http://dx.doi.org/10.20396/rbi.v17i2.8652020

ARTIGO

Applying an IPC network to identify the bioenergy technological frontier

*Carolina da Silveira Bueno**

*José Maria Ferreira Jardim da Silveira***

*Antonio Marcio Buainain****

*Maria Ester Soares Dal Poz*****

Received: 25 March 2018 Revised version: 16 May 2018 Accepted: 18 May 2018

ABSTRACT

This paper proposes to identify the formation of technological frontier clusters. In the present study, we applied the method for bioenergy technologies. The first methodological step uses patents extracted from the Derwent Innovation Index, in which a data mining treatment is done. The second step is the construction of the IPC (International Patents Classification) network of patents from a co-occurrence matrix. Three clusters of technological knowledge are identified. The first and main cluster is the ethanol technologies, and two clusters in the training phase, one of which uses by-products and the other of technologies based on cellulosic material. The work shows how a general knowledge base is appropriate and favors the configuration of new frontier technologies that amplify the knowledge base of the bioenergy paradigm.

KEYWORDS | IPC; Networks; Bioenergy; Bioeconomy; Technological Frontier

JEL CODES | D85; B41; C38; C82; D01.

* Universidade Estadual de Campinas (Unicamp), Campinas (SP), Brasil. E-mail: carolinnasilveira@yahoo.com.br.

** Universidade Estadual de Campinas (Unicamp), Campinas (SP), Brasil. E-mail: jmsilv52@gmail.com.

*** Universidade Estadual de Campinas (Unicamp), Campinas (SP), Brasil. E-mail: buainain@eco.unicamp.br

**** Universidade Estadual de Campinas (Unicamp), Campinas (SP), Brasil. E-mail: ester.dalpoz@fca.unicamp.br

Aplicando rede de IPCs para identificar a fronteira tecnológica da bioenergia

RESUMO

Este artigo propõe um método para identificar clusters de fronteira tecnológica. No presente estudo, aplicamos o método para tecnologias da bioenergia. A primeira etapa metodológica utiliza patentes extraídas da Derwent Innovation Index, em que é feito um tratamento de *data mining*. A segunda etapa é a construção da rede de *International Patents Classifications* (IPCs) a partir de uma matriz de co-ocorrência. Foram identificados três clusters de conhecimento tecnológico. O primeiro é o de tecnologias de etanol e os outros dois em etapa de formação, sendo um de tecnologias que utilizam subprodutos e outro de tecnologias a partir de material celulósico. O trabalho mostra como uma base de conhecimentos gerais vai sendo apropriada e favorece a configuração de novas frentes tecnológicas que amplificam a base de conhecimento representada pelo paradigma da bioenergia.

PALAVRAS-CHAVE | IPC; Redes; Bioenergia; Bioeconomia; Fronteira Tecnológica

CÓDIGOS-JEL | D85; B41; C38; C82; D01.

---

1. Introduction

Bioenergy is an area of great interest to the economics of innovation. As several scientific and technological activities are combined by different agents with the aim of innovating, its study has been conducted following various methodological lines. For instance, bioenergy can be depicted through a layered approach, in which processes involving radical innovation are articulated with ongoing incremental processes, in one dimension or another, and a range of innovations, from far-reaching ones to others that create a market niche, are considered (WINSKEL et al., 2011).

Geels (2004) proposes an approach to cover high-complexity systems which identifies their hierarchical structures, addressing from broad issues, which involve a wide array of social actors, to issues that demand an understanding of the dynamics of technologies that stem from different knowledge fields and are articulated throughout the process of innovation generation.

Interpreting Winskel's (2011) and Geels' (2004) ideas, one of the justifications for the importance of bioenergy refers to environmental impacts (fuels that are seen as advanced because they contribute to the reduction of greenhouse gases emission). However, at a lower level, the competition with other power generation systems and similar products (such as electric cars and chemical synthesis products) must be considered. The different modalities and alternatives connected to bioenergy articulate various technological components, and as such they demand knowledge on which arrangement variations are more profitable in a given timeframe. At another level, farther from the market and closer to technological prospection, lies what allows for the understanding of how enabling technologies are constituted and unfolded, technologies which allow new concepts and innovating ideas to be truly implemented from a standpoint of production scales which are compliant to market demands.

In a study developed by Bueno, Silveira, and Buainain (2018), the building of technological knowledge from a paradigm in its initial developmental phase, in this case related to bioenergy, occurs through an international collaborative network (CN), due to a heavy dependence on international authorship of papers that is a trait of the network. If the opportunity to innovate is related to the development of such co-authorship networks and the same can be said of the interdependence of areas of knowledge, a matter which is dealt with on section 2 of this paper, we ask: are the networks a key parameter to understanding the development of technological frontiers in the emerging area of bioenergy?

The present study focuses on the analysis of applications of class classifications of technologies contained in patents (International Patent Classification – IPC[1]). The timeframe is long enough to enable the evaluation of the hypothesis that the research in biotechnology has become complex in terms of scientific knowledge requirements and the design and arrangement of building blocks for the construction of their technological platforms. From algorithms applied to the co-occurrence matrix of IPC subclasses, which can be interpreted as "knowledge areas" that occur jointly, clusters of greater interest to this study are generated. Knowledge connections forming complex networks are obtained, in the sense that some IPC classes are so tightly associated to others that they cannot be removed without profoundly altering the interpretation and, thus, the meaning of the network.

The study touches on the identification of the complex nature of technology flows in bioenergy: it is assumed that bioenergy is still in evolution and the areas of greater dynamism, which constitute its technological frontier, are increasingly formed by technologies with greater scientific and technological content, corroborating what other studies have pointed out in relation to biotechnology in general.

The more emerging the knowledge of a given technological system is, the bigger its level of complexity will be, since the manner and the extent to which its evolution is manifested is unknown. The present paper is based on the ideas of Kraft, Quatraro, and Saviotti (2009), who use patents as measure and social network analysis (nodes and links) to represent the flows of knowledge as an interpretative structure, by means of measurements and analysis of structural changes in knowledge itself. This set of elements (nodes) and their interactions (links) in the network suggest that its structure is related to the manner by which knowledge flows among actors, which is characterized by a specific architecture (knowledge basis, forms of transmission, and forms close to the market) that can evolve over time. In turn, time is a decisive effect in the interactions among economic agents involved in the collective process of knowledge creation.

In complex systems, the whole is greater than the sum of its parts. Overall, the properties of complex systems involve three dimensions: collective unity, functional organicity, and emerging property (SOLÉ et al., 1996). From these dimensions, this paper relies on complexity studies to understand the dynamic of networks:

1 All published patent requests are classified under the technological area they belong to. IPC is the international classification system, whose technological areas are divided from classes A to H. Classes follow a hierarchical system, containing subclasses.

- the areas of knowledge involved in knowledge flows are comprised of a set of interrelations among actors;
- the knowledge network exhibits a set of parts or subparts with internal individual relations connected to each other, thus forming a collective unity with its own dynamic, which is characteristic of a functional organicity;
- the interactions among knowledge areas that are required by the generation of innovation consist of an exteriorization of the network as a whole. This dynamic produces a set of unities that interact with each other and this set stands as a new level of evolution of the network, that is, it contains emerging properties.

The general hypothesis of this study is that the subnetworks that have been found by the methodological procedures here adopted point to characteristics of greater scientific and technological content in bioenergy as an emerging property of this system (dynamic in a, b, and c).

Our results confirm what has been observed by patents studies: a rising degree of complexity in research and technological development activities in bioenergy results in greater interdependence of areas that are, in turn, characterized by greater scientific and technological content in relation to the ones that prevailed three decades ago. This study confirms, with higher precision, the results obtained by Kraft, Quatraro, and Saviotti (2009) for the large area of biotechnology, using similar methodology.

