Scientometrics (2015) 103:355–372
DOI 10.1007/s11192-015-1553-2

Collaborative Networks as a measure of the Innovation Systems in second-generation ethanol

Luiz Gustavo Antonio de Souza · Márcia Azanha Ferraz Dias de Moraes · Maria Ester Soares Dal Poz · José Maria Ferreira Jardim da Silveira

Received: 6 April 2014 / Published online: 13 March 2015

Abstract Ethanol obtained from the conversion process of different types of biomass is a renewable source of fuel and since 2010 it has been classified as an "advanced fuel" by the EPA, due to its contribution to the reduction of the impacts of GHG emissions. Recent literature stresses the importance of the use of second-generation fuels to reduce the impacts of the direct and indirect use of land, mostly on agricultural prices. Although these demands constitute a clear clue to R&D activities, there are an impressive number of alternatives, regarding different kinds of biomass, processes and byproducts, a complex matrix of technological opportunities and the demands that generates a clear incentive for collaboration. This paper uses both the Bibliometry and Scientometry approach and the Innovation System (IS) literature under the perspective of Social Networks Analysis (SNA) to build Collaborative Networks (CNs) to the second-generation ethanol (lignocellulosic) using ISI Web of Science database. The adopted procedure emerges once authors, countries and institutions related to bioenergy have incentives to share information in the process of creating a new role in partnership—a network point-of-view. The results show that the United States is in a better position than other countries, improving the role of the university in their IS while China proves to be a great ally of the United States regarding the production of technology to produce lignocellulosic ethanol. Brazil however, does not appear well placed in the network, despite being the second largest producer of first-generation ethanol in the world.

L. G. A. de Souza (✉) · J. M. F. J. da Silveira
Institute of Economics – IE/UNICAMP, University of Campinas, Rua Pitágoras, 353, Barão Geraldo, Campinas, São Paulo, CEP 13083-857, Brazil
e-mail: luizgustavoeco@gmail.com

M. A. F. D. de Moraes
Escola Superior de Agricultura "Luiz de Queiroz" – ESALQ/USP, University of São Paulo, Piracicaba, São Paulo, Brazil

M. E. S. Dal Poz
Faculty of Applied Sciences – FCA/UNICAMP, University of Campinas, Campinas, São Paulo, Brazil

Keywords Ethanol · Second-generation · Lignocellulosic · Networks · Innovation

Mathematics Subject Classification 91D30

JEL Classification O13 · O32 · D85

Introduction

The projected demand for energy increases as well the mitigation of the effects of climate change and the expected depletion in stocks of fossil fuels, highlights the Innovation Systems (IS) of many countries concerning the key importance of the development of alternative sources of energy, particularly those related to the use of biomass. There is increased concern about energy policies and the need to include Science, Technology and Innovation (S, T&I) policies within the body of those policies (HLPE 2013; Babcock and Pouliot 2013).

The advantages of renewable sources includes energy security, favorable environmental impacts, job creation (in industry and agriculture) as well as laying foundations for the development of biotechnology, and chemical and materials engineering, all of which are of key importance to the productive sector, as Rabelo (2010) have shown.

Criticism of the impact of energy polices on food prices and the need to strengthen bioeconomy—reducing costs and improving economies of scope of the production units (Wielen and Breugel 2014; Willems 2015; Liu et al. 2014)—have drawn attention to the second-generation biofuels.

The main requirements for the low-cost and high-productivity production of second-generation ethanol are[1]:

(a) Feasible and disposable raw material to attend the year-round requirements of the process unit. Seasonality and climate disturbances that can affect the plant are problematic;
(b) Producers that are continually seeking greater productivity and efficiency, generating a continuous path of searching, learning, adapting and adopting technologies within the industry;
(c) Available knowledge and technology, obtained from scientific papers that scientists published as a result of the cumulative process of knowledge such as the advances in basic and applied science; in addition, this knowledge used by the industry, will generate technologies to convert into new products and processes that can be patented through disposable agencies; and;
(d) An institutional environment that makes efforts to reduce negative externalities from pollutants to correct market failures.

