biotech to bioeconomy has been conducted so far.

Apart from that, it is yet not clear, what technologies exactly can be seen as bioeconomy-related. Although several policy documents clearly state the importance of biotechnology applications for this societal area, not all subfields of biotech seem to be

From biotech to bioeconomy
New empirical evidence on the technological transition to plant-based bioeconomy based on patent data

welcome in this new field. According to some visions (e.g. bio-ecology vision, outlined in Bugge et al., 2016, which focuses on ethical issues and organic agriculture), such important biotechnology developments as transgenic crops cannot be included in bioeconomy (Bugge et al., 2016). Therefore, understanding of the dominant technological paths within biotechnology helps to reveal, whether more weight is added to sustainability topics over time, which will signal the transition to bioeconomy. Taking into account issues stated above as well as the importance of bioeconomy for sustainability agenda, an in-depth understanding of the field is important for both managers and political decision makers alike. Hence, in this paper we seek to explore the technological field evolution of biotechnology from various angles in order to answer the following research question: do we observe a transition from biotechnology towards bioeconomy on the example of pBE¹?

To do so, we compile a unique data set for the German biotech industry, covering a period of 20 years (1995-2015), along the cycle of industry development, from the initiation in mid 90s to the last year, for which patent population could be constructed. The dataset covers the full set of all plant-based biotechnology (pBT) patents and all direct forward citations.

There are several good reasons for focusing on the German case. First, the pharmaceutical and chemical industry, strongly involved in biotechnological research, have a longstanding tradition in Germany. Second, funding schemes – promoting the emergence of biotechnology as an entirely new technological field – were initiated and continuously supported by the Federal Ministry of Education and Science (BMBF) in the mid-90s. Third, the German scientific landscape is characterized by high density of universities, applied universities and research institutions covering the full spectrum from applied research (e.g. Fraunhofer Society) to basic research (e.g. Max Planck Society). This, in turn, provides a fertile ground for the emergence of bioeconomy.

To gain an in-depth understanding of the technological transition from biotech to bioeconomy in Germany, we apply a three-stage research design. In line with Verspagen (2007), Nomaler, Verspagen (2016) and Fontana et al. (2009), we reconstruct the technological landscape of biotechnology based on patent application data by using a graph theoretical approach. In addition, we include direct patent citation data to account for a broader range of bio-related patents. This, in turn, allows us to gain a comprehensive picture of the technological field of biotech and beyond. We employ network analysis methods on direct graphs to measure and analyze the system's topology from various angles and identify main paths and peripheral areas. Finally, we combine our empirical network approach with a content-related text mining analysis.

¹ Although the definition of bioeconomy is rather broad and it stems not only from biotechnology, but also from other fields, including agriculture, waste disposal etc. In this paper, however, we are limiting ourselves to the part of bioeconomy, coming from plant biotechnology.

Latter allows looking deeper in context dimension of chosen patents and understand deeper where the field is going.

The remainder of the paper is organized as follows: we start in section 2 with some theoretical considerations and an assessment of the literature on technological trajectories as well as the main path identification. Next, we introduce the German biotechnology and provides an overview of the data and methodology used for the purpose of this paper. Section 4 provides the results of our empirical analyses. The paper ends with a discussion of our key findings, limitations and concluding remarks.

2. Theoretical background and state of the art

Within the scope of neoclassical economics firms are typically considered as a black box in which production processes takes place under fixed technology and against the backdrop of a highly idealized set of assumptions. It was Schumpeter (1912, 1939) who emphasized the role of innovation and technological advancements as the main driver of economic change and prosperity (Witt, 2008). Schumpeter was convinced that economic change is generated within the economic system and that there must be some kind of inherent force – today typically referred to as innovations - that destroys every equilibrium state (Hanusch and Pyka 2007).

This notion was embraced several years later and initiated the emergence of a new and influential strand of economics (Nelson and Winter 1974, 1982), known as evolutionary economics or neo-Schumpeterian school of thought. This approach differs in various aspects from mainstream economics. First, economic actors are considered to be heterogeneous, bounded rational agents following individual strategies, which are rather the result of trial-and error heuristics than fully rational and deterministic optimization procedures. Second, knowledge and learning become main cornerstones of the approach. It's the agents' ability to generate novelty that allows them to cope with uncertainty and complexity in highly volatile and rapidly changing environments. Third, market equilibria are rather an exception than the rule. Hence, the status quo is typically highly fragile and endogenous drivers – e.g. entrepreneurs or established companies initiating radically new ideas – fuel the ongoing systemic change. Finally, technological developments at higher aggregation levels mirror the individual innovation efforts of actors at the micro level. Literature in this field shows that technological change processes are context-specific (Boschma 2005), path-dependent (Arthur, 1989; David, 1985) and cumulative in nature (Dosi, Grazzi, 2010), as they are building upon previous experience, successes and failures. As a consequence, the destruction or radical change of existing industries due to technological progress is considered from a neo-Schumpeterian perspective to be a natural constituent of an economy.

For the purpose of this paper, we are particularly interested in the structural dimension of technological change. The literature on technological paradigms (Dosi, 1982, Dosi et al. 1988) provides a fertile ground to analyze how path-dependent innovation processes at the micro-level shape the emergence and structuration of technological landscapes. The technological paradigm here is seen as a “model” and a “pattern” of solution of selected technological principles (Dosi, 1982). The selective and cumulative character of paradigm is reflected by technological trajectories as “the path of improvement taken by technology” within it. (Dosi & Nelson, 1994).

Besides these mostly conceptual contributions, a rich body of empirical and methodological literature emerged over the past years and enhanced our understanding of how and why technological change occurs. Patent data and publication data turned out to be highly valuable in this context. For instance, Verspagen (2007) and Fontana et al. (2009) used patent data for analyzing technological graphs, since patent documents contain a broad range of information about technological characteristics of an invention, thus providing knowledge embedded in it. At the same time, the studies of Hummon and Doreian (1989) and Liu, Lu (2012) demonstrate that bibliometric data has a high potential to gain comprehensive insights on technological change as well.

Thus, several authors have applied the idea of technological trajectories in order to identify main paths of development within a specific industry with the help of patent data. Most notably, the seminal paper of Hummon and Doreian (1989) paved the way for a number of follow-up studies by exploring the developments in DNA research. This triggered series of empirical studies, which can be grouped into three overlapping streams.

The first stream of literature uses main path analysis to describe the developments within important societal areas. To start with, Verspagen (2007) depicted technological trajectories in fuel cells research. He asks, whether there is a convergence within technological space. The results along the network of main paths confirm a selective character of technological trajectories within the field. The work of Fontana et al. (2009) extended previous research by providing a fine-grained exploration of technological paths using the example of Ethernet. The findings of the conducted main path analysis confirmed the initial predictions of experts from economics and engineering. One specific advancement contributed by Fontana et al. (2009) is the identification of several distinct subsequent stages along with main path development. Mina et al. (2006) enriched the literature by investigating co-evolutionary processes in medical research by exploring two complementary datasets, namely patent data and bibliometric data.

The second stream of literature focuses on methodological and technical improvements of the main path analysis and contributes significantly to a critical assessment of previous research. Barbera-Tomas et al. (2011) checked the reliability of

connectivity analysis by implementing and conducting external and internal validation procedures based on artificial disc patents data. Connectivity analysis shows that the main path contains reliable knowledge of the field. Liu, Lu (2012) in turn, focused on citation graphs of scientific research papers, dealing with the Hirsch index. The authors distinguish between 'local' and 'global' main paths. They show empirically that the global main path, which has the overall largest traversal count, has to be calculated in order to identify the overall importance of the knowledge flow. One more alternative approach is proposed by Park and Magee (2017), who introduced genetic backward-forward path analysis. This method is based on identifying high persistence patents, which preserve important knowledge of the field. As a result, network complexity may be decreased with dominant knowledge not being lost.

Finally, within the third stream of literature previous technology-centered contributions are extended by applying the main path methodology to other related themes. Most notably Nomaler and Verspagen (2016) applied an evolutionary economic perspective and analyzed first and second-order citations across county boundaries within the US. The main research interest here lied in investigating under which conditions knowledge flows globally or locally. Results showed the co-existence of technological clustering in regions and turned attention to the necessity of connecting locally embedded knowledge via cross-spanning ties.

In this paper, we contribute to the above-outlined literature in at least two ways. On the one hand, to the best of our knowledge, we are the first to go outside the borders of one field and look at the societally initialized normative paradigm change. In doing so, we extend our knowledge of the transition process in a highly topical and relevant technological field. Here transition towards sustainability plays a central role. On the other hand, we complement the methodological toolbox by combining well-established network analysis methods for main path analysis with content-based analysis via using text mining tools. Whereas network analysis in that sense helps us to identify the most influential patents, text-mining tools help to dive deeper in the context of given applications.

3. Technology, data and methods

3.1. Plant-based biotechnology in Germany

The intensification of research and development as well as the beginning of commercial use of pBT dates back to the early 1970s. In 1973 the r-DNA technique was invented (Parayil, 2003). In the following year, 1974, the first patent related to biotechnology was applied for. The patent dealt with the topic of genetic manipulations (Feldman, 2000). The 1980s were characterized by increased interactions between

scientific and commercial sides of the industries. Large biotechnology corporations intensified their investments in research on transgenic organisms by using constantly advancing scientific and technological base (Parayil, 2003).

However, all the initial developments in biotechnology in general – and in pBT in particular – took place in the US. The first noteworthy developments in the field of biotechnology in Europe date back to the early 1990s. Firstly, the period was distinguished by improvement of the legislation in the field of genetic research: regulatory regime for the work of genetic laboratories was introduced along with the legislation by European Union on authorization of medical products (Kaiser, Prange, 2003; Häussler, 2007). Apart from that, research and development activities were strongly supported by several federal programs. In Germany these programs included BioRegio Competition and BioProfile targeting particular regions and BioChance opting at the support of particular projects. These opportunities of financial support led to the founding of many SMEs, which – alongside with established biotech and pharmaceutical companies – constitute the core of German biotechnology.

Nevertheless, not all operating fields of biotechnology have received equal support and development. Whereas red (medicine-related) and white (industry-related) biotechnology were not restricted in their development, green (agricultural) subfield was lagging behind. The reason for that is the extensive inventions, which relate to genetically modified plants, which the EU is negative about (Székács, 2017). These open systems, according to Székács (2017), relate to openness to environment, e.g. to agriculture or fishery. Apart from that, the differences between four generations of GM crops are of high importance here. Whereas the first generation relates mostly to the use of pesticides, second deals with improvement or modification of plants qualities, third deals with industrial and pharmaceutical products and fourth produced with new methods in molecular biology (Székács, 2017). Thus, newer generations of genetic crops may be looked at more favorably from the EU perspective.

In this context, it is important to emphasize that green biotechnology, as well as the white one, are exactly the fields where most of the potential for sustainable development is hidden (McCormick & Kautto, 2013). Many authors consider biotechnology to be an important predecessor and the source of bioeconomy (Bugge et al., 2016). This notion is also supported by several policy documents and initiatives initiated on different levels. According to the National Research Strategy BioEconomy 2030 of German Federal Government: "... A combination of approaches in breeding, plant biotechnology, and other disciplines will be necessary to provide technical solutions for these requirements. This includes not only methods such as genome analysis – including epigenetic studies, proteomic- and metabolomic research, bioinformatics, and the system-biological integration of these approaches – but also the responsible application of genetic engineering" (BMBF, 2011). Apart from that, the importance of biotechnology inventions for bioeconomy introduction was voiced by OECD (2009).