The results also show a more intense drive in patenting in specific areas, which indicates that some targets of industrial interest have begun to consolidate, not only related to the improvement and enhancement of productivity in traditional products, but also aiming at the protection of what is, at once, the frontier of knowledge and technological advancement.

Section 2 summarizes some of the results from papers that act as a basis for this study and that have led to the identification of necessary matters of concern to further our examination, namely, the need to analyze the characteristics of the co-occurrence matrix and the networks that derive from it, in different levels of IPC classification, and the formation of clusters from the network analysis technique. Section 3 presents the methodology used and some secondary research results, mainly the ones related to the networks and subnetworks obtained in the study of Bueno et al. (2018). Section 4 analyzes the main results of this study and section 5 presents its conclusions.

2. Bioenergy knowledge networks

The emergence of innovation occurs from the interaction of various distinct components of knowledge which demand complex technological research and development structures, organized in layers, as presented by Winskel et al. (2014). Keasling (2017) shows that this is the case of most platforms related to biotechnology products, R&D--intensive activities (research and development) which seek to obtain liquid energy, and other economically interesting products from the transformation of biomass.[2]

Despite there being a large number of platforms and possibilities of technological development in bioenergy (CHERUBINI et al., 2010) in the period from 1970 to 2000, the core of innovative countries was kept small and stable. From 2000 on, new countries have been consistently incorporated to the research in bioenergy, whatever the level of aggregation which is considered (BUENO et al., 2016), a process combining technological advances in the field of distinct biomasses with the exploration of opportunities stemmed from the complementarity of assets (SOUZA et al., 2015).

Thus, the field of bioenergy is no longer interesting only to leading countries in biofuel production, involving now countries which dominate distinct production chain links, from enabling technologies to products that substitute the ones already produced by conventional ways (drop in) and those that are radically new (drop out) (SALLES FILHO et al., 2017; ARAUJO, 2016; FERRARI et al., 2016). Concurrently, institutional transformations, involving regulation and incentive policies, have been shaping the productive configurations of the segment.

Referring to Bueno, Silveira, and Buainain's work (2018), CNs exhibit intense international connections which involve bioenergy research. Their study also indicates the constitution of new forms of association that correspond to broader collective organizations, combining regions, different types of knowledge that is continually accumulated, and expansion of publication in several connected areas, a type of knowledge cumulativity.

The central argument is that both to accomplish the mapping of future possibilities, given the evolution of the knowledge and technological frontier (exploration), and to decide as to the areas of technology application (exploitation), varied forms of knowledge arrangement are necessary, retrieving, at times, knowledge and

2 The use of biomass for energy is connected to the exploitation of available resources in regions, such as sub-products from agropastoral activities (CORTEZ; BALDASSIN, 2016; HLPE, 2013). It is precisely in biomass that countries have sought solutions to reduce CO2 emission (greenhouse gas) (WINSKEL et al., 2014; HLPE, 2013), as well as to develop bioeconomy products (TRIGO et al., 2013).

technology that had been somewhat forgotten. It is what Winskel et al. (2014) term layers which define learning pathways.

The information on Table 1 is essential to the argument presented here: it is important to obtain an evolution over time in respect to publications, collaboration, and patenting. When a bibliometric survey of the period from 1970 to 2014 is undertaken, the hierarchy in terms of publication reflects research in agriculture and plant sciences. We can suspect that even biotechnology reflects techniques and laboratory procedures which do not involve knowledge that is more advanced than molecular biology, genetics, and microbiology. The correlation between number of publication and the presence of collaboration between countries in the elaboration of papers can be seen. Souza et al. (2015) also point to similar conclusions.

TABLE 1
Collaboration of areas of knowledge – bioenergy
1970-2014

Areas	Papers (total)	Collaboration (total)	%
Agriculture	4,760	2,675	56.20
Plant Sciences	2,290	1,360	59.39
Biotechnology	1,830	1,488	81.31
Energy fuels	732	312	42.62
Biochemistry/Molecular biology	603	352	58.38
Chemistry	595	271	45.55
Environmental Science	392	312	79.59
Engineering	331	280	84.59
Genetics	304	292	96.05
Microbiology	283	175	61.83

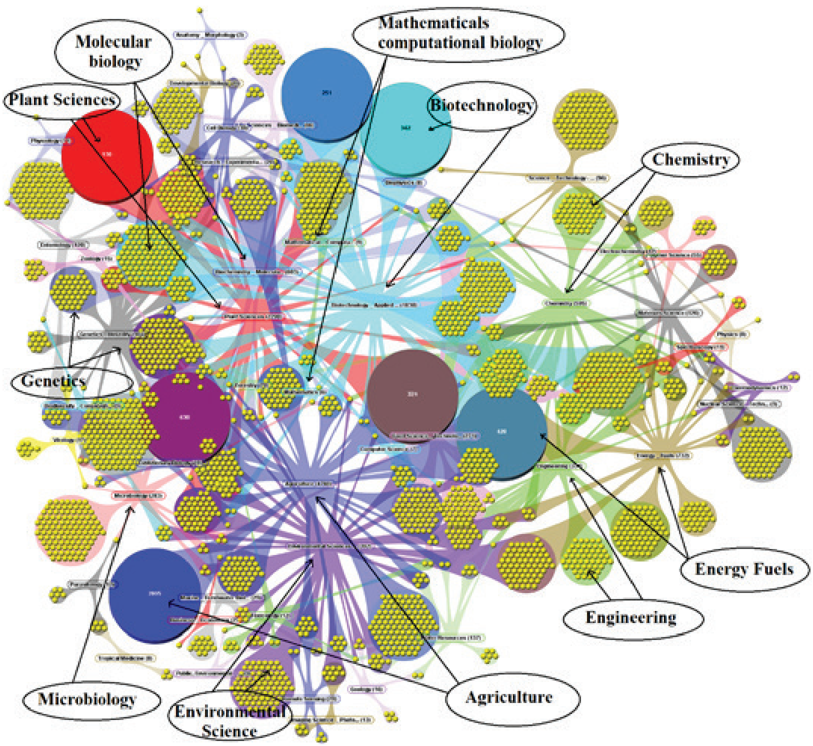
Source: Bueno, Silveira, and Buainain (2018).

Bueno, Silveira, and Buainain (2018) highlight differences between the patterns among collaboration networks related to a given main product but with different basic processes, as a result of different raw materials. Some results are summarized below:

- almost a hundred countries are part of the knowledge network, with Brazil being the degree centrality for the development of knowledge about sugar cane ethanol and the USA being the degree centrality in the knowledge network of corn bioethanol;

- Brazil and the USA were similar in co-authorship connection links and there was a significant degree of semantic relation in international collaboration with other countries; the study on patents revealed, to some extent, a strong connection between scientific activity and innovation generation;
- the topological grouping of knowledge areas from the papers, departing from the co-occurrence matrix, exhibits a high degree of interdependence, as is clear on Figure 1. Biotechnology and bioengineering are highly connected, by the same token as plant sciences and biochemistry.