[1] In more simple terms, the process for obtaining second-generation ethanol consists of "breaking" the lignocellulosic plant material (which may be done physically or through chemical or enzymatic reactions) to obtain the cellulose. In this case sucrose is obtained, and one of the destinations is the production of ethanol. To convert lignocellulosic materials into other products the following steps must be performed: (1) pre-treatment of lignocellulosic material in order to increase the exposure of the pulp fibers, facilitating the action of acids or enzymatic hydrolytic agents; (2) use of enzymes from microorganisms such as fungi and bacteria, obtaining sugars by the enzymatic hydrolysis process; and (3) fermentation process of the sugar mixture. See for more details, Brown and Brown (2012), Lee (1997), Sun and Cheng (2002) and Rabelo (2010).

Considering these four aspects, the third will be the focus of this paper, since it is not possible advance to increase the commercial production of second-generation ethanol without disposable knowledge and technology. The research activities are composed complex mix of general purpose technologies like molecular biology and technologies based on problem solving devices, called "local technologies" (Antonelli 2003), as enzymes dedicated to specific processes to the largest producer of conventional ethanol in the world.

The local availability of raw materials modulate the processes of exploration and exploitation of R&D in each country. For instance, the US seek mainly to use the straw and corncobs for the production second-generation ethanol. Brazil, second largest producer of conventional ethanol, aim to use straw and bagasse from sugarcane, byproducts (or residuals) from first-generation ethanol.[2] Some countries follow the "fast pyrolysis route" (Brown and Brown 2012), other countries, enzymatic convert biomass using a broad range of materials including crop residuals, wood and derivatives and waste (HLPE 2013). The combination of general and local technologies motivates the cooperation between countries and their research groups, what has generated the motivation of this paper.

The next section present a short review of the literature related to our methodological option to apply network analysis to understand the emergence of second-generation of biofuels in an international perspective. The third section describes the methodology used followed by the results and discussion in the fourth section, covering views on collaborative networks by countries, institutions, Keyword Plus and citations. The final section draws conclusions and highlights implications for policy and energy research.

Innovation Systems and the second-generation ethanol

The approach of Innovation Systems (IS) has emerged in the 80's, with their diffusion in England, Denmark and other countries. This theoretical perspective allows a better understanding of how occur the processes of acquisition, use and dissemination of knowledge as well as how productive and innovative capabilities are developed and generated. Once the innovation process is cumulative, it means, depends on endogenous capacities and tacit knowledge, the innovative capacity of a country or region stems from the relationship between economic, political and social agents. In this case, innovation means the processes that firms use to introduce and spread new products and processes (Dal Poz 2006).

The triple—University-Government-Enterprise—permeates characteristics for IS under different sights described by Carlsson and Stankiewicz (1991); Freeman (1995); Freeman and Soete (2008); Foray and Lundvall (1996); Patel and Pavitt (1994); Mytelka and Farinelli (2000); Edquist (2001) and Etzkowitz and Leydesdorff (2000). Each of the described literature have such specific sights, but, they agree on the importance of an interaction among the triple—University-Government-Enterprise—as a role in the development of a knowledge based economy.

This paper uses a general concept of IS that focus on the University as an important force to determine the success of a knowledge-based economy, specifically for the bioenergy sector (second-generation ethanol production).

[2] Consists of obtaining ethanol through a fermentation and distillation process from disposable and significant sugars that are in the plants. The main commercial crops are sugarcane, maize, sugar beet, potato and wheat.

This article proposes the use of the evolutionary approach of the economics of innovation and the Social Network Analysis (SNA) as a measure to understand and compare the different Innovations Systems based on lignocellulosic ethanol. There are two basic approaches combined in the paper: bibliometrics and scientometrics. The use of bibliometric approach and scientometric indicators, as a proxy of the science generated in a specific sector from the international scientific papers, allow to understand the development of the IS based on the existing relations among authors direct and indirect evolved to second-generation ethanol research.

Bibliometrics and Scientometrics approach: linking scientific publication to technology

The conception of intellectual connections between the ideas of scientists is established through social relations, and in certain areas, the collaboration is crucial to development. The scientific and technological production is directly related to its agents. Such as, they behave and relate how to give to the organization and how to transmit information among themselves (Wagner and Leydesdorff 2005).