Despite the role of biotechnology in bioeconomy establishment, the issue of GM crops usage is still far from being clear. Some more ethically-oriented (bio-ecology) visions of bioeconomy exclude GM plants from it (Bugge et al., 2016). Furthermore, the replacement of fossil fuels with biomass may lead to other environmental problems, such as land erosion or biodiversity loss, which is critically seen by some NGOs (McCormick, Kautto, 2013). All these things call for the necessity of the identification of how biotechnology is developing and distinguishing the ones, which can be related to bioeconomy.

So where does the field of pBT go in Germany? How is the conflict between set goals towards necessity of providing nutrition and ethical concerns are resolved? Do other connected fields (medicine, chemical engineering etc.) influence this new field? These questions are going to be answered within the following subchapters of this article.

3.2. Data

Data used for further analysis comes from several sources. Patent data used in the paper was taken from PATSTAT Database, Version Autumn 2017. PATSTAT presents consolidated data on patents, registered in different patent offices around the world. Apart from that, applicant/inventor data along with information on technological spheres related to each particular patent is provided. PATSTAT also allows to track forward and backward citations of selected patents.

The information on firm attributes, such as funding date, turnover, number of employees as well as the ownership data was taken from the ORBIS database, which provides a broad range of firm information. Data stem from various sources (e.g. from Creditreform, Handelsregister etc.). In the scope of this paper, such kind of data is used only when analyzing the attributes of the applicants/inventors, who contributed to the field.

One relevant initial issue is the identification of patent classes related to pBT. The choice of the relevant patent classes, was performed as follows. Firstly, the IPC classes which were considered as biotechnology by either OECD (2009) or WIPO (2008) or both were taken². Secondly, these classes were analyzed in order to isolate plant-related classes within the full set of biotechnology patents. As international patent classification has a rather detailed structure, this could be realized by looking at the title of relevant classes, subclasses, main groups or subgroups and filtering out those where the term "plant" appears. Thus, following units of IPC could be identified as being pBT: A01H 1*,

² This means following IPC classes: A01 H1/00, A01 H4/00, A61 K38 /00, A61 K39 /00, A61 K48 /00, C02 F3/34, C07 G, C07 K, C12 M, C12 N, C12 P, C12 Q, C12 R, C12 S, G01 N27 /327, G01 N33 / (53 *, 54 *, 55 *, 57 *, 68 ,74 ,76 ,78 ,88 ,92) as well as later added C40B 40/00 - 50/18, C40B 70/00 - 80/00, C40B 10/00. Later classes can be found in Eurostat indicators on High-tech industry and Knowledge.

A01H 4*, A61K 38/56, C07K 4/10, C07K 14/415, C07K 16/16, C12M 3*, C12N 5/04, C12N 5/14, C12N 9/32, C12N 15/05, C12N 15/29, C12N 15/82, C12Q 1/6895.

Finally, we check this initial choice by analyzing titles and abstracts of biotechnology classes. To proof that this method gives reliable results, we have screened abstracts and titles of patents in order to count, in which biotechnology classes term “plant” appears at most. We are counting patent families, filed between 1995 and 2015 by German applicants or inventors. Here we follow the procedure offered by Parisi et al. (2012). However, as we are interested in the subfield in general and not in its specific technologies, no extra keywords were added. As a result of this keyword search, the initially identified classes could be supported. The unconditional leader of pBT patent classes is the main group C12N 15/82 - *Mutation or genetic engineering; DNA or RNA concerning genetic engineering, vectors, e.g. plasmids, or their isolation, preparation or purification; Use of hosts therefor... Vectors or expression systems specially adapted for eukaryotic hosts ... for plant cells*. Apart from that, many plant-related patents are included in classes C12N 5/10, C12N 15/09, A01H 1/00, C12N 15/29, A01H 4/00, C12Q 1/68. Some of these classes however are rather broad and include many patents, only some of which by coincidence or for a reason include “plant” in them. Thus, for example, the class C12N 5/10 - *Microorganisms or enzymes; Compositions thereof... Undifferentiated human, animal or plant cells, e.g. cell lines; Tissues; Cultivation or maintenance thereof; Culture media therefor... Cells modified by introduction of foreign genetic material, e.g. virus-transformed cells* have more than 10000 patents for the period, where the term “plant” appears in the abstract or in the title. However, in relative terms these patents constitute less than 10% of the overall number of patent applications in this class. Thus, it was not taken into consideration. After filtering out such classes, as well as the ones where less than 100 patents were applied for during the observation period³, the following classes were taken into consideration: A01H 1*, A01H 4*, A61K 38/56, C07K 4/10, C07K 14/415, C07K 16/16, C12M 3*, C12N 5*, C12N 9/32, C12N 15/05, C12N 15/29, C12N 15/82, C12Q 1/6895⁴.

3.3. Sample identification

After sorting out pBT patent classes, a sample population, which was later used for analysis, was constructed. Accordingly, chosen patents should have their earliest filing date between 1995 and 2015, with 1995 being taken as the approximate initiation of biotechnology in Germany, while 2015 is the latest year, for which the PATSTAT 2017 Autumn edition provides whole patent coverage. Apart from that, as we are interested in the German pBT, at least one actor (applicant or inventor), who is related to the patent, should be based in Germany. Being genuinely interested in the flows of knowledge, we

³ As it is hard to drive conclusions about the relatedness of the class to pBT basing only on the small amount of patents.

⁴ Titles of respective patent classes are provided in Appendix A.

do not differentiate between applicants and inventors. Thus, we do not miss important patents from individuals, or patents, performed inside of German organizations by individuals, based elsewhere.

As the focal aim of this paper lies in covering what is happening within pBT in Germany, an additional filter was put on place of patent registration. Thus, only German, European and International patents were taken into account. This helped us to get rid of the possible strategic patenting or corporations, creating something specifically for a particular market, which may not reflect real German pBT development. Apart from that, we have focused on the level of DOCDB simple patent family⁵. According to the European Patent Office, patents, belonging to one patent family, cover a single invention (EPO, 2017). Good example for it is an invention, which is subsequently registered in different countries. These patents, however, have one priority filing date, which is the subject to our interest. After applying these filters, we have obtained sample of 1434 patent families.

Next, citations to these patents were collected. As well as the focal patents, citations should have at least one inventor or applicant, who is based in Germany and be registered at the German, European or International patent office. As the result, 1682 patent families were identified.

3.4. Main path analysis

The analytical part of the paper is based on the 'main path analysis' method, originally proposed by Hummon and Doreian (1989). The method allows identification of dominant trends with the help of directed graphs. Mina et al. (2007) applied the main path analysis to show the existence of main paths in the case of Coronary artery disease. Verspargen (2007) analyzed technological evolution in the field of fuel cell technology. Fontana et. al (2009) complemented this strand of research by providing new evidence on technological change patterns in LAN technology. Liu, Lu (2012) extended the analysis by choosing Hirsch Index as a field of interest. This method was applied in more or less stylized way by Park, Magee (2017), No, Park (2010) and others.

The key idea of Hummon and Doreian (1989) was to find a systematic way of tracing the main streams of field evolution in a graph. In the case of patent data, each vertex of the graph represents a single patent while the edges stand for citations. Thus, main path on patent data allows for the identification of the dominant trajectory/trajectories, representing the most distinct strand of innovations in a well-specified technological field. Graphs typically show a highly complex structure. Hence,

⁵ We are using earliest filing date (the filing date of priority filing) for our analysis, as it is the year of knowledge generation. Apart from that, as we are interested in the process of knowledge diffusion, we are focusing on patent applications.

the identification of several dominant paths is not uncommon (Verspagen, 2007). To identify and isolate the main path, several measures were introduced and discussed in the pertinent literature, typically referred to as ‘traversal counts’. The idea behind these measures is straight forward. They are designed to measure the connectivity between vertices in a graph. Hummon and Doreian (1989) proposed three of these measures:

- node pair projection count (NPPC), reflecting the number of times the link is connecting all possible node pairs. The drawback of this indicator is the overestimation of the points in the middle. Apart from that, here a specific subnetwork with start and end nodes is implied (Liu, Lu, 2012). This algorithm is not applied when analyzing large graphs (Batagelj, 2003).
- search path link count (SPLC), defining the number of times the link appears on all possible search paths, from each node to each end node. The drawback of the indicator can be seen in assigning later links higher values because of them being counted several times (Triulzi, 2011), thus it can be used only when the importance of initial nodes is low (De Nooy et al., 2005);
- search path node pair (SPNP), counting all node pairs, connected by a link. The drawback here is the overestimation of the middle-located nodes (Triulzi, 2011).

These measures were further extended e.g. by forward citation node pair (FCNP), defined as product of forward citations of two connected nodes (Choi, Park, 2009). The most used index, however, was introduced by Batagelj (2003). This measure is called search path count (SPC). It is calculated by counting the number of paths that a particular path connects, trying to eliminate thus disadvantages of SPNP and SPLC. The latter index along with SPNP and SPLC was implemented in main path identification technique of the Pajek software, which allows reducing large graphs. After considering all pros and cons with regards to the research question, we thus apply SPC for main path analysis.

Apart from the difference between several measures, one should also differentiate between global and local main paths. By the local main path method, the highest traversal count is chosen on every step (with “step” meaning new starting point), which does not mean that the local main path receives the highest overall traversal count. The latter is exactly depicted by the global main path (Liu, Lu, 2012). Two specifications may but should not lead to identical results. As Liu and Lu (2012) mention that global path can be used for the identification of the most important flow of thoughts overall, whereas local main path emphasizes the important technologies’ progress over separate periods. Apart from these well-established methods, key-route was proposed by Liu and Lu (2012) as an alternative. Its peculiarity is that a defined number of top links should be included in a corresponding path. As we are trying to follow a step-by-step transition of the one field into another, local main path is put forward.

When it comes to local main path, forward and backward options can be distinguished. As Liu and Lu (2012) mention, forward main path can help to identify from where important patents from the past come, whereas backwards search provides origins of the currently important patents. Apart from that, Park and Magee (2017) combine both methods by introducing genetic backward-forward path analysis, which is based on identifying high persistence patents, which preserve important knowledge of the field. Basing on our research question, we are mostly focusing on the forward main path.

3.5. Text mining applications

As already stated, main path analysis focuses on the structural configuration of a technological field and is still the method of choice for identifying the main directions of the field advancements. However, additional information provided by patent documents are typically neglected. Patent classes and titles of the patents on the main path provides us with general information on the main trends, however, not showing in the detail the topics and keywords, which are included in these patents. Text mining mitigates this limitation by providing a rich tool-box for analyzing patent abstracts and technical descriptions.