FIGURE 1
Scientific collaboration network in research areas – bioenergy – opportunity 1970-2014

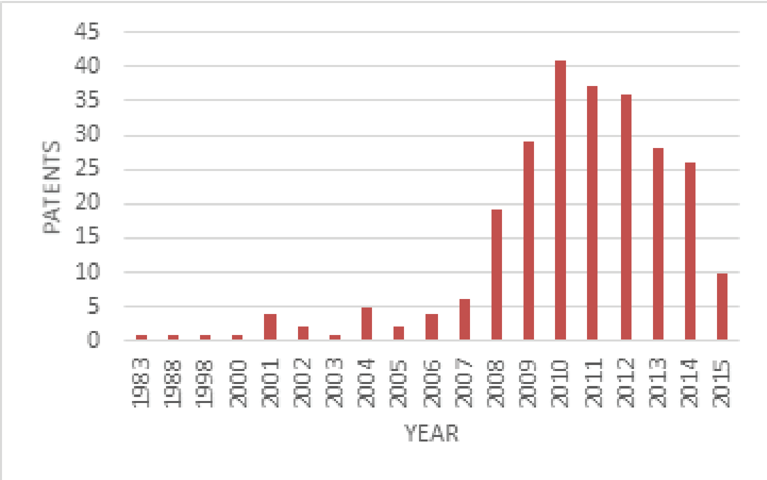


Source: Bueno, Silveira, and Buainain (2018).

Figure 1 displays all cases of co-authorship, grouped by knowledge area, from 8,243 papers on bioenergy. The collaboration is seen as a network. Note that biotechnology shows more than 80% of collaboration with other research areas, indicating that the level of interdisciplinarity of the areas is high, as it is the interdependence among them.

Graph 1, a sample of 263 patents related to second-generation ethanol, with emphasis on lignocellulosic biomass, shows that there was a substantial increase in patenting starting from the year 2000, but the bioenergy crisis – mainly the challenges posed to the second-generation of ethanol – affected the continuity of this process.

GRAPH 1
Yearly distribution of patents obtained by the search procedures – patents sample
1983-2015



Source: Bueno, Silveira, and Buainain (2018).

Bueno, Silveira, and Buainain's study (2018) on IPC classes in these patents (Graph 1) revealed 544 different technology classes, three of which exhibiting great power of articulation to other classes. However, it is interesting to realize their mobilization capacity, since they are connected with a large number of other IPC classes. This is the idea proposed here. Thus, an indication taken from the study mentioned is to understand changes in the way IPC classes in the bioenergy patents base are articulated over time with the purpose of identifying technological frontiers.

The next section presents the methodology developed to understand the structure and interdependence of clusters and to interpret them in light of the incorporation of new knowledge, which comes to constitute the bioenergy frontier.

3. Methodology

The methodological starting point is the work of Batagelj et al. (2014): in its 4th chapter, the book presents the methodology for working with collaboration networks and the following chapter discusses patents citation networks and identification of islands that can reveal patterns of technological development (page 190, for instance).

Our methodology was conducted through four stages: delimitation of patents sample; obtention of the co-occurrence matrix of 4-and 8-digit IPC classes; application of cluster formation procedures and identification of blocks per knowledge area in time sections; construction of a timeline for patents that had been already attributed and which belong to the main classes of 8-digit IPC classes, identified in the two previous stages.

Stage 1

It is begun by forming the patents database, with three methodological steps:

- construction of query: the tool IPC STATS Search, available from the website of World Intellectial Property Organization (WIPO) was used. This tool allows, with the insertion of one keyword, or a combination of them, for IPCs which are related to the proposed words to be recovered. The terms used with these instruments were: *ethanol, biofuels, bioenergy, sugarcane, corn, biomass, lignocellulosic, cellulose, hemicellulose, biochemical conversion, bioconversion, enzymatic hydrolysis, hydrolysis, enzyme, fermentation, saccharification;*[3]
- the IPC classes from the sample were analyzed by using Information Technology (IT) procedures;
- the patents database underwent a data mining treatment, carried out with the help of VantagePointtm,[4] with the objective of removing inconsistencies, repetition, and other problems derived from the previous steps (WU et al., 2014).

Stage 2

It involves the construction of the network stemming from the co-occurrence matrix in the IPC network. A major point is that, in order to obtain an IPC network,

3 Keywords were chosen with the assistance of specialists (from NIPE – Núcleo de Estudos de Planejamento Energético, Embrapa Agroenergia, and researchers from the Bionenergy Program at FAPESP).

4 Information on the TI program can be accessed on the website: <https://www.thevantagepoint.com/>.

there is no need to build a patent collaboration network; having the patents database constructed in the previous stage suffices.

The square co-occurrence matrix has the function of showing the relation between two or more elements: a referential node and what is called a neighbor. It represents in each element the number of times there was a transition of a neighbor, taking into account both distance and direction.

The same program generates the visualization of the matrix as a network image. This network is represented by graphs, which are illustrations of the network represented by a set of vertices (or nodes) V, a set of edges (or links) E, E → V, in which (e) is the source and (v) is the target of the directed edge (GOYAL, 2007). Consequently, the set of elements to which some are connected by links will be represented and enable the analysis both of each vertex and of the network as a whole (BATAGEJL et al., 2014).

The software VantagePoint[tm] generates the graph with the main classes and their links in accordance to the variables chosen by the user. In this case, subclasses with co-occurrence >0.75 were chosen. Thus, the areas with a lower level of correlation, that is, <0.75, were excluded from the network visualization.[5]

Stage 3

The next step is the obtention of IPC clusters. The software algorithm applies, to the network, an interactive scheme oriented by a tree of minimum network scope, with the aim of generating coordinates for the nodes expressing intercluster and intracluster links. This indicator displays technological areas and the mutual dependency among relevant technological paths. This stage occurs in accordance with the following principles (ACHTERT; BÖHM; KRÖGER, 2006):

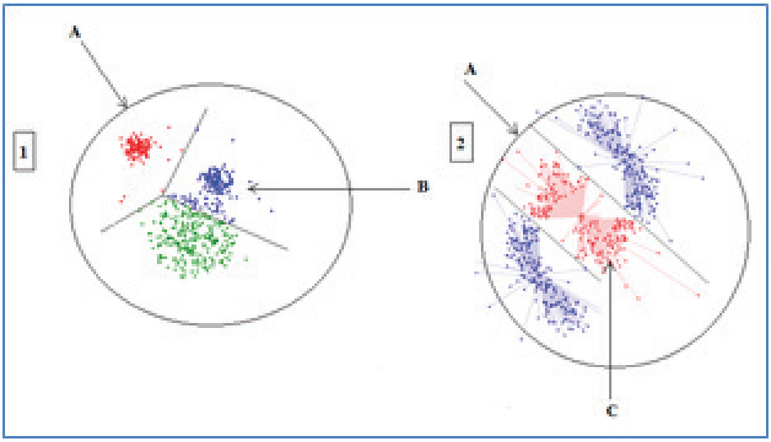
- a cluster (or grouping) represents a set of properties in relation to which the elements in a group are more similar to each other (in one way or another) than to the elements of other groups (or other clusters);
- the grouping of properties based on connectivity, also known as hierarchical grouping, represents an assemblage set that can be described by the farthest distance which connects the parts of the property set. That is, different distances, different groups, will form in accordance to the connection between them;

5 The choice of a correlation higher than 0.75 is justified by network visualization. If we had considered a smaller number, the visualization, due to the quantity of nodes being too large, would not have been possible.

- in hierarchical linkage clustering, algorithms do not provide one single partition of the data set, but a wide hierarchy of clusters that fuse with each other at certain distances. That is, this method will not produce a single data set, but also the direction of extreme values, such as additional agglomerates or even other clusters, which is known as chaining phenomenon.

Analyzing intra- and intercluster chaining phenomena is important, since they allow for the investigation of precisely the internal and external linkage chaining on the whole of a data agglomeration. For instance, note Figure 2: in **A** (total dataset) we show examples of clusters; in **B**, we have an intracluster; in **C**, an intercluster. That is, **B** and **C** have a level of interdependence upon **A**, explained by two possible causes: **A** has direct links that form cluster **B** (image 1); or cluster **A** originated cluster **C**, which is in-between clusters (image 2). The intercluster exhibits characteristics of grouping with the intracluster, i.e., it might be the cluster which generated a large dataset which in turn formed intraclusters, or it might be in-between clusters, as an intermediary or not.