There is a robust evidence showing that the scientific collaboration—authors of co-authored publications and citations that reference other authors in their publications—are positively correlated with the diffusion of scientific knowledge. Therefore, scientific collaboration is, from this perspective, a very good proxy of innovation (Wagner and Leydesdorff 2005). This measures—by means of indicators—the knowledge recorded through articles published in scientific journals and patents registered in institutions dedicated to organizing this information (Glänzel and Schubert 2005). Such documents registered and certified by the community, which produced, serve as relevant and useful indicators to understand the qualities of such knowledge, in terms of its relations with the developing countries.

According to Geisler (2000), to evaluate science and technology, there is only one viable method: the measurement of indicators. The verifiable data measured are represented by indicators such as: (a) incidence of articles in a given period of time; (b) publication of specific areas, measurement of impacts and influences on literature; (c) authorship and collaboration, incidences of citation; (d) relationship between citations and audiences, such as for example, the interdisciplinarity; and (e) classification, and area, as well as the characteristics that are observed in patents to evaluate the production of technological innovation.

The bibliometrics and scientometrics studies have specifications, approaches and different roles among themselves, and in this respect, their application depends on the objectives to which they relate. Bibliometrics and scientometrics, for example, differ in relation to the object and purpose of each application: the objective of the bibliometrics is to study the books or journals, in order to understand the activities of science of information. The aim of scientometrics is to study the quantitative aspects of the creation, dissemination and use of scientific and technical information and the objective is based on understanding search engines such as social activity.

According to Konur (2012), there was few “full-scientometrics” study published on bioenergy production using biomass, despite efforts, which began in the 1980s, to the development of the scientific knowledge necessary for second-generation technology. This fact highlights a gap in studies on the subject.

The models of information delivery specify how content and lexical representations of documents are intrinsically related (Croft 2000). According to Zitt and Bassecoulard (2006), the design of technological fields is essential for studies of decision support, to evaluate the positions of the industrial development of institutions or countries, in order to understand the dynamics of S, T&I and the strategic position of certain actors.

Geisler (2000) advocates the use of multiple indicators in order to increase the chances of understanding the many complex aspects of science and technology, which this work fully assumes. For Leydesdorff (2001) the constant search for a theory of citation in quantitative studies of science itself can be considered as an indicator to explore more systematically the relationship between the use of scientometric methods and qualitative approaches. The lexicographical order analysis of content of texts—articles and patents, applied under analysis of keywords, can illustrate the dynamics of emergency—or consolidation of particular field of studies (Bonaccorsi and Thoma 2007). Katz and Martin (1997) assert that the scientific collaboration is crucial to the development of generic technologies, such as biotechnology and new materials. This remarks point the opportunity and convenience to combine both in a methodology dealing with bioethanol.

At this point it is important to highlight that the activity regarding science, is a complex social construction, which increasingly partnerships and work carried out in collaboration, reflecting the very evolution of the scientific framework. In this scenario, the role of national and international collaboration is a requirement of modern science of quality and one of the main factors of success in scientific research (Leclerc et al. 1992).

Social Network Analysis applied to Scientometry

The methodology of this study is based on an analysis of networks applied to citations from scientific articles in the field of bioenergy. According to Freeman (2004), a network analysis is characterized as structural when it is possible to analyze the links between objects of a study; in the case of this research, citations from scientific articles in the field of bioenergy. The four elements that characterize this area of scientific knowledge can be applied to this research study.

The first element, the structural character of network analysis, makes it clearly possible to study the multidisciplinary character of bioenergy research: advancements in knowledge and how they are applied to influential technologies depend on the progression of various areas of knowledge, found in distinct research organizations and countries.

Contrary to many areas of economics, in which data is scarce and not made public, databases on scientific publications and those to identify citations are available, thus conform to the characteristics of network analysis, i.e. they are based on empirical data and can be adapted to various theoretic visions, as discussed at length by Freeman (2004). The complexity of links that form the networks requires the possibility to utilize not only indicators, but also systems of graphic visualization that can be easily manipulated and interpreted, precisely due to the application of mathematical and computational models, the two other characteristics of network analysis.

Recent evolutionary literature on economics points the importance of networks to knowledge acquisition (Morone and Taylor 2010; Saviotti 2009). Therefore, the process of building knowledge will occur as faster as the information can be shared among the authors. If considered that to make a scientific paper have to evolve a sort of co-authors, and they spread the knowledge in other papers with other co-authors, the knowledge could grow faster as the links within authors inside the network emerges.