As patent documents are well-structured and provide sufficient information about the invention along with their abstracts and titles, text mining techniques are of help here. This method allows identifying patterns in the large scope of data (Yoon and Park, 2003). Text mining includes a large number of procedures and tasks and their implementation may vary from the type of data in use as well as aims of the analysis. As we are interested in the technological change of German pBT over time, analysis of frequencies of specific terms and topics over time is of high significance. In applying text mining we are following below-stated procedure (elaborated based on Yoon and Park, 2003 and Liang and Tan, 2007):

1. Choice of the relevant sample of patent data. In performing this step, we are taking the same patent data, as was used for main path analysis. In this, only German and English titles and abstracts from both focal and citing patent population were filtered out, with abstracts and titles written in German language being translated with the help of DeepL and Google Translate. As some patents' titles or abstracts are missing in PATSTAT and some patents were applied in languages, different from English or German.
2. Data preprocessing and transformation. This step included "tidying" and structuring data. Thus, stop words were removed from both titles and abstracts and stemming of the document was performed. Besides standard stop words, typical words of patent documents that are not describing the content of the

invention, were deleted from text, so that they do not bias results of the analysis (such as “invention”, “current”, “present” etc.⁶) Apart from that, numbers, being irrelevant for our case, were excluded from texts. On the next step words within each title and abstract are separated to make count easier.

3. Identification of relevant patterns and their analysis. At this stage, term frequencies are calculated for the whole dataset as well as over time. Apart from that, trends of occurrence for several keywords identified with the help of main path and bioeconomy-related literature to be relevant or crucial for the transformation, are observed. These terms for our case included: "plant", "nucleic", "fatti", "acid", "transgen", "gene", "yield", "resist", "environment", "agricultur", "econom", "bioreactor", "cell", "tissu", "wound" and "vitro".

Visualization of the results. In order to make results illustrative, they are visualized with the help of R Studio. The relevant codes are elaborated based on the book of Silge and Robinson, 2017.

4. Main path analysis

4.1. Descriptive statistics

To start with, we turn our attention to general characteristics of patent data used for the purpose of this analysis. First of all, we are interested in analyzing the overall trends of patenting over time. Figure 1 presents the distribution for focal and citing patents for the entire observation period. The peak of initial patents, related to pBT, lies between 1998 and 2001, which corresponds to peak in the whole industry. The peak in citations corresponds to 2008 to 2012, with the most often citation lag of 5 to 8 years. This lag can explain the drop in the citation numbers in the last observation years. The citing peak occurs after decline in pBT patents. In the middle of 2000s the number of citing patents is starting to exceed pBT patent numbers. One possible explanation for that may be that citing patents spread in the fields other than pBT.

⁶ Full list of the words: one, "also", "said", "say", "composition", "process", "method", "invention", "current", "present", "relate", "summary", "use", "comprising", "sequence", "formula", "allow", "allowing", "alter", "altering", "additionally", "agent", "agents", "activity", "activities", "usage", "well", "methods", "thereof", "encoding", "least", "absence", "comprises", "novel", "provided", "using", "including", "provide", "producing", "containing", "particularly", "preferably", "represent", "constructs", "useful", "optionally", "provided", "interest"

From biotech to bioeconomy

New empirical evidence on the technological transition to plant-based bioeconomy based on patent data

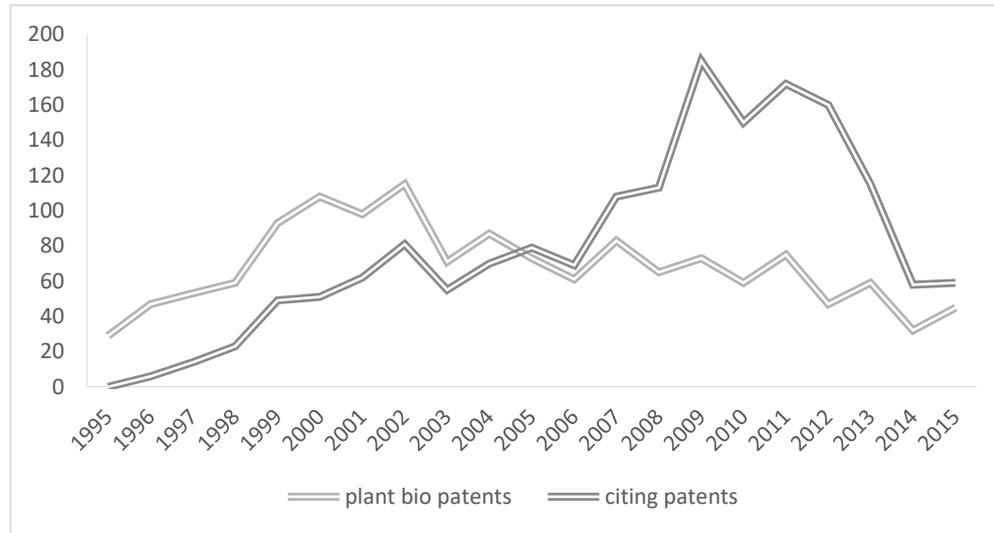


Figure 1: Focal and citing patents over years

To gain in-depth insights into the composition and technological origin of citing patents, we checked, whether the citing patents originate from biotechnology. Thus IPC classes, assigned to citing patents, were checked for being biotechnology-related (list of biotechnology-related patent classes was presented in 3.2). Patents were classified as biotechnology patents whenever at least one biotech-related technology class appeared on the document. The results of the analysis are shown in figure 2.

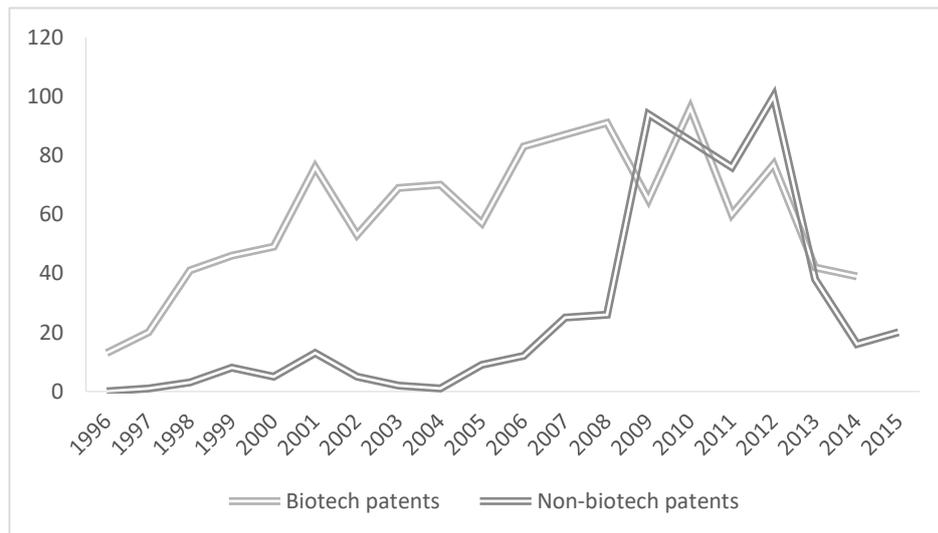


Figure 2: Biotechnology vs. non-biotechnology patents in citing population over years

At the onset of the observation period, the large majority of citing patents is related to biotechnology. However, after the middle of the 2000s the picture is starting to change. For biotechnology-related patent citations, we observe a volatile trend line while non-biotech patent citations increase continuously. In 2008 the number of non-biotech patent citations exceeds the number of biotech patent citations for the first time. The reason for

that may lie in expected increased competition, coming from regulation change in the field of pharmaceuticals (Vermeulen, 2018). Thus, many pharma patents, which for sure can cite biotech ones, were applied at that time. The dominance of non-biotech patent citations in recent years implies that other technologies increasingly affected biotechnology in Germany. Figure 3 provides a fine-grained exploration of technological orientation for all non-biotechnology patent citations (figure 3).

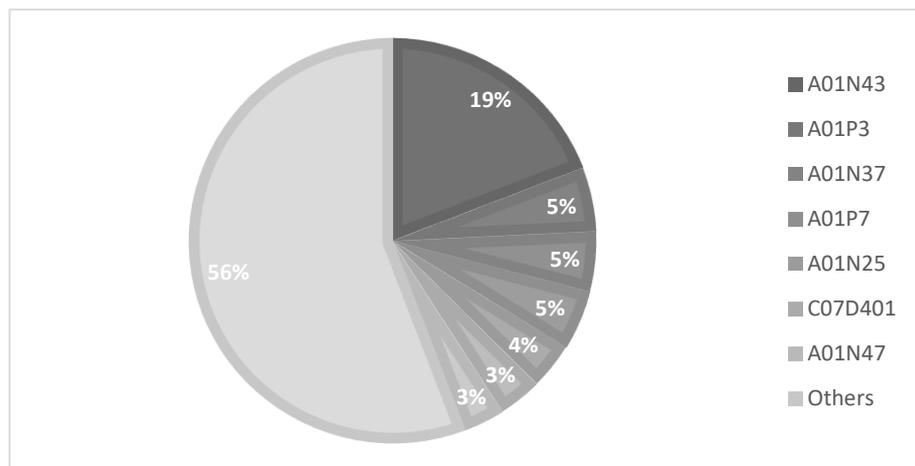


Figure 3: The most popular technological classes within non-biotechnology citing patents¹

Interestingly, many of these technological classes relate to biocides and plant growth regulators (A01N43, A01N25, A01N47), i.e. classes which are highly relevant for increasing plant production and increasing yield. In other words, we can observe already at this stage of our exploration patterns which mirror a transition of biotechnology towards bioeconomy.

Furthermore, we looked at the applicants, that contribute to the largest number of patents within the population. These actors include: BASF Plant Science GmbH (about 22% of patents without accounting for subsidiaries). Apart from that, active was Bayer Cropscience AG (about 4,5% of patents), Max Planck Society (3,5% of patents) and Fraunhofer Society (about 3,2% of patents). When it comes to different types of SMEs, the most active here appeared to be Icon Genetics AG, which is now part of the Denka Company (more than 1,5% of patents) and Greenovation Biotech GmbH, which is now part of KAWA Beteiligungen GmbH (around 1% of patents). In addition, several new actors appear within the citing sample. Thus, Sungene & Company KGaA, a subsidiary of BASF Plant Science, has applied for more than 1% of patents or large corporation Suedzucker (0,5% of patents) were put forward along with SMEs Greenovation Biotech, Nomad Bioscience GmbH or Zellwerk GmbH (both around 0,1-0,7% of patents).

Summing up, although the majority of patents were applied by big corporations, giants in the field of life sciences, some SMEs also put forward a notable number of

important patents. In the next session, the main path within technological field will be defined and analyzed.

4.2. General characteristics of the directed patent citation graph

Insight from our descriptive analysis indicates a shift in pBT towards more sustainability-oriented topics. To substantiate this initial observation, we explore now the overall patent citation graph for the German pBT.

Hence, we first construct a directed patent citation graph. Here, nodes are initial pBT patents identified, which are connected to citing patents' nodes via citation tie. In general, the whole patent citation graph is constructed of 2138 nodes and 12198 ties. Not all initially identified patents were included in this graph. 581 patent families (40,5% of the overall population) did not receive any citation, thus being isolates. Figure 4 illustrates the general citation graph, which, however, does not provide fine-grained information on the field structure. Nonetheless, it reveals some insightful patterns. The exploration shows one big component (1204 nodes, 56,31% of the whole graph), followed by a second relatively big component (303 nodes, 14,17% of the whole graph) and several smaller components (which in sum constitute 29,52% of the whole graph). Thus, one can state, that there are at least two large completely disconnected technological sub-fields, which dominated plant biotechnology in Germany over the last 20 years. Therefore, further main path analysis would focus on these two big components.