FIGURE 2
Clusters illustrations



Source: Adapted from Achtert, Böhm, and Kröger (2006).

Stage 4

To answer the question of which cluster the path of formation of possible technological clusters is being manifested from, it was necessary to examine the scope of the IPC network individually for each cluster. This indicator reveals the presence (or absence) of a dependence of interest path and the respective clusters.

The results of using VantagePoint[tm] in the obtention of clusters were checked by the use of the software Pajek for the eight-digit IPC classes. It allowed for better understanding how the main classes are articulated and can be "separated" by classes of less importance in the analysis period, but which might serve as bridges.

Finally, the patents of the main 8-digit IPC classes were identified and a timeline was constructed so as to visualize the most intense period of patenting, which reveals the emergence of the technological frontier.

4. Results

4.1. Construction of the 4-digit IPC network and analysis of its characteristics

263 patents were obtained to form the patents database, according to the restrictions laid on the keywords. Therefore, this work refers to the bioenergy technological frontier and not to the different platforms that have been pointed out by the literature (CHERUBINI, 2010). In the database, 93 IPC classes (4 digits), forming a total of 544 IPC subclasses (8 digits), were identified. Since it is a large number of classes, the quantity of all possible subclasses of "bioenergy content" is also large.

The result refers to the total period of survey. Thus, the classes which reflect the general phenomenon indicated by Kraft, Quatraro, and Saviotti (2009) could lack prominence, in the event of the phenomenon of "technological sophistication" being too recent. However, this was not observed. There was predominance of subclasses C12P and C12N, whose classes and subclasses are presented on ATTACHMENT I. A careful reading of their descriptions demonstrates a strong affinity with the procedures of modern biotechnology applied to the production of liquid energy (BROWN; BROWN, 2012). This is an early significant result of this study.

The 4-digit IPC network is very dense; 19% of the link possibilities of the 4-digit IPC classes are accomplished. The average degree is 36.1. The data on Table 2 show that there is not a scale-free behavior (few with high degree and many with very low degree). Even less prominent classes exhibit a higher than 10 co-occurrence level (53 classes). The table also shows that there is a hierarchy among IPC classes, but also great thematic variety, a characteristic of frontier research in bioenergy, the main point of this study.

TABLE 2
4-digit IPC network – Distribution of the network centrality indicator – degree (n=263) in the period 1970-2015

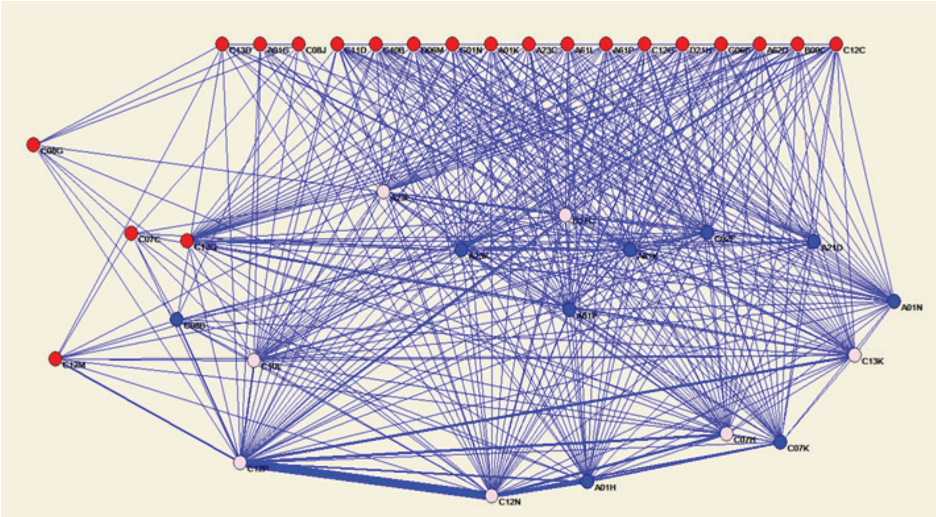
Estrato	Freq	CumFreq	REPRESENT
0-10	40	43.0108	B01J
10-20	16	60.2151	C12R
20-30	21	82.7957	C12M
30-40	9	92.4731	A01H
>40	7	100.0000	C12P

Source: Authors' own. Software Pajek.

An important fact of IPC networks is the asymmetry in link weight. For instance, the C12R class is seen linked to 20 other classes, from the 93 that exist, and is the limit of the second stratum. Still, 40% of the classes have more links than C12R. However, in 45 patents (17.7% of the total) there was co-occurrence of C12R and C12P, i.e., the weight of links varies significantly in IPC networks.

The IPC network is presented on Figure 3. The partition was carried out based on the index of degree centrality and reconverted to 5 strata that correspond to the ones presented on Table 1. The thickness of vertices shows the asymmetry of links. The difference in link weight is the basis for the formation of clusters that will be observed ahead. The network is characterized in a 4-digit IPC subnetwork, for classes with a degree above 20. The correlation between degree and weight exist, but there are exceptions. Therefore, classes that are very "popular" may not occur in a significant percentage of patents. The A23L class, which refers to food preparation, occurs longside only C12P in six patents (2.2%). In most cases, the partnership occurs in one or two patents and is no longer repeated. In the opposite case, C12P and C12N have very substantial weights (reflected in the thickness of the arc): these classes are prevalent in bioenergy frontier research, indicating the importance of the theme "fermentation and microorganisms".

FIGURE 3
IPC subnetwork for the three largest classes based on degree values (n-37)



Source: Authors' own. Software Pajek.
Caption: degree: pink: very high degree; blue: high degree; red: medium degree.
The thicker the link, the more substantial de presence of the IPC class in the total number of patents.

According to Figure 3, the classes that connect to many classes (high degree) and are present in a high percentage of patents (link thickness) are: C12P, C12N, C07H, respectively: fermentation, microorganisms and enzymes, and sugars, in addition to molecular biology methods. So far, the expected scientific basis of processes is demonstrated.

Next, more specialized, but appearing in a significant number of patents, are A01H and C07K, referring, respectively, to biotechnological methods for the obtention of new plants and their auxiliaries (tissue culture) and the obtention and manipulation of peptides (nitrogen compounds). Finally, connected to other patents, but with less prominence in relation to the total percentage, are D21C and C13K, which refer to lignocellulosic materials and industrial processes with saccharides.

Other centrality and impact indicators (betweenes centrality, proximity prestige) only endorse the findings attained above. The analysis of 4-digit IPC networks exhibits a coherent association of themes in which the development of enabling technology (enzymatic processes, fermentation, microorganisms improvement) is intensely articulated and resulta in a patenting process which can indicate an efficient protection on the part of their holders.

4.2. Analysis of 8-digit networks and identification of their main hubs

When we move from 4 to 8 digits, there is no change in the characteristics of the network. It is still very dense, with each 8-digit IPC class being linked to 41.5 others, in average. There is a decrease in density to 0.14, meaning that 14% of the possible links were achieved, which is a very high number for a network of 544 vertices.

Table 3 shows that the peak frequency is in lesser degree classes, but the decay factor does not allow for a scale-free behavior.

TABLE 3
8-digit IPC network – Distribution of the network centrality indicator – degree (n=544) in the period 1970-2015

Estrato	Freq	CumFreq	REPRESENT
0-50	411	75.6	C12N-001/16
50-100	88	91.7	C12M-001/00
100-150	23	96.0	C12N-015/82
150-200	16	98.9	C12P-019/14
>200	6	100.0	C12P-007/10

Source: Authors' own. Software Pajek

The first clusters already show an above-average cut in distribution. These are less prominent clusters, possibly present in a smaller number of patents as well. Figure 2 shows a subnetwork formed by classes that make up the three most prominent strata in terms of degree. The emerging links, the new areas that have started to connect to consolidated ones, are thus lost, but there is a greater understanding of bioenergy as an area in which technologies that are close to reaching the market incorporate knowledge of modern biotechnology, in the sense conveyed by Kraft, Quatraro, and Saviotti (2009).