In terms of aspects of scientific collaboration, Wagner and Leydesdorff (2005) studied the determinants of the rapid growth of international scientific collaboration, the authors adopted the hypothesis of the behavior of networks of collaboration that is described by the behavior of the preferential attachment, which is based on the reputation and rewards. The actors decide to collaborate to gain visibility, reputation, complementary capabilities and access to resources.

Wagner and Leydesdorff (2005) analyzed the scientific international publications structured in knowledge networks—in which authorship and co-authorship play the role of links between these actors. The choice of individual scientists to cooperate is motivated by a reward structure, where co-authorship, citations and other forms of professional recognition lead to additional work and the reputation to a virtuous circle. Researchers with high visibility and productivity, who are able to choose, work with people who are more likely to increase their productivity and credibility. These same authors show that the activities of scientific collaboration is a self-organized activity, and that such collaboration is highly correlated with innovative activities, which collaborates to generate trends leading to innovation in bioenergy, which is the focus of this this article proposes (Wagner and Leydesdorff 2005).

Since research on bioenergy begun, it has been linked to policy formulation. In the last 20 years, themes related to the exploration of technological opportunities have included discussions about the impacts on the environment and the price of food. One prominent result of this discussion was the emphasis on the importance of scientific research and technology in all fields of alternatives related to renewable energy, as pointed out by Rausser and Papineau (2008). In order to formulate policies, it is important to have more than a surface-level understanding of the nature and interconnection of disciplines related to development in bioenergy.

Getting closer to the network approach, the empirical study of production in S, T&I of second-generation ethanol is based on methods of document clustering, and on their relations intrinsic grouping of words. The groups are obtained by algorithms of frequency and similarity of presence of words in different documents. This type of tool is used as the basis for clustering the concepts of Han and Karypis (2000), who demonstrate that words that occur with high frequency in a set of words of texts summarize the center of a cluster. The majority of approaches to clustering are based on models of spatial vectors, which orientate to supplement the frequency and distance of words that are similar or identical, in collections of texts.

Shibata et al. (2008) highlight the methodologies that not only support policy makers, but also detect new and emerging areas. These methodologies focus more on the emergence of paradigms and technological trajectories, and less on the analysis of pre-existent trajectories. They define three groups: (1) an approach based on the analysis of texts, (2) data mining (DM), and (3) database tomography (DT), "that assisted the forecaster to identify the taxonomic structure of research-domain" (Kajikawa and Takeda 2008).

Studies such as those carried out by Kajikawa and Takeda (2008) and Souza (2013), which use the DT methodology, are based on keywords and capture the wide range of areas that encompass bioenergy. This makes it possible to identify which of the segments form specific clusters, that form dense networks of article citations, and which of these are related to other areas of knowledge. These networks cover a large number of publications and focus on the "age" of the studies that make up the clusters, which are more important when they can be clearly characterized as forming areas of interest, such as the biofuel cluster, and the sub-clusters (derivatives of clusters) pyrolysis and lignocellulose, among others.

There are articles that, benefitting from the advancements in scientometrics, limit the range of keywords (by using sub-clusters, as identified by the studies above). These articles focus on characterising the areas in terms of countries that carry out a greater deal of research, and in terms of the identification of cooperation between them and their mechanisms. Regarding the field of bioenergy, it is worth citing the work of Konur (2011) and Konur (2012). The first is based on an emerging area, which uses microalgae to produce bioenergy and derivative products, and the second, focuses on biomass.

The present paper, focused in second-generation has a similar results as those obtained by Liu et al. (2014), that used a broader range of keywords. These keywords include fields of investigation that, in focusing more on petrochemical derivatives (such as the production of ethanol from syn-gas), move away from the areas related to biotechnology and the use of biomass, thus resulting in a vision that focuses too heavily on the research and cooperation in bioenergy.

They conclude that "academic research publications in this area have grown dramatically over the last two decades" However they also conclude that "publication activity in most countries is still dominated by domestic research" (Liu et al. 2014). They also conclude that scientific activity in bioenergy is progressively scattering around the world and have stressed the role of Academy of Sciences in centralizing bioenergy research in China, the country that is now in second position, following the USA in number of publications, but far from the leader when impact is considered.