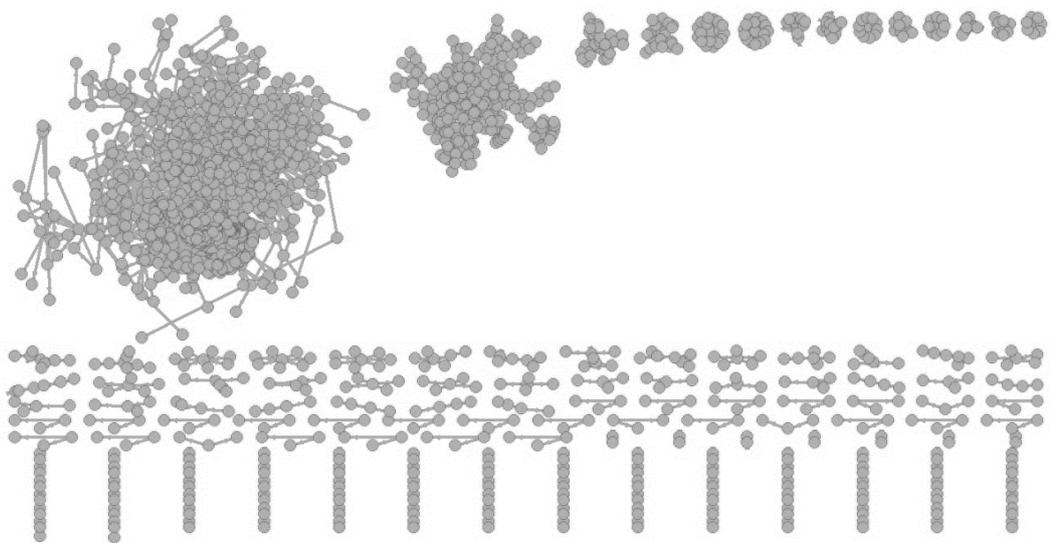


Figure 4: Patent citation graph in German biotechnology citation

Table 1 presents a selected number of key measures that allow us to draw conclusions about how the patents and their citations are concentrated and centralized.

Table 1: General measures

MEASURE	VALUE
DEGREE CENTRALIZATION	0.148
OUT-CENTRAL	0.148
IN-CENTRAL	0.034
AVERAGE DEGREE	4.952
DENSITY	0.002
CONNECTEDNESS	0.003
BREADTH	0.997

The overall patents citation graph is characterized by a rather low density which is typical for larger graphs. The average degree, defined as the average number of ties a node has, is almost five, meaning that outgoing ties are more frequent than incoming.

The node level measure provides insights on the role of individual patents. In this, the looking at betweenness centrality is important. This measure reflects the number of times, the node lies on the shortest path between two other nodes (Borgatti et al., 2018). Thus, we can identify patents, which are working as the connectors in the citation graph, by bringing together two other inventions. These are 26007459- Monocotyledon plant cells and plants which synthesise modified starch; 38462357 - Process for the production of a fine chemical; 7867091 - New enzyme with starch synthase activity, useful for producing starch for foods and packaging materials; 42111699 - Soybean event 127 and methods related thereto. Thus, topics of modified plants and fine chemicals appear to be important within the scope of biotechnology.

Considering two largest components and frequencies of patenting within them over time (Figure 5), one can see that the first (largest) component gains importance at the end of 2000s – beginning of 2010s. On the contrary, the importance of the second component remains relatively stable over time with a slight increase at the end of the 2000s. By combining this data with patenting and citing statistics, one can state, that many of these patents are already the ones, coming outside of pBT. Thus, the investigation of main path may bring important insights here regarding where the field goes and to which extent this direction corresponds to a sustainability shift.

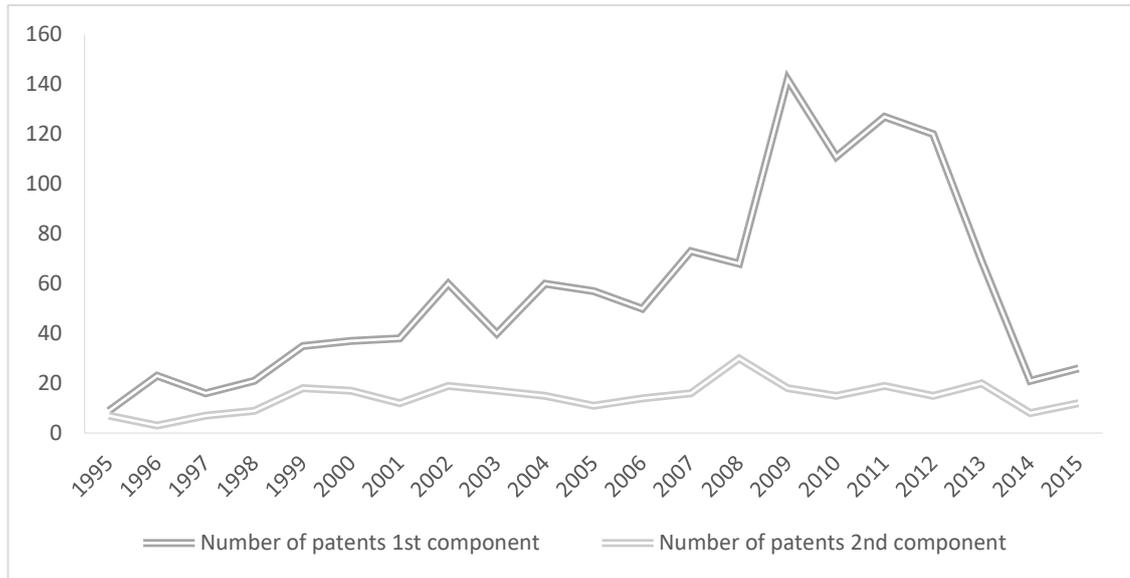


Figure 5: Trend of number of patents across components over time

Thus, we can generally see two main topics, that developed within the field. But what subjects exactly received the highest attention within the field and how did it develop over time? To answer this question main path analysis is applied in the next section. It allows us to zoom into the technological core development within one well-specified sub-graph.

4.3. Main path specification

Now, we are interested in identifying the most influential trajectories along this patent citation graph. As already stated in part 3.4, we are using the search path count measure in order to identify the main path. Along with the aim of this paper, we are focusing on forward local main path⁷. As two big components could be found, main trajectories have to be found for both of them. In that case, we are expecting to have two separate topics, representing each of the components. Grouping of the main path patents over time accompanied by an evaluation of patents' contents, allows us to identify three sequential phases of development inherent in each of two components. Figure 6 contains a forward local main path whereas table 2 contains titles of these applications.

⁷ Backward main path as well as global main paths are also presented in the appendix B. As can be seen there, in general global main path corresponds to the identified forward local main path, whereas backward local main path is different. However, the topic of the path, as well as stages of its development are always identical.

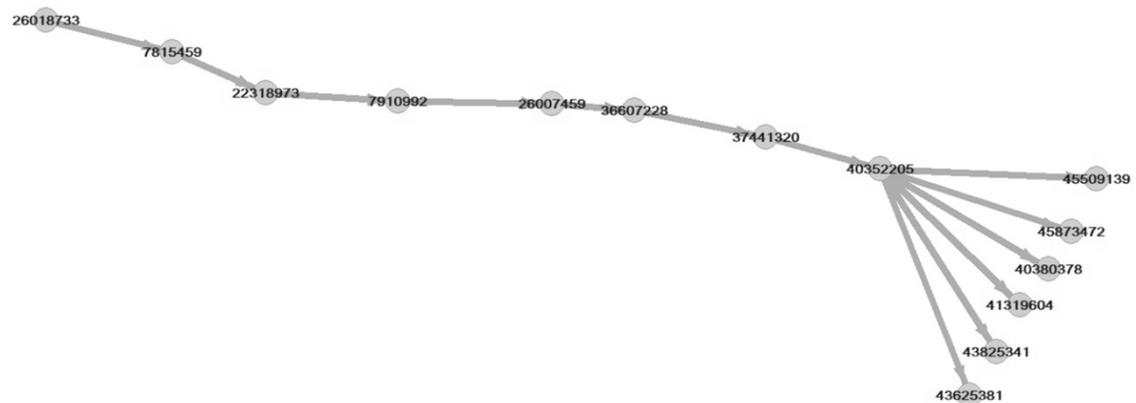


Figure 6: Forward local main path

Table 2: Titles of patents on forward local main path

PATENT FAMILY	YEAR ⁸	TITLE
26018733	1995	Process for the production of a modified starch
7815459	1996	Neue Nucleinsäuremoleküle aus Mais und ihre Verwendung zur Herstellung einer modifizierten Stärke (<i>English: Nucleic acid molecules from corn and their use in the preparation of a modified starch</i>)
22318973	1998	Nucleic acid molecules from rice encoding an r1 protein and their use for the production of modified starch
7910992	1999	Nukleinsäuremoleküle aus Weizen, transgene Pflanzenzellen und Pflanzen und deren Verwendung für die Herstellung modifizierter Stärke (<i>English: Nucleic acid molecules from wheat, transgenic plant cells and plants and their use in the production of modified starch</i>)
26007459	2000	Monocotyledon plant cells and plants which synthesise modified starch
36607228	2004	Plants with increased activity of a starch phosphorylating enzyme
37441320	2006	Genetically modified plants synthesizing starch with increased swelling power
40352205	2009	Method for improved utilization of the production potential of transgenic plants introduction
40380378	2009	Use of carboxamides on cultivated plants
41319604	2009	Use of succinate dehydrogenase inhibitors for increasing the resistance of plants or parts of plants to abiotic stress
43625381	2010	Use of fluopyram for controlling nematodes in nematode resistant crops
43825341	2010	Active ingredient combinations comprising pyridylethylbenzamides and other active ingredients
45509139	2010	Use of fluopyram for controlling nematodes in crops and for increasing yield
45873472	2010	Use of active ingredients for controlling nematodes in nematode-resistant crops

⁸ Here earliest filing year is included.

By looking at application titles and abstracts, one can distinguish several stages:

- **Stage one. End of 1990s-beginning of the 2000s. Researching on nucleic acid molecules and modified starch**

At the beginning of trajectory more general patents dealing with the description of nucleic acid molecules as well as methods of starch modification and production of transgenic plants. Thus, it can be stated that companies in Germany only started working on the topics of transgenic plants. Patents relate to maize, rice and wheat proteins and can thus be mostly related to green and not to white or red biotechnology.

- **Stage two. Middle of the 2000s. Evoking extraordinary properties.**

These patents start to work out how the developed genetically modified plants will be used in order to increase specific properties, e.g. “increasing swelling power” or “increase in the activity of a starch-phosphorylating OK1 protein”, being thus the middle ground between general initial research on nucleic acids and specific environment-related research.

- **Stage three. End of the 2000s-2010s. Increasing yield, tolerance and resistance**

These patents come from just investigating specific plant properties towards their usage in specific climatic conditions, e.g. “drought” or increasing plant tolerance towards “abiotic stress” and increasing the production potential of a plant. Thus, this stage can be seen as the answer to the calls for nutrition provision and to harder climatic conditions.

As for actors of the field, it can be seen, that it is created by big corporations (BASF, Aventis) and their subsidiaries. No SMEs or research institutions can be found along the path. Therefore, at this stage big players can be seen as technological agenda setters within the biggest biotechnology patent citations’ component⁹.