The split from four to eight digits confirmed the importance of class C12P (fermentation and secondary synthesis processes) but specifies the predominant subclasses: C12P-007/02 (fermentation with organic compounds, hydroxy group); C12P-007/06 (ethanol) e CP-007-10 (cellulosic material substrates). There is great closeness to CP12-019/00 (preparation using enzymes, containing sucrose radicals) and CP12-019/02 (preparation using enzymes with monosaccharides). It is clear that the breakdown of C12P shows how some specific types of fermentation technology for biomass treatment aiming at ethanol production are present in the patents of the area. As we have stated above, it is no different from what was expected.

Class A01H-005/00 (angiosperms) appears alongside many IPC classes, but exhibits a preferential connection to class C12N-015/82 (gene transformation methods). The latter is linked to a smaller number of 8-digit classes (100-150), suggesting a level of specialization in plant gene manipulation. The link between A01H and C12N-029, which refers to vector and viruses used in gene manipulation, confirms the scenery. However, a significant number of patents ties plant improvement and manipulation to research on enzymes and microorganisms, which clearly forms the frontier on bioenergy methods and materials.

Subclass C13K-00/02 (connected to between 150 and 200 other classes) is strongly linked to the fermentation of organic compounds class (C12P-19), which refers to saccharides compounds that are originated from the treatment of cellulosic material. It is strongly connected to the preparation and separation of processes originated from fermentation.

4.3. Cluster formation based on IPC classes and in the area classification of patents

Figure 4 presents the main 8-digit IPC classes and their most important links, following the rule of only selecting correlation values from the co-occurrence matrix which are larger than 0.75. Thus, it is possible to see the importance of each class (represented by the granulated ball, which conveys the weight of each class, that is, the number of patents in which it is present) as well as the main classes to which it is linked. It is noteworthy that the graph allows for the identification of IPC classes that are closer and, therefore, define which research areas are complementary, but which also indicate specialization.

Cluster A gravitates around ethanol technologies. The main class is CP12 007-06, present in 105 patents and linked to 353 other 8-digit IPC classes (65%). It is a class that links to other classes, in and outside of its 4-digit level (using the criteria defined by Corredoira and Banerjee, 2014). It is a cluster of broad areas, from traditional technologies to highly advanced ones. It is explained by the fact that it contains traditional bioethanol production technologies that were patented as of the 1980s. A reading of the patents related to the cluster showed technologies in the areas of enzymatic hydrolysis, fermentation procesesses, genetic modification, microorganisms, and enzymes. There has been an increase in the number of patents from 2008, indicating that there is no stagnation.

As was shown in the section on methodology, VantagePoint™ identifies intermediary clusters which "separate" two clusters that are more prominent. Here, cluster B separates clusters A and C, which are based on IPC classes of strong presence in patents and high level of connection to the areas identified by the research. Cluster B is formed by technologies to convert biomass into ethanol. It is the smallest cluster, in comparison to the others, with patents starting from 2004. The central class in the cluster, C12P007-08, refers to sub-products from residue and material that contain cellulose, hemicellulose, and lignin. It is linked to 72 other classes (class 2 degree, that is, relatively low) and present in 21 patents (8%); its preferential connection is to patents of big clusters, A and C, a link that is stronger than the in-cluster connections, and it can be considered a sort of "bridge" between more traditional clusters, with older patents, and more recent developments, taking place in cluster B. It is also linked to classes related to enzymatic processes, which are also emergent, i.e., are still present in few patents. These technologies are also present in cluster A. There is indication that the patents which cite classes from cluster B are receiving information and knowledge from the other two main clusters.

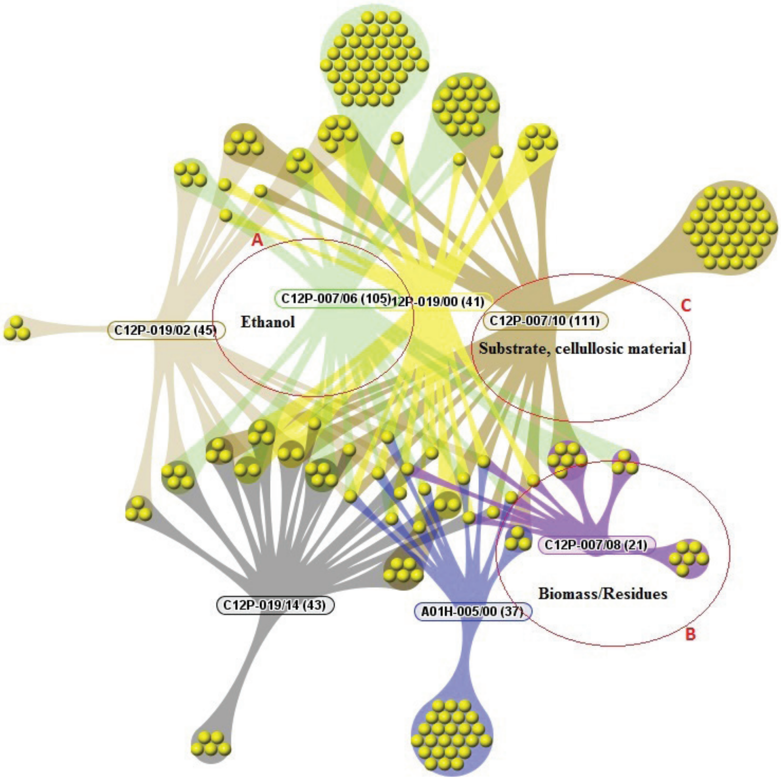
Cluster C uses technologies to draw sugar from cellulosic material, mainly from wood residue. Its main technological areas are organic chemistry and biotechnology. Patenting started in the 2000s and the patents are also linked to the classes present in cluster A.

Figure 4 also shows closeness among substrates and cellulosic material, saccharification processes using enzymes (C12P019-02 and C12P019-14, in this case hemicellulose and alpha-amylase) with more proximity to cluster C but with an important link to more traditional technologies (cluster A) as well. The themes are relevant, occurring in 16% of patents, which shows the links between second-generation processes and the treatment of sugars from enzymatic hydrolysis.

The results confirm the research on co-occurrence and entropy index calculation (how much the co-occurrence of terms can reveal the organization of a patents database) carried out by Murakami et al., (2015).

At last, we note the closeness to A01H (present in 15% of the patents), involving plant transformation and with strong connection to classes related to microorganisms transformed by genetic engineering (classes that were not included in the analysis due to the 0.75 criterium), and less closeness to defining classes of more important clusters, which were present in a greater number of patents. It is one more frontier, combining new plants and processes that use microorganisms that run parallel to the main research frontier.