Methodology

This paper uses both the Bibliometry and Scientometry approach and the Innovation System literature under the perspective of Social Networks Analysis to build Collaborative Networks to the second-generation ethanol subject.

The methodology used assumes that the combination of techniques to search for information[3] can increase the efficiency of the method, when the task is to categorize the content of texts (Lewis and Hayes 1994; Larkey and Croft 1996). The methods of delineation and mapping (Laurens et al. 2010) can be strengthened by the use of hybrid approaches, including those that involve the cooperation of experts in the fields investigated are essential. It is evident that the reliability of the results when adopting this methodology depends on adequate the choice of indicators is in relation to the aspects that we want to measure, the level of aggregation of the data and the relevance of the implemented operations.

Social Network Analysis

Social Network Analysis (SNA) emerges from graph theory. A graph (or network) is composed of three basic elements:

1. Nodes are people or groups of people who come together with a common goal. Visual representation in the units of analysis can be actors, elements, countries, research institutes, companies, friends, papers, etc.;
2. Edges indicate the interactions or links between two or more nodes, i.e., connecting two adjacent vertices. In a network with n players, one particular node can have links $(n - 1)$; and
3. Flow indicates the direction of the bond that is with an arrow which may be unidirectional or bidirectional.

From the networks created with a specific objective, the indicators for the network that requires further interpretation can be obtained:

[3] Using the same queries for both approaches—bibliometrics and scientometrics—with the content analysis of scientific papers.

1. Average Geodesic Distance: the geodesic distance (or social distance) is an indicator of network cohesion. Defined as a minimum number of links (or edges) that separates two distinct actors in a network. That is, given the shortest path between two nodes, the length of this shortcut in number of intermediate links, is called geodesic distance. This indicator is used as an indicator of specificity, i.e., the further two actors are, the smaller the connection is between them. In networks with a smaller distance between the actors, cohesion is greater, i.e., there is a greater bond strength between the actors, and the information is passed more quickly. In this case, geodesic distance is a clear proxy of authors' connectivity, showing that the more important the collaboration in science is, the closer to technological diffusion the knowledge is;
2. Average Density: The density of the network indicator measures the relative amount of existing connections. Is also a network cohesion indicator. Networks are considered dense if there is a high number of links between actors and considered sparse if there are few links. This indicator is the proportion of ties that occurs in relation to all possible links. It makes it possible to analyze the intensity of relations between actors (weakness/strength) in a network. A fully connected network is called a click and has a specific gravity of 1. If the network has no links, it is called empty and its density is zero; and
3. Average Centrality Degree: The centrality measures assist in verifying the relative importance of a vertex in a network. In this case this indicator is specific and allows verification of the centrality of the actors. The degree of centrality measures the number of actors to which an actor is directly linked. In this case a decision must be made: to connect an actor or not. Thus, the degree of centrality can be analyzed as: In-degree: the sum of the interactions that we have with the other actor; Out-degree: the sum of the interactions that the actor has with others; and Degree to a symmetric network: i.e., when the relationship between the different actors is reciprocal, the adjacency matrix is symmetric and therefore the degree of input is equal to output and represents the sum of interactions that actors have with others (the matrix of interactions adds to the row or column specific to the actor in question).

The following section describes the procedures for constructing the database.

Database

Data collection

A first-step is regarding the election of the keywords of the research parameter, to construct the research parameter, a bibliometric procedure, trying to combine words selected from the state-of-art literature in lignocellulosic ethanol, and expected to be present in many scientific articles, like energy, and specific words like enzymatic hydrolysis. After, adopting an analysis of the words from the perspective of three panels of experts[4] in bioenergy, particularly in bioethanol, to elect the essential words to construct the research parameter.

A second-step is performing the scientometric approach. Zitt and Bassecoulard (2006) and Glänzel and Schubert (2005) show that the use of scientometric approaches can be effective for analyzes of macro-level, as the adopted here for and "energy" and specific words, related to "second-generation ethanol", since the system of delivery of information is quantitative, a social network based analysis.

[4] The panel sessions were held in a meeting coordinated by BIOEN-Research team during the early 2012 with researchers from bioenergy research centers from the University of São Paulo, State Paulista University (UNESP), State University of Campinas and CTBE.