Thus, by just looking at the patents and actors staying behind these applications, several conclusions can be made. First of all, the main topic of the biggest component concerns the topic of genetic plants. This topic is rather controversial when talking about sustainability issues and cannot be included into bioeconomy according to some streams of the views, which point out ethical issues concerning the field. However, here the development of the path is of a high interest. Whereas initial patents reflect rather basic research on the topic, which is further enriched by the applications of created genetically modified plants, later patents directly address the creation of more climatically-prone

⁹ Apart from their power, these firms also possess financial resources, which make high amount of patents possible. They are also having higher number of large projects, which include several subsequent (therefore citing) patents.

plants, and this issue is without a doubt quite topical. Secondly, as the patents are created by big corporations, it can be suggested, that, although being registered also internationally/in Germany, these inventions are not directly oriented towards the German market. Despite this fact, the topic of genetically modified plants, set by big corporations, plays an important role for the German technology space.

Thus, although giving the idea about the development of trajectory within the timeframe, at this point the question of transition towards bioeconomy based on the patent data cannot be answered univocally. Thus, we look at the second largest component within a patent citation graph to complement the picture gained so far.

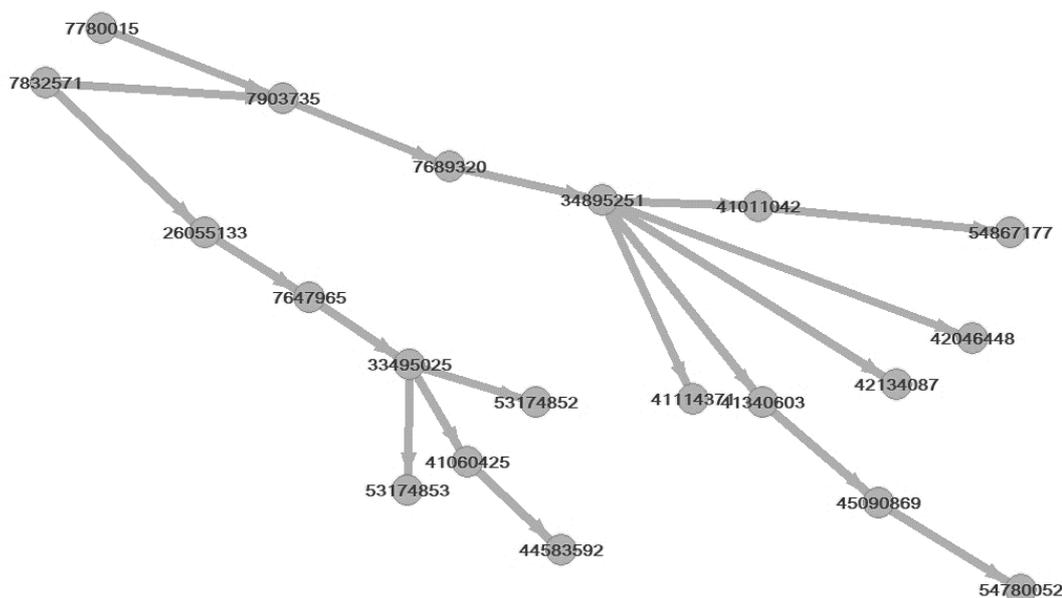


Figure 7: Forward local main path, 2nd component

Table 3: Titles of patents on forward local main path, 2nd component

PATENT FAMILY	YEAR	TITLE
7780015	1995	Cell cultivation process and assembly incorporates spirally-wound substrate belt
7832571	1997	Cell culture apparatus for three-dimensional cell or tissue cultivation
7903735	1999	Method for populating substrates with biological cells and populating devices that can be used therefor
26055133	1999	System for the cultivation of cells or tissue has a culture container with capillary netting or matrix material around the cultivation zone to be fed with a consistent and gas-free culture medium
7647965	2000	Bioreactor growing tissue or complete organ implants by assembly of differing cell types in specific forms and patterns, includes chambers with cavities of differing shape
7689320	2001	Vorrichtung zur Druckperfusion für das Züchten und/oder für das Behandeln von Zellen (<i>English: Device for pressure perfusion for culturing and/or treating cells</i>)
33495025	2003	Perfusionseinheit und Perfusionsstation zur Hautwundbehandlung sowie dessen Verwendung (<i>English: Perfusion unit and perfusion station for treating skin wounds, and its use</i>)
34895251	2004	Inventive bioreactors and bioreactor systems
41011042	2008	Perfusable bioreactor for the production and/or cultivation of a human or animal blood vessel and/or a human or animal tissue
41060425	2008	Microbioreactor and microtiter plate comprising a plurality of microbioreactors
41114371	2008	Perfusable bioreactor for the production of human or animal tissues
41340603	2008	Bioreactor and method for cultivating cells and tissues
42046448	2008	Device for physiological, dynamic in-vitro cell stretching
42134087	2008	Method and device for the stimulation of cells
44583592	2010	Microfluidic system and method for the operation thereof
45090869	2010	Perfusion device
54780052	2014	Device and method for identification of microorganisms
53174852	2015	Wound care system
53174853	2015	Wound dressing system
54867177	2015	Modulares bioreaktorsystem (<i>English: Modular bioreactor system</i>)

Figure 7 displays the main path for the second largest component. The main path is divided into two streams. The first stream is purely related to medical research, in particular with wound systems and in-vitro devices. The patterns identified here point to the close relatedness between biotechnology and medical research.

The second stream is particularly interesting with regards to the aim of this paper. Along the trajectory we can find patents related to the topic such as operations on cells and tissues as well as bioreactors, each of which closely linked to medicine research. In most of the patents it is stated, that provided inventions can be important for oncology, obesity and wound healing. Apart from that, tissue cultivation may be important for the production of plants with new properties. Thus, this path provides a good example of

biological material usage for different industries, in this case medicine. This insight is in line with SDG3: Good health and well-being (Fraunhofer, 2018).

Most interestingly, the technological advancements constituting the technological path displayed in Figure 7 also follow the stages identified for the largest component. General investigations of cell structure apparatus and bioreactors at the end of the 1990s-beginning of the 2000s were further substituted by investigation of specific bioreactor systems and specific use of inventions identified before. At the end of the 2000s these inventions were used for specific purposes – tissue engineering, cells stimulation. Thus, the same aims of increased yield or plant resistance here can be achieved with the help of alternative to transgenic plants methods, which means higher relevance for bioeconomy. Also important to mention is that main actors of the patents on this main path are not corporations, but universities, individuals and smaller firms. As consequence, here the topics of transgenic plants do not have such relevance as for the large corporations' patents. What seems to be important for the German biotechnology space, is the development of medical research-related biotechnology into advanced usage of biomass for the purposes of healing of specific relevant illnesses. However, apart from that, tissue and cell engineering may be of high importance for the increased plant resistance.

In summary, main path analysis allows us to detect and explore the technological core within pBT. The analysis can be applied to the largest components (but also other sub-graphs) of the patent citation graph. We found technological paths in the two largest components follow the same development stages, starting with the basic research, followed by application-based research, and finally sustainability-related inventions.

4.4. Exploration of patent space with text mining tools

In order to detect topics, which occur in the patents at most, abstracts are analyzed using text mining tools implemented in R¹⁰. In the first place, we are interested in analyzing the whole patent space in order to find out particular terms that are of major importance for the industry and their trends over time. Thus, all identified 2083 patent families were analyzed (titles + abstracts). The most frequent terms (more than 1000 occurrences) are present in figure 8¹¹.

The aim of this analytical exercise is to conduct a content-based exploitation using the entire patent population to identify the most relevant topics and their development over time. Focus on two biggest components allows us to separate most influential technological trends within the industry, based on the limited amount of

¹⁰ <https://www.rstudio.com/>

¹¹ Correspondent wordcloud can be found in appendix C.

From biotech to bioeconomy
 New empirical evidence on the technological transition to plant-based bioeconomy based on patent data

patents. To be sure, that our claim regarding trends towards increased plant tolerance and resistance as well as tissue engineering techniques, are also relevant for the industry in general, we have to take periphery patents into account as well.

Figure 8 presents the most frequently appearing terms along with patent population. Our exploration indicates a high relevance of the topic ‘transgenic plants’. This is supported by the appearance of such terms as “dna”, “transgen”, “express”. Apart from that, “yield”, “enhanc...”¹², “increas...” also appear to be among the terms with high frequencies, which outlines the overall importance of the patents along the first component’s main path. No terms related to the topic of tissue engineering or cells cultivation are present. However, it is interesting to look at the appearance of these terms over time and to see, whether their relative importance has changed.

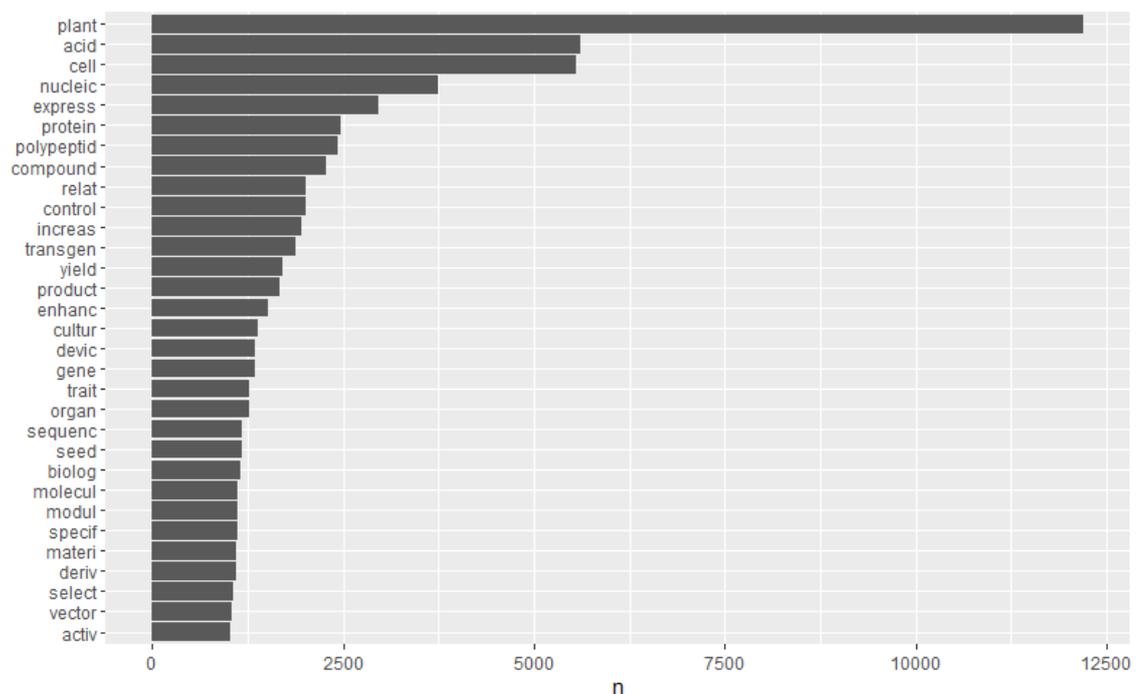


Figure 8: Most frequent terms, all patent families

To do so, we look at the terms, relevant for topics of our interest (two main paths topics) as well as sustainability-related terms. The latter would favor the development of the field towards bioeconomy. Thus, we choose the terms which appeared to be important after descriptive and main path analysis, and analyzed changes of their relevance over time. The term "plant" points to the relevance of selected patents for pBT. Such terms as "nucleic", "fatti", "acid", "transgen" and "gene" are relevant for the main path coming from the largest component. Terms "yield", "resist", "environment" and "agricultur", which play significant role along later stages of the path. The term "econom" may reflect importance of economic view on bioeconomy, which outlines the necessity

¹² Terms obtained as the result of stemming.

of inventions to be economically relevant. Terms "bioreactor", "cell", "tissu", "wound" and "vitro" represent two branches of the second component's main path. Figure 9 presents frequencies of these terms over time. All frequencies presented are shown in relative values by dividing the number of patents, in which a specific term occurs in a specific year, through all patents, that were filed in that year.