FIGURE 4
Bioenergy: Subnetworks from the main hubs (8-digit IPC classes)



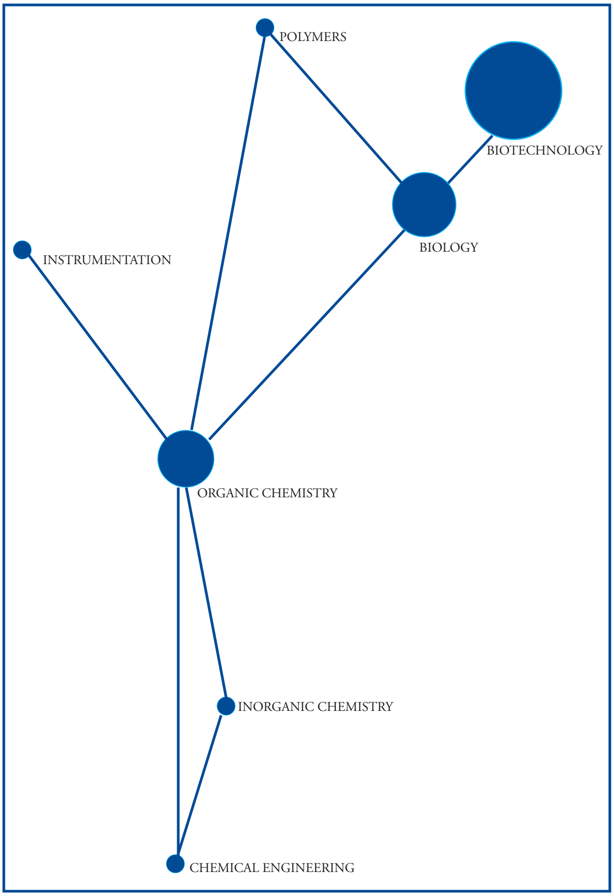
Source: Authors' own, software VantagePoint[tm.]
Caption: A: Ethanol; B: waste, biomass; C: substrates and cellulosic material.

In short, three 8-digit C12P classes (02, 06, and 10) are present in a great number of patents and are articulated to other classes, forming sets in many patents. However, there are signs of specialization that merit thorough investigation in subsequent research.

VantagePoint[tm] also allows for the obtention of classes consolidated by wide knowledge areas (Figure 5), representing their relative importance in the patent database and the closest links, with emphasis on the ones which present correlation levels that are higher than 0.75 (taking the co-occurrence matrix as a basis, as described in the methodological session). Nonetheless, the method preserves the areas that act as a connection among the most prominent ones. It is noteworthy that the bigger the circle is, the higher the frequency of the areas in the patents is.

POLYMERS

FIGURE 5
Main specialization classes which comprise the patents sample



Source: Authors' own, Software VantagePoint[tm]

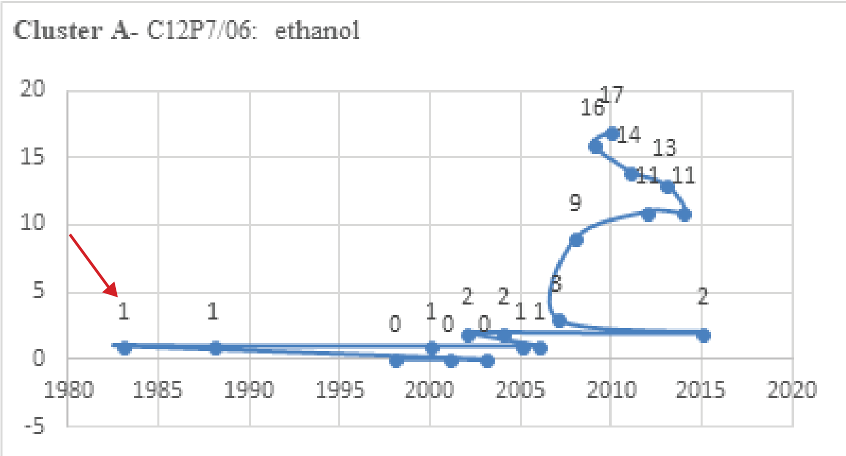
Considering the importance of the two central classes in the clusters presented on Figure 4, we will now turn to the timeline of the clusters.

4.4. Timeline of the three most important IPC classes and two emerging classes

The analyses suggest that, in the period from 1970 to 2015, there are more important subperiods, which reveal a greater combination of various types of scientific knowledge that constitute the frontier of the area, particularly the research on second-generation ethanol (BUENO et al., 2018).

The first cluster to be analyzed was cluster A. Graph 2 covers a period up to 2015 and shows that the technologies were initiated in 1983. At first, this would be the "most traditional" subclass, farther from the knowledge frontier, and it is present in a large number of patents, many of them using the knowledge frontier.

GRAPH 2
Timeline of class C12P7/06 – Center of cluster A – Patents

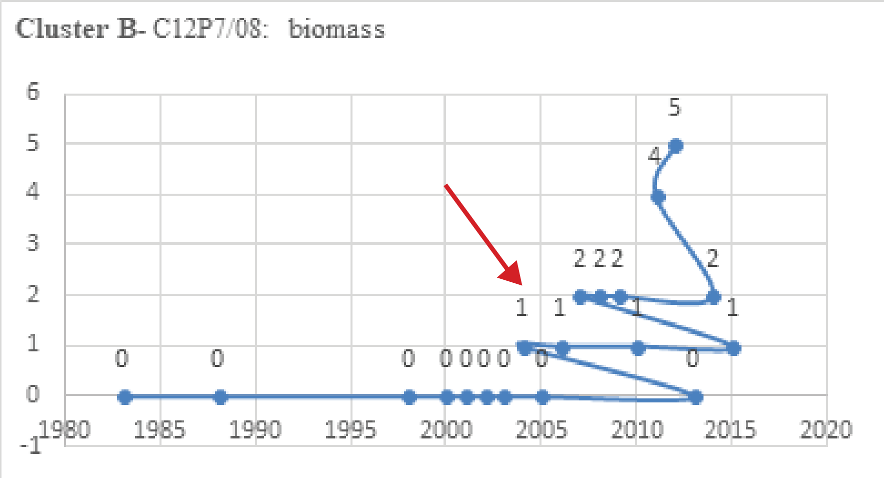


Source: Authors' own.

There is a resemblance between Graphs 3 and 4: the patenting period coincides, with the one from 2010 to 2015 being slightly more concentrated than the one before. As was seen in the last subsection, this is the class that relates treatment of cellulosic residue to biotechnology and organic chemistry. It appears connected to a great number of other classes, but the closeness to patents related to enzymatic and saccharification processes renders it a key factor in the identification of a second-generation technological frontier based on the dominance of biotechnology, the transformation of microorganisms, and the development of new enzymes and processes associated to them.

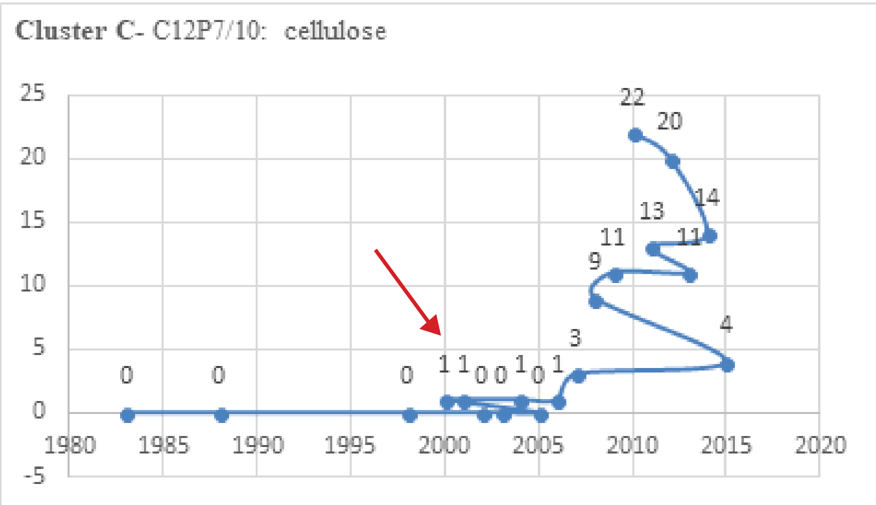
In Graph 3, cluster B entails technologies to convert biomass into ethanol. It is the smallest cluster, in comparison to the others, with patents initiated in 2004. In this cluster, technologies come from the areas of organic chemistry, organic physics, lignin, genetics, microorganisms, enzymes, pre-treatment, second-generation saccharification, recombinant DNA, and biotechnology. These technologies are also present in cluster A. That is, cluster B has captured technologies from cluster A, but not from cluster C.