RESEARCH PARAMETER
TS=(*ethan* OR *energ*) AND TS=(*sugar* OR *cane* OR bagas* OR straw* OR cogener*) AND TS=(*conversion* OR *lign* OR *cellul*) AND TS=(*hydrolys* OR *ferment* OR *enzym* OR fung* OR *bac* OR *pressur* OR steam* OR chem* OR sacch* OR microb* OR clostrid* OR thermocell* OR *spor* OR *cocc* OR erwinia* OR strept* OR sclerot* OR phaneroch* OR trichod* OR asperg* OR schizoph* OR *penicill* OR SCP OR "Single Cell" OR *xyl*) Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years Lemmatization=On

Fig. 1 Research parameter for scientific publications in selected topics: keyword, title and abstract. *Source*: ISI (2012). 6053 papers retrieved until 24th october, 2012

A search of scientific publications at only one database, a multidisciplinary base of reference and with high level of international integration, in this study, the Web of Science, that is integrated to the ISI Web of Knowledge. This database has referential data that summaries all areas of knowledge.

Figure 1 shows the query (research parameter) used for searching the ISI WoS. The terms combines words (i.e. hydrolysis) and radicals of words (i.e. *energ*) to capture variations in written words.

Data procedure

To deal with the data obtained during the extraction process of the information it is necessary to use a program that can translate data previously extracted in a format (example, format '.txt' with no quotes) and then enables analysis or even export data filtered.

The VantagePoint[5] program makes it possible to import the previously information obtained from the ISI WoS through filters developed by that company. After importing the data, pre-cleaning took place that aims to extract information and even duplicate the grouping of terms that for some reason the original base had wrongly drafted.

With the filtered files you can create subfiles from some criterion (e.g., country, keyword, etc.). In which this new file only expresses the relations that have this criteria. This is interesting from the moment that there is a general and new base can be generated up to the researcher without the need to return to the original database.

The following procedures presented describes the construction of Collaborative Networks (CN).

For the construction of networks,[6] it is necessary to create the adjacency matrices that express some relation, i.e., from important relationships, it is possible to analyzable networks.

[5] The VantagePoint version 7. http://www.thevantagepoint.com/.

[6] For this procedure, three different programs were chosen: Microsoft Excel 2013—Data tabulation of The VantagePoint version 7 and exported to UCINET version 6 (import the data, graph building, analysis of indicators for networks and nodes and visualization of relationships among key stakeholders and Gephi version 0.82 beta—This program allows artistic visualization of networks. For the display the Fruchterman-Reingold algorithm was chosen as it best represented the data because of the enormous quantity of relationships. This algorithm represents a force-directed layout because it considers a force between any two nodes. In this algorithm, the nodes are represented by steel rings and the edges have springs between them. The attractive force is analogous to the spring force and the repulsive force is analogous to the electrical force. The basic idea is to minimize the energy of the system by moving the nodes and changing the forces between them (Fruchterman and Reingold 1991). Because of the enormous quantity of data there was a superposition of nodes, even using the algorithm, so the second-step was avoid the superposition of most relevant nodes of the network, setting to the nodes be placed at the bound of the sphere.

With these procedures, it was possible to express the following relationships in CNs:

1. Relationship between countries—International Collaboration Networks (macro level);
2. Partnerships between institutions, i.e., universities, government and enterprises—International Collaboration Networks (micro level);
3. Relationship between KeyWord Plus[7]—variable created by an algorithm of Thomson Reuters and allowing stress adjacencies between areas of knowledge formed from the second-generation ethanol, is interpreted as an indicator of the amount, or counter clockwise to search—KeyWord Plus Networks; and
4. The co-occurrence networks of authors and publications that appear in the references, indicate that the most relevant authors are the most frequent—Citations Networks.

The following section show the main results from the proposed methodology.

Results and discussion

The research on ISI WoS database shows that in more than 30 years, 103 countries took part directly or indirectly in the knowledge production to second-generation ethanol. Table 1 summarizes the first ten countries in number of scientific publications and respective share of papers in national and international collaboration.

Analyzing the scientific publications, the United States occupies a prominent position, with almost 23 % of total papers in second-generation and related areas belonging to this country, followed by China and Brazil (see Table 1). It is interesting to note that some non-traditional biofuel producers are doing research on new technologies on biofuel production.

Figure 2 represents the International Collaborative Network (macro level) for countries based on scientific