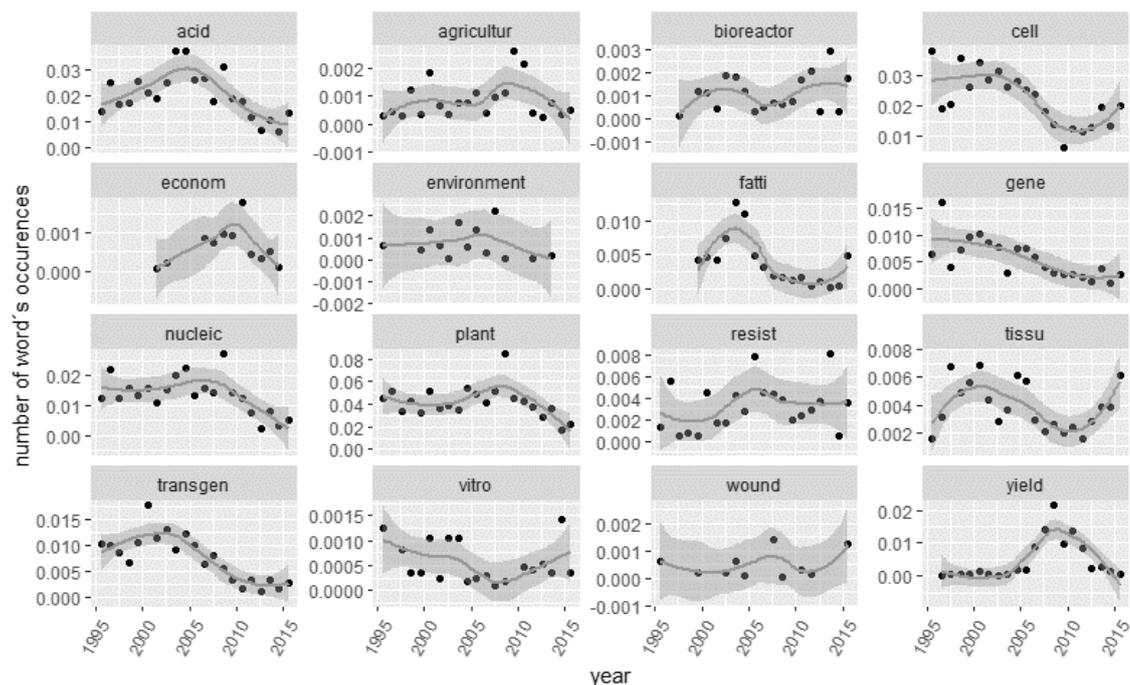


Figure 9: Frequencies of specific terms over time

One important conclusion is the decrease of all terms, connected to the transgenic plants, thus, reflecting main path of the first (largest component). Terms "Nucleic", "fatti", "acid", "transgen" and "gene" lose importance in the mid-2000s. Additionally, the topic of yield increase (terms "yield", "resist", "environment" and "agricultur"), gain in importance by the end of 2000s-beginning of 2010s. All terms, related to this group ("bioreactor", "cell", "tissu", "wound" and "vitro"), are the only ones that increased in relevant terms over last period.

The patterns described above provide us with some interesting insights, which become not directly apparent when conducting a main path analysis. First, the main component related to the topic of transgenic plants, seems to lose its dominance over time. This happens, although the main component in general consists of more than half of the overall graph's nodes. The only sub-topics, which are still developed within this path, relate to 'increased yield' and 'tolerance of the plants'. However, even these themes did not receive further attention in Germany during the last several years. Second, research institutes and small firms gained in importance, which is reflected in the relative increase of attention towards the topic of cell cultivation. Derivatives, produced along this

path, may present real importance for the future of medicine and in its broad sense to bioeconomy.

Thus, the overall analysis shows that pBT in Germany changed its research orientation. At the onset, we uncover a strong research focus in pBT on transgenic plants. The patent documents substantiating this finding are held by a small number of very large corporations. Now the field gains diversity by putting forward such topics as sustainability, resistance towards hard climatic conditions, healing with the help of biological material. Main path analysis revealed the existence of two dominant components within the field: the largest component related to transgenic research, and the smaller one, related to cell cultivation. These different paths, however, follow the same stages of development: from general investigations through first applications towards evoking of new extraordinary properties of research subjects. Both paths seem to develop in line with SDGs, which suggests their importance for the German bioeconomy.

5. Discussion and conclusion

This paper provides new empirical evidence and insights on technological field evolution by exploring the transition from biotechnology innovations towards the emergence of bioeconomy over time. Main path analysis can bring important insights into the evolution of a particular field by pointing out main trajectories. The results of this paper may not only be used for the particular field of pBT but also enrich main path methodology by distinguishing the differences in the technological path that specific categories of firms (here SMEs) main take. Apart from that, the paper adds on the discussion on transgenic plants and inventions regarding them made by German firms.

Main path analysis, extended by content analysis, has shown, that there exist two main components of patents, orchestrated by different kind of actors. The larger component, probably with orientation on international markets, has the main topic of transgenic plants research. It is created by a small number of large corporations. Smaller component is led by smaller firms and research institutions. It relates to the topic of cells cultivation and bioreactors and applies to the usage of biomaterials in medical research. Over time the second topic seems to gain more relative weight. All in all, it can be said, that the patents of the firms are moving towards answering the challenges, voiced by Sustainable Development Goals (SDGs), in particular goal two “Zero hunger” and goal 3 “Good health and well-being” as well as National Research Strategy BioEconomy 2030, which sees Securing global nutrition as one of the most important actions. The paper thus provides important implications to policy-makers regarding the execution of strategies action plans. In addition to that, it also gives insights into the differences in patenting patterns of SMEs and purely German research institutions and big corporations. SMEs thus may profit from engaging in cooperation with research

institution as well as from finding niche topics within biotechnology spectrum. The paper also provides insights to the discussion on technological trajectories. It shows the importance of going beyond main component, especially in the case of the presence of several big players in the field. Apart from that, it outlines value added from conducting content analysis on the whole patent population, which helps to identify relevant trends.

Nevertheless, there still remains enough room for future research. First, several methods to define technological classes related to pBT can be used to prove the robustness of results. As noted in Shubbak (2018) apart from combined keywords and IPC analysis, firm population or expert advice can be used. Thus, a broader picture of a not yet officially defined subfield can be brought. Furthermore, although the research questions stated in the article, already limits itself to a specific subfield within biotechnology, pBT may still be too broad for the main path identification. Additional research may be performed, which limits itself to a specific technology within pBT, e.g. tissue culture. Therefore, more specific trajectories and important patents may be observed. As this paper aims at providing the overall picture of the field, no deeper analysis of each patent class was provided, however, it may serve as an interesting implication for those, willing to go deeper in the subfield analysis. Besides, the paper is limited to the usage of patent data as the indicator of innovative activity of the field. It is also stated, that for the field of biotechnology in general patents can serve as such an indicator. However, the complementary usage of alternative indicators and data sources, e.g. citations of scientific literature, may enrich the picture of pBT. Apart from that, important limitation of the paper lies in the absence of commercial value of patents in the analysis. The presence of this data would enable realizing whether the patents belonging to second component (medical and more bio-related) bring the same end-value as DNA-related ones. Also worth mentioning are the possible extensions of text mining tools, including clustering of given abstracts into topics and their detailed analysis.

References

- Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The economic journal*, 99(394), 116-131.
- Barberá-Tomás, D., Jiménez-Sáez, F., & Castelló-Molina, I. (2011). Mapping the importance of the real world: The validity of connectivity analysis of patent citations networks. *Research Policy*, 40(3), 473-486.
- Batagelj, V. (2003): Efficient Algorithms for Citation Network Analysis. *arXiv preprint cs/0309023*
- Borgatti, S.P., Everett, M.G., Johnson, J.C. (2018). *Analyzing Social Networks*. 2nd Edition. SAGE Publications Ltd.
- Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional studies*, 39(1), 61-74
- Bugge, M.; Hansen, T.; Klitkou, A. (2016): What is the Bioeconomy? A Review of the Literature. *Sustainability*, Vol. 8 (7), 1-22.
- Choi, C.; Park, Y. (2009): Monitoring the organic structure of technology based on the patent development paths. *Technological Forecasting & Social Change*, Vol. 76, 754-768.
- David, P. A. (1985). Clio and the Economics of QWERTY. *The American economic review*, 75(2), 332-337.
- De Nooy, W., Mrvar, A., Batagelj, V., 2005. *Exploratory Social Network Analysis with Pajek*. Cambridge University Press, New York.
- Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research policy*, 11(3), 147-162.
- Dosi, G., Freeman, C., Nelson, R., Silverberg, G., & Soete, L. (1988). *Technical change and economic theory*. Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy.
- Dosi, G., & Grazzi, M. (2010). On the nature of technologies: knowledge, procedures, artifacts and production inputs. *Cambridge Journal of Economics*, 173-184.
- Dosi, G., & Nelson, R. R. (1994). An introduction to evolutionary theories in economics. *Journal of evolutionary economics*, 4(3), 153-172.
- European Patent Office (EPO) (2017). *Patent families at the EPO*
- Federal Ministry of Education and Research (BMBF) (2011): *National Research Strategy BioEconomy 2030 – Our Route towards a biobased economy*.
- Federal Ministry of Education and Research (BMBF) (2014): *Bioökonomie in Deutschland Chancen für eine biobasierte und nachhaltige Zukunft*.
- Feldman, M. P. (2000). Where science comes to life: university bioscience, commercial spin-offs, and regional economic development. *Journal of Comparative Policy Analysis: research and practice*, 2(3), 345-361.
- Fontana, R.; Nuvolari, A.; Verspagen, B. (2009): Mapping technological trajectories as patent citation networks. An application to data communication standards. *Economics of Innovation and New Technology*, Vol. 18 (4), 311-336.
- Hanusch, H., & Pyka, A. (Eds.). (2007). *Elgar companion to neo-Schumpeterian economics*. Edward Elgar Publishing.