GRAPH 3
Timeline of patents of class C12P7/08 – Center of cluster B



Source: Authors' own.

GRAPH 4
Timeline of class C12P7/10 – Center of cluster C – Patents



Source: Authors' own.

In Graph 4, cluster C displays technologies to draw sugar from cellulosic material, mainly from wood residue. The intermediary class, related to patents that refer to sub-products from the processing of certain types of biomass (mainly residues), maintains shape, but with weaker importance. It is clear from the analysis of clusters that this is not about a subset of clusters A and C only, but that these patents are connected to other classes, related to products as carboxylic acid esters.

Still, there is evidence of a joint movement of greater patenting intensity from 2005 to 2015, with more emphasis in the recent period. These results support a task of keeping track of changes over the following years, from the confirmation of the characteristics of the technologies related to the bioenergy frontier to their reversion, due to problems observed in market-scale production.

5. Conclusion

The method that applies the IPC network to identify the formation of technological knowledge clusters has led to the acknowledgement that the network structure exhibits three main areas: cluster A was identified as the intercluster which drives new areas, represented by ethanol, and the intraclusters B and C were detected as clusters in a formation stage, represented by ethanol stemming from cellulosic material and biomass. That is, cluster A is directing the formation of technologies in clusters B and C. That means that clusters B and C, by developing their trajectories, are further strengthening the bioethanol cluster. The intensive use of new scientific and technological knowledge, which contribute to the development of these biology-based processes for the transformation of natural resources into energy (B and C), characterizes B and C as technological frontiers of bioenergy, which is a product of bioeconomy.

Figure 3 showed various areas in co-occurrence in bioenergy. If we consider that incremental processes are attached to processes of radical innovation in the shape of research networks (as shown in Figure 5), it can mean that bioenergy is in symbiosis with the biotechnology paradigm, which makes the grasp on the development of these networks even more complex. Was it the biotechnology paradigm what led to the development of radical innovations for bioenergy technologies? Is the development of the bioenergy paradigm a strengthening of the bioeconomy paradigm?

At first, the phenomenon that characterizes the development of new technological fronts in this area of bioenergy is the interdependence of knowledge areas (Figure 4). This interdependence is also what brings about an imbalance in knowledge, with a reaction of new connections between new or old variables, which, in turn, form new clusters in technological areas that differ from the corresponding ones in the network. This is the same phenomenon observed by Krafft, Quatraro, and Saviotti (2009).

Some remarks are emphasized:

- the structure of technological knowledge (represented by patents) shows the interdependence of cluster A; that is, the technological knowledge developed in the shape of new links which generate other technology clusters (B and C). This result signals that knowledge interdependence would be a pattern in the development of new technology fronts in bioenergy, interdependence being an organizational unity, in which the formation of new clusters is an indissociable factor that amplifies the bioenergy paradigm;
- clusters B and C do not exhibit variables between each other, but are interdependent on cluster A. In this case, cluster B did not capture the knowledge of cluster C, and vice-versa. If we consider that the nodes in cluster A form new clusters (B and C), these spillover effects cause an imbalance in knowledge inside the network, which, in turn, creates variety. This imbalance could be related to the functional organicity of the network, which enables a collective unity. In other words, even if a node is not directly linked to another, as in the case of B and C, they manifest in the whole of the network, generating, thus, the phenomenon of emerging properties, that is, the phenomenon of technological frontiers.

Remarks (i) and (ii) reveal a new phenomenon: the properties of a network are not a natural consequence of its constituent elements when they considered in isolation. The complexity which is entailed in the formation of new links and arrangements produces elements which are comprised of different parts. These parts interact in order to generate new arrangements in the collective behavior of the network. Such characteristics are typical of complex systems. Are cumulativity and interdependence of knowledge areas the essential mechanisms to the formation of new technological fronts and, consequently, to the interest of new countries in taking on the pursuit (via international collaboration) of technological frontiers?

Knowledge accumulation seems to indicate technological advances, which, in turn, foster a fundamental source of technological variety in the initial move of the bioenergy paradigm. However, we can already infer, albeit not completely, that the unfolding of new IPC classes (Graphs 2, 3, and 4) showed that, over time, new technology clusters come to exist, which, clearly and unequivocally, points to technological variety. It is hard to assess exactly what weighs more in order for this technological variety to take place, nonetheless, the phenomenon, in this case, is in the interdependence of areas as a cause of formation of new knowledge arrangements (Figure 5).

The notion of technological innovation has become more and more systemic and sequential, configuring modular complex systems in which bottlenecks related to the variety of innovation appear as interdependent on old and new variables of knowledge generated in the network. This study has shown how a basis of general knowledge (cluster A) is appropriated (in this case, via patents) and favors the configuration of new technological areas (clusters B and C) which amplify the knowledge represented by the frontiers of the bioenergy paradigm.

Acknowledgements

Financial support FAPESP - São Paulo Research Foundation - nº 16/13820-4.

References

ACHTERT, E.; BÖHM, C.; KRÖGER, P. DeLi-Clu: boosting robustness, completeness, usability, and efficiency of hierarchical clustering by a closest pair ranking. In: NG, W. K. et al. (Ed.). *Advances in knowledge discovery and data mining*. Germany: Springer, 2006. p. 119-128.

ARAÚJO, W. A. Ethanol industry: surpassing uncertainties and looking forward. In: SALLES-FILHO, S. et al. (Ed.). *Global bioethanol*: evolution, risks, and uncertainties. London: Elsevier, 2016. p. 1-33.

BATAGELJ, V.; MRVAR, A. Pajek. *Encyclopedia of social network analysis and mining*. New York: Springer, 2014. p. 1245-1256.

BIC Press Release. European Bioeconomy 2013: € 2.1 trillion turnover and 18.3 million employees. Bruxelas, 2016. Available in: <https://goo.gl/kX6VfE>. Access in: 10 jul. 2017.

BROWN, T. R.; BROWN, R. C. Techno-economics of advanced biofuels pathways. *RSC Advances*, n. 17, 2013.

BUENO, C. S.; SILVEIRA, J. M. F. J.; BUAINAIN, A. M. International collaborative networks: sugarcane ethanol versus corn ethanol. Is there a room for technological convergence? In: 20TH ICABR CONFERENCE. *Proceedings…* Ravello, Italy: International Consortium on Applied Bioeconomy Research, 2016.

_________. Innovation, networks and the paradigm of biofuels. *Int. J. Entrepreneurship and Small Business*, v. 35, n. 3, 2018.

CHERUBINI, F. The biorefinery concept: using biomass instead of oil for producing energy and chemicals. *Energy Conversion and Management*, v. 51, n. 7, p. 1412-1421, Jul. 2010.

CORREDOIRA, R.; BANERJEE, P. *Measuring patent's influence on technological evolution*: a study of knowledge spanning as driver of inventive activity. College Park, MD: Robert H. Smith School, 2014 (Research Paper n. RHS 2428452). Available in: <https://ssrn.com/abstract=2428452>. Access in: 14 jan. 2018.

CORTEZ, L. A. B.; BALDASSIN JR., R. Policies towards bioethanol and their implications: case Brazil. In: SALLES-FILHO, S.; BARBOSA CORTEZ, L.; SILVEIRA, J. M.; TRINDADE, S. (Ed.). *Global bioethanol*: evolution, risks and uncertainties. Academic Press, 2016. p. 142-162.

EASLEY, D.; KLEINBERG, J. *Networks, crowds, and markets*: reasoning about a highly connected world. Cambridge: Cambridge University Press, 2010.

FERRARI, V. E.; DAL-POZ, M. E. S.; SILVEIRA, J. M. J. F. The importance of intellectual property rights to the plant biotechnology industry In: THE WINIR SYMPOSIUM ON PROPERTY RIGHTS. *Proceedings…* Bristol, UK, 2016.

GEELS, F. From sectoral systems of innovation to socio-technical systems. *Research Policy*, v.33, n. 6-7, p.897-920, 2004.

GOYAL, S. *Connections*: an introduction to the economics of networks. Princeton and Oxford: Princeton University Press, 2007.

HLPE. *Biofuels and food security*: a report by the high-level panel of experts on food security and nutrition. Roma: Committee on World Food Security, 2013.

KEASLING, J. D et al. Rhodosporidium toruloides: a new platform organism for conversion of lignocellulose into terpene biofuels and bioproducts. *Biotechnology for Biofuels*, n. 10:241, 2017.

KOPPEJAN, D.; VAN EST, R. Bio-economy policy: inspiring but not leading. In: ASVELD, L.; VAN EST, R.; STEMERDING, D. (Ed.). *Getting to the core of the bio-economy*: a perspective on the sustainable promise of biomass. The Hague: Rathenau Institute, 2011. p. 35-52.

KRAFFT, J.; QUATRARO, F.; SAVIOTTI, P. *The evolution of knowledge base in knowledge--intensive sectors*: social network analysis of biotechnology. Italy: Università di Torino, 2009 (Working Paper Series, n. 9).

MURAKAMI, T.; SILVEIRA, J. M.; SOUZA, L. G. A.; BUENO, C. S. Applying entropy indexes to identify technological trajectories: second-generation bioethanol production. In: ENCONTRO NACIONAL DE ECONOMIA, 43. 2015, Florianópolis. *Anais*... Rio de Janeiro: Anpec, 2016. Disponível em: <https://ideas.repec.org/p/anp/en2015/146.html>. Acesso em: 06 out. 2017.

NEWMAN, M. The structure and function of complex networks. *SIAM Review*, v. 45, p. 167-256, 2003.

SALLES-FILHO, S. Conclusions: Futures of bioethanol – main findings and prospects. In: SALLES-FILHO, S.; BARBOSA CORTEZ, L.; SILVEIRA, J. M.; TRINDADE, S. (Ed.). *Global bioethanol*: evolution, risks and uncertainties. Cambridge, MA: Academic Press, 2016. p. 238-248.

SCOTCHMER, S. The political economy of intellectual property treaties. *Journal of Law, Economics, and Organization*, v. 20, n. 2, p. 415-437, 2004.

SOLÉ, R. V.; MANRUBIA, S. C.; LUQUE, B.; DELGADO, J.; BASCOMPTE, J. Phase transitions and complex systems: simple, nonlinear models capture complex systems at the edge of chaos. *Complexity*, v. 1, n. 4, p. 13-26, 1996.

SOUZA, L. G. A.; ARANHA, G. M.; DAL POZ, M. S.; SILVEIRA, J. M.; Collaborative networks as a measure of the innovation systems in second-generation ethanol. *Scientometrics*, v. 103, n. 2, p. 355-372, 2015.

TRIGO, E.; HENRY, G.; SANDERS, J.; SCHURR, U.; INGELBRECHT, I.; REVEL, C.; SANTANA, C.; ROCHA, P. *Towards bioeconomy development in Latin America and the Caribbean*. Alcue KBBE, 2013 (Bioeconomy Working Paper, n. 2013-01). Disponível em: <https://goo.gl/Rb3LFg>. Acesso em: 23 set. 2017.

UNITED STATES. Department of Energy. *Federal activities report on the bioeconomy*. Washington, DC, Feb. 2016. Disponível em: <https://goo.gl/zOWxsB>. Acesso em: 05 dez. 2016.

WINSKEL, M.; MARKUSSON, N.; JEFFREY, H.; CANDELISE, C.; DUTTON, G.; HOWARTH, P.; JABLONSKI, S.; KALYVAS, C.; WARD, D. Learning pathways for energy supply technologies: bridging between innovation studies and learning rates. *Technological Forecasting and Social Change*, v. 81, p. 96-114, 2014.

WINSKEL, M.; ANANDARAJAH, G.; SKEA, J.; JAY, B. Accelerating the development of energy supply technologies: the role of research and innovation. *Energy 2050*: making the transition to a secure low-carbon energy system. London: Routledge, 2011. p. 187-218.

WU, X.; ZHU, X.; WU, G.; DING, W. Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, v. 26, n. 1, p. 97-107, 2014.

ATTACHMENT

CHART 1
4-digit IPC technological classes which compose the patents sample

IPC classes	IPC desciption
IPC classes (4 digits)	C12P; C12N; C12R; A01H; C07H; C13K; C07K; C12M; C10L; C08B; D21C; A01N; C12Q; C07C; A23K; B01D; A23L; B09B; A01P; A61K; C08H; C08G; C13B; B01J; C12S; C13D; A01G; C02F; C07B; C07G; C08F; C08J; C08L; C10G; C11D; C12F; C40B; D06M; G01N; A01K; A21D; A23C; A61L; A61P; B65D; C09K; C10J; C11C; C12G; D01F; D21H; F23G; G06F; A01C; A01D; A23D; A46D; A61Q; A62D; B04B; B04C; B05B; B05D; B09C; B29B; B30B; B32B; B82Y; C01B; C01C; C01F; C05D; C05F; C05G; C08K; C09B; C09J; C10M; C12C; C12J; C13C; C13F; D06P; D07B; D21B; D21F; D21J; F02B; F16L; G06K; G09C; H01H; H01M.

Source: Authors, from Derwent Innovation Index.

CHART 2
Summary of some 8-digit IPC groups which compose the patents sample

Subclass		Description
C12P7	06	(…) Ethanol, i.e. non-beverage
C12P7	08	(…) produced as by-product or from waste or cellulosic material substrate
C12P7	10	(…) substrate containing cellulosic materials
C12P7	14	(…) Multiple stages of fermentation; Multiple types of microorganisms or reuse for microorganisms
C12P	19/14	(...) produced by the action of a carbohydrase, e.g. by alpha-amylase
C12P	39/00	Processes involving microorganisms of different genera in the same process, simultaneously
C12N	1/15	(...) modified by introduction of foreign genetic material
C12N	1/16	(...) Yeasts; Culture media therefor
C12N	1/18	(...) Baker's yeast; Brewer's yeast
C12N	1/19	(...) modified by introduction of foreign genetic material
C12N	1/20	(...) Bacteria; Culture media therefor
C12N	1/21	(...) modified by introduction of foreign genetic material
C12N	1/22	(...) Processes using, or culture media containing, cellulose or hydrolysates thereof
C12N	9/02	(...) Oxidoreductases (1.), e.g. luciferase
C12N	9/04	(...) acting on CHOH groups as donors, e.g. glucose oxidase, lactate dehydrogenase
C12N	9/14	(...) Hydrolases
C12N	15/1	(...) Preparation of mutants without inserting foreign genetic material therein (…)
C12N	15/2	(...) Preparation of hybrid cells by fusion of two or more cells, e.g. protoplast fusion
C12N	15/3	(...) Bacteria
C12N	15/4	(...) Fungi
C12N	15/5	(...) Plant cells
C12N	15/10	(...) Processes for the isolation, preparation or purification of DNA or RNA (...)
C12N	15/11	(...) DNA or RNA fragments; Modified forms thereof

Source: IPC. Available at: <http://ipc.inpi.gov.br>.