From biotech to bioeconomy
 New empirical evidence on the technological transition to plant-based bioeconomy based on patent data

- Häussler, C. (2007). Proactive versus reactive M&A activities in the biotechnology industry. *The Journal of High Technology Management Research*, 17(2), 109-123.
- Hummon, N.; Doreian, P. (1989): Connectivity in a citation network: the development of DNA theory. *Social Networks*, Vol 11 (1), 39-63.
- Kaiser, R.; Prange, H. (2003): The reconfiguration of National Innovation Systems – the example of German biotechnology. *Research Policy*, Vol. 33 (3), 395-408.
- Liang, Y., & Tan, R. (2007). A text-mining-based patent analysis in product innovative process. In *Trends in computer aided innovation* (pp. 89-96). Springer, Boston, MA.
- Liu, J. S.; Lu, L.Y.Y. (2012): An Integrated Approach for Main Path Analysis: Development of the Hirsch Index as an Example. *Journal of the American Society for Information Science and Technology*, Vol. 63 (3), 529-542.
- McCormick, K.; Kautto, N. (2013): The Bioeconomy in Europe: An Overview. *Sustainability*, Vol. 5 (6), 2589-2608.
- Mina, A.; Ramlogan, R.; Tampubolon, G.; Metcalfe, J.S. (2007): Mapping evolutionary trajectories: Application to the growth and transformation of medical knowledge. *Research Policy*, Vol. 36 (5), 789-806.
- Nelson, R. R., & Winter, S. G. (1974). Neoclassical vs. evolutionary theories of economic growth: critique and prospectus. *The Economic Journal*, 84(336), 886-905.
- Nelson, R. R., & Winter, S. G. (1982), *An Evolutionary Theory of Economic Change*. Belknap Press/Harvard University Press: Cambridge.
- No, H. J., & Park, Y. (2010). Trajectory patterns of technology fusion: Trend analysis and taxonomical grouping in nanobiotechnology. *Technological Forecasting and Social Change*, 77(1), 63-75.
- Nomaler, Ö.; Verspagen, B. (2016): River deep, mountain high: of long run knowledge trajectories within and between innovation clusters. *Journal of Economic Geography*, Vol. 16 (6), 1259-1278.
- Parayil, G. (2003): Mapping technological trajectories of the Green Revolution and the Gene Revolution from modernization to globalization. *Research Policy*, Vol. 32 (6), 971-990.
- Parisi, C.; Rodriguez-Cerezo, E.; Thangaraj, H. (2012): Analysing patent landscapes in plant biotechnology and new plant breeding techniques. *Transgenic Research*, Vol. 22 (1), 15-29.
- Park, H.; Magee, C. L. (2017): Tracing Technological Development Trajectories: A Genetic Knowledge Persistence-Based Main Path Approach. *PLoS One*, Vol. 12 (1), 1-18.
- Schmoch, U. (2008). Concept of a technology classification for country comparisons, final report to the World Intellectual Property Organisation (WIPO). *Fraunhofer Institute for Systems and Innovation Research, Karlsruhe*.
- Shubbak, M. (2018). *Forging Ahead: Technology Development & Emerging Economies* (Doctoral dissertation, Universität Bremen).
- Silge, J., & Robinson, D. (2017). Text mining with R: A tidy approach. " *O'Reilly Media, Inc.* "
- Székács, A. (2017): Environmental and Ecological Aspects in the Overall Assessment of Bioeconomy. *Journal of Agricultural and Environmental Ethics*, Vol. 30 (1), 153-170.
- Triulzi, G. (2011): Technological Trajectories and the Role of (Former) Developing Countries' Firms in the Semiconductor Industry: a Patent Citation Analysis over the last 50 years.
- Van Beuzekom, B., & Arundel, A. (OECD) (2009). *OECD Biotechnology statistics-2009*.

From biotech to bioeconomy
New empirical evidence on the technological transition to plant-based bioeconomy based on patent data

-
- Vermeulen, B. (2018). Geographical dynamics of knowledge flows: descriptive statistics on inventor network distance and patent citation graphs in the pharmaceutical industry. *International Journal of Computational Economics and Econometrics*, 8(3-4), 301-324.
- Verspagen, B. (2007): Mapping technological trajectories as patent citation networks: a study on the history of fuel cell research. *Advances in Complex Systems*, Vol. 10 (1), 93-115.
- Witt, U. (2008). *Evolutionary economics* (pp. 1873-1879). Palgrave Macmillan UK.
- Yoon, B., & Park, Y. (2007). Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information. *IEEE Transactions on Engineering Management*, 54(3), 588-599.

Appendix A: Chosen technology classes, IPC

Plant biotechnology patent classes

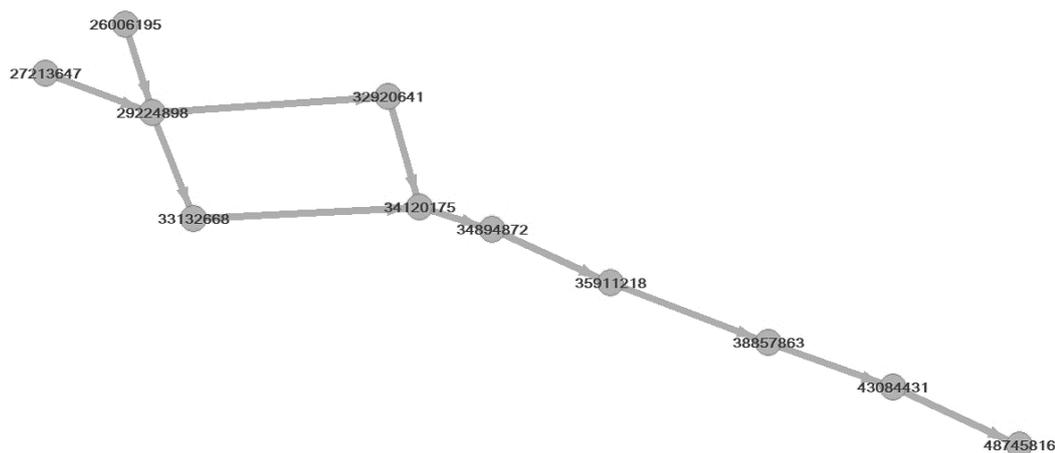
PATENT CLASS	TITLE
A01H 1*	Processes for modifying genotypes (A01H 4/00 takes precedence)
A01H 4*	Plant reproduction by tissue culture techniques
A61K 38/56	Medicinal preparations containing peptides (peptides containing beta-lactam rings A61K 31/00; cyclic dipeptides not having in their molecule any other peptide link than those which form their ring, e.g. piperazine-2,5-diones, A61K 31/00; ergoline-based peptides A61K 31/48; containing macromolecular compounds having statistically distributed amino acid units A61K 31/74; medicinal preparations containing antigens or antibodies A61K 39/00; medicinal preparations characterised by the non-active ingredients, e.g. peptides as drug carriers, A61K 47/00) ... from plants
C07K 4/10	Peptides having up to 20 amino acids in an undefined or only partially defined sequence; Derivatives thereof... from plants
C07K 14/415	Peptides having more than 20 amino acids; Gastrins; Somatostatins; Melanotropins; Derivatives thereof... from plants
C07K 16/16	Immunoglobulins, e.g. monoclonal or polyclonal antibodies... against material from plants
C12M 3*	Tissue, human, animal or plant cell, or virus culture apparatus
C12N 5*	Undifferentiated human, animal or plant cells, e.g. cell lines; Tissues; Cultivation or maintenance thereof; Culture media therefor (plant reproduction by tissue culture techniques A01H 4/00)
C12N 9/32	Enzymes, e.g. ligases (6.); Proenzymes; Compositions thereof (preparations containing enzymes for cleaning teeth A61K 8/66, A61Q 11/00; medicinal preparations containing enzymes or proenzymes A61K 38/43; enzyme containing detergent compositions C11D); Processes for preparing, activating, inhibiting, separating, or purifying enzymes... Alpha-amylase from plant source
C12N 15/05	Mutation or genetic engineering; DNA or RNA concerning genetic engineering, vectors, e.g. plasmids, or their isolation, preparation or purification; Use of hosts therefor (mutants or genetically engineered microorganisms C12N 1/00, C12N 5/00, C12N 7/00; new plants A01H; plant reproduction by tissue culture techniques A01H 4/00; new animals A01K 67/00; use of medicinal preparations containing genetic material which is inserted into cells of the living body to treat genetic diseases, gene therapy A61K 48/00; peptides in general C07K... Plant cells
C12N 15/29	... genes encoding plant proteins, e.g. thaumatin
C12N 15/82	... for plant cells
C12Q 1/6895	Measuring or testing processes involving enzymes, nucleic acids or microorganisms (measuring or testing apparatus with condition measuring or sensing means, e.g. colony counters, C12M 1/34); Compositions therefor; Processes of preparing such compositions... for plants, fungi or algae

The most popular patent classes within non-biotechnology citing patents

PATENT CLASS	TITLE
A01N 43	Biocides, pest repellants or attractants, or plant growth regulators containing heterocyclic compounds (containing cyclic anhydrides, cyclic imides A01N 37/00; containing compounds of the formula fig11, containing only one heterocyclic ring, wherein $m \geq 1$ and $n \geq 0$ and fig12 is unsubstituted or alkylsubstituted pyrrolidine, piperidine, morpholine, thiomorpholine, piperazine or a polymethyleneimine with four or more CH ₂ groups A01N 33/00-A01N 41/12; containing cyclopropane carboxylic acids or derivatives thereof, e.g. esters having heterocyclic rings, A01N 53/00)
A01P 3	Fungicides
A01N 37	Biocides, pest repellants or attractants, or plant growth regulators containing organic compounds containing a carbon atom having three bonds to hetero atoms with at the most two bonds to halogen, e.g. carboxylic acids (containing cyclopropane carboxylic acids or derivatives thereof, e.g. cyclopropane carboxylic acid nitriles, A01N 53/00)
A01P 7	Arthropodocides
A01N 25	Biocides, pest repellants or attractants, or plant growth regulators, characterised by their forms, or by their non-active ingredients or by their methods of application (fungicidal, bactericidal, insecticidal, disinfecting or antiseptic paper D21H); Substances for reducing the noxious effect of the active ingredients to organisms other than pests
C07D 401	Heterocyclic compounds containing two or more hetero rings, having nitrogen atoms as the only ring hetero atoms, at least one ring being a six-membered ring with only one nitrogen atom
A01N 47	Biocides, pest repellants or attractants, or plant growth regulators containing organic compounds containing a carbon atom not being member of a ring and having no bond to a carbon or hydrogen atom, e.g. derivatives of carbonic acid (carbon tetrahalides A01N 29/02)

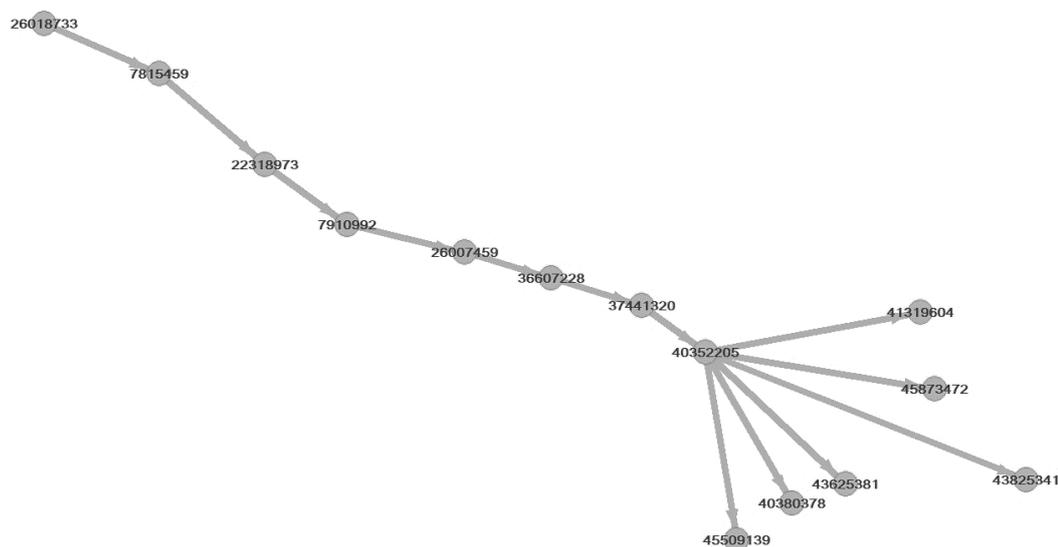
Appendix B: Additional main path explorations

Backward local main path, component 1



PATENT FAMILY	YEAR	TITLE
26006195	1999	Plants expressing delta 6-desaturase genes and oils from these plants containing pufas and method for producing unsaturated fatty acids
27213647	2000	Novel elongase gene and method for producing multiple-unsaturated fatty acids
29224898	2002	Verfahren zur Herstellung mehrfach ungesättigter Fettsäuren in Pflanzen (<i>English: Process for the preparation of polyunsaturated fatty acids in plants</i>)
32920641	2003	Method for the production of polyunsaturated fatty acids
33132668	2003	Novel plant acyltransferases specific for long-chained, multiply unsaturated fatty acids
34120175	2003	Method for the production of multiply-unsaturated fatty acids in transgenic organisms
34894872	2004	Method for producing unsaturated omega-3 fatty acids in transgenic organisms
35911218	2004	Method for producing polyunsaturated fatty acids in transgenic organisms
38857863	2006	Processes for producing polyunsaturated fatty acids in transgenic organisms
43084431	2009	Regulatory nucleic acid molecules for enhancing seed-specific gene expression in plants promoting enhanced polyunsaturated fatty acid synthesis
48745816	2013	Gene expression or activity enhancing elements

Global main path, component 1



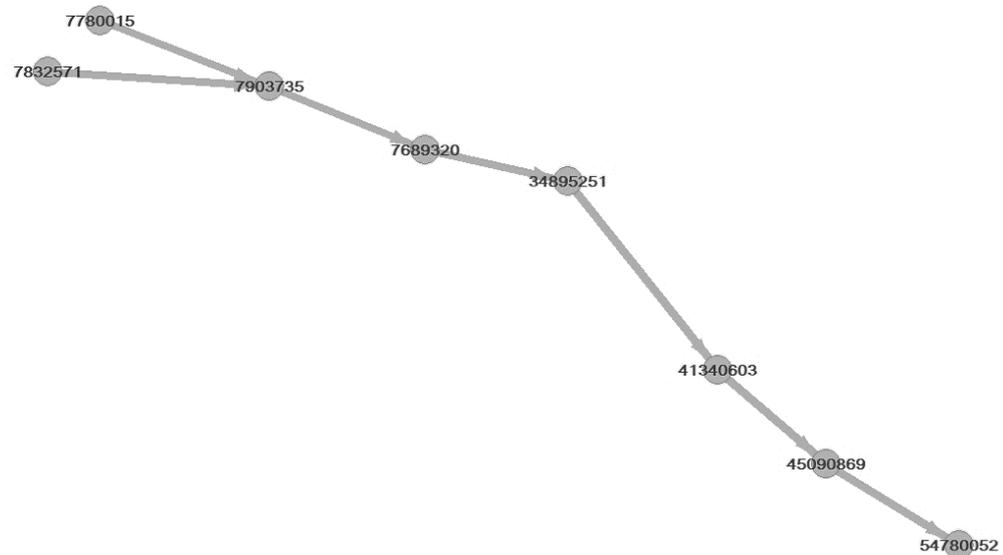
PATENT FAMILY	YEAR	TITLE
26018733	1995	Process for the production of a modified starch
7815459	1996	Neue Nucleinsäuremoleküle aus Mais und ihre Verwendung zur Herstellung einer modifizierten Stärke (<i>English: Nucleic acid molecules from corn and their use in the preparation of a modified starch</i>)
22318973	1998	Nucleic acid molecules from rice encoding an r1 protein and their use for the production of modified starch
7910992	1999	Nukleinsäuremoleküle aus Weizen, transgene Pflanzenzellen und Pflanzen und deren Verwendung für die Herstellung modifizierter Stärke (<i>English: Nucleic acid molecules from wheat, transgenic plant cells and plants and their use in the production of modified starch</i>)
26007459	2000	Monocotyledon plant cells and plants which synthesise modified starch
36607228	2004	Plants with increased activity of a starch phosphorylating enzyme
37441320	2006	Genetically modified plants synthesizing starch with increased swelling power
40380378	2008	Use of carboxamides on cultivated plants
40352205	2009	Method for improved utilization of the production potential of transgenic plants introduction
41319604	2009	Use of succinate dehydrogenase inhibitors for increasing the resistance of plants or parts of plants to abiotic stress
43625381	2010	Use of fluopyram for controlling nematodes in nematode resistant crops
43825341	2010	Active ingredient combinations comprising pyridylethylbenzamides and other active ingredients
45509139	2010	Use of fluopyram for controlling nematodes in crops and for increasing yield
45873472	2010	Use of active ingredients for controlling nematodes in nematode-resistant crops

Backward local main path, component 2



PATENT FAMILY	YEAR	TITLE
7666720	2000	Verfahren und Vorrichtung zur Herstellung von biologischem Gewebe in einer Wachstumskammer (<i>English: Method and apparatus for producing biological tissue in a growth chamber</i>)
7712173	2002	Vorrichtung zum Züchten oder Kultivieren von Zellen in einem dosenartigen Behälter (<i>English: Device for growing or cultivating cells in a can-like container</i>)
7751205	1995	Appts. for measuring contractile force of muscle tissue
7802112	1996	New micro-gravitational cell culture assembly
7823408	1997	Biotechnological reaction vessel for breeding microorganisms
7858958	1998	Tissue engineering procedure extracting and culturing cells with mechanical stimulation to simulate in-vivo growth
34484964	2003	Verfahren und Bioreaktor zum Kultivieren und Stimulieren von dreidimensionalen, vitalen und mechanisch widerstandsfähigen Zelltransplantaten (<i>English: Method and bioreactor for culturing and stimulating three-dimensional, vital and mechanically resistant cell transplants</i>)
41011042	2008	Perfusable bioreactor for the production and/or cultivation of a human or animal blood vessel and/or a human or animal tissue
54867177	2015	Modulares bioreaktorsystem (<i>English: Modular bioreactor system</i>)

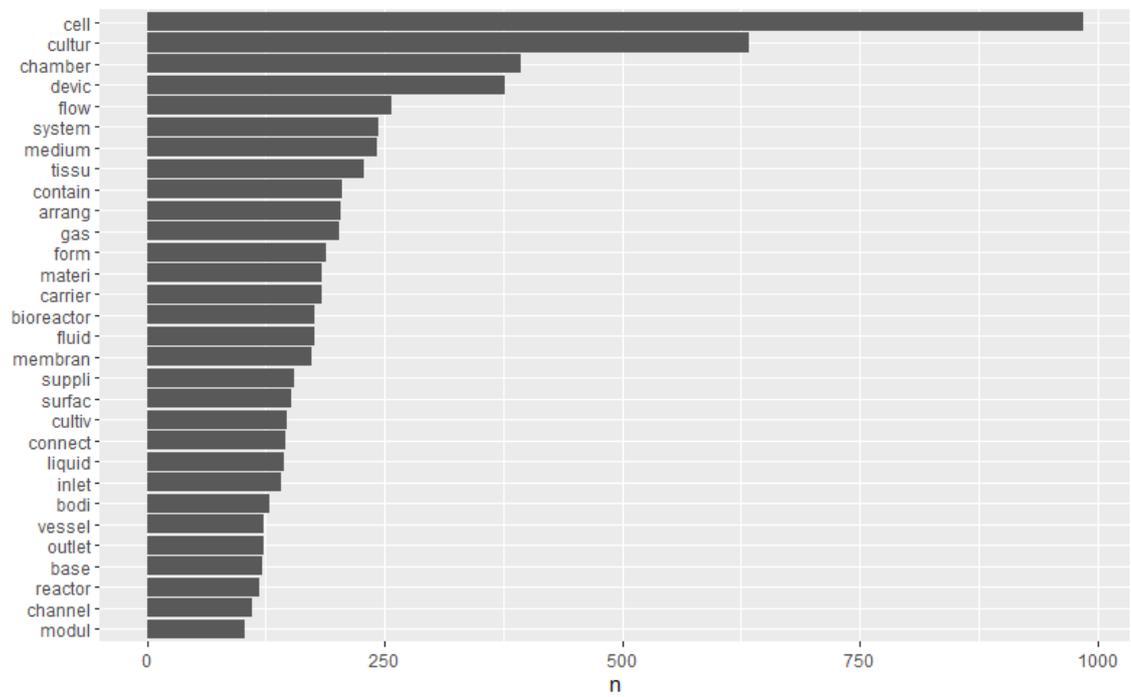
Global main path, component 2



PATENT FAMILY	YEAR	TITLE
7780015	1995	Cell cultivation process and assembly incorporates spirally-wound substrate belt
7832571	1997	Cell culture apparatus for three-dimensional cell or tissue cultivation
7903735	1999	Method for populating substrates with biological cells and populating devices that can be used therefor
7689320	2001	Vorrichtung zur Druckperfusion für das Züchten und/oder für das Behandeln von Zellen (<i>English: Device for pressure perfusion for the cultivation and/or treatment of cells</i>)
41340603	2003	Bioreactor and method for cultivating cells and tissues
34895251	2004	Inventive bioreactors and bioreactor systems
45090869	2010	Perfusion device
54780052	2014	Device and method for identification of microorganisms

From biotech to bioeconomy

New empirical evidence on the technological transition to plant-based bioeconomy based on patent data



Imprint

Bremen Papers on Economics & Innovation

Published by
University of Bremen, Faculty of Business Studies & Economics,
Institute for Economic Research and Policy (ierp)
Max-von-Laue-Straße 1, 28359 Bremen, Germany

Editors

Prof. Dr. Christian Cordes
Evolutionary Economics
Phone: +49 (0)421 218 66616, e-mail: c.cordes@uni-bremen.de

Prof. Dr. Dirk Fornahl
Regional and Innovation Economics
Phone: +49 (0)421 218 66530, e-mail: dfornahl@uni-bremen.de

Prof. Dr. Jutta Günther
Economics of Innovation and Structural Change
Phone: +49 (0)421 218 66630, e-mail: jutta.guenther@uni-bremen.de

Prof. Dr. André W. Heinemann
Federal and Regional Financial Relations
Phone: +49 (0)421 218 66830, e-mail: andre.heinemann@uni-bremen.de

Prof. Dr. Torben Klarl
Macroeconomics
Phone: +49 (0)421 218 66560, e-mail: tklarl@uni-bremen.de

Bremen Papers on Economics & Innovation #2002

Responsible Editor: Prof. Dr. Jutta Günther

All rights reserved. Bremen, Germany, 2019

ISSN 2629-3994

The working papers published in the series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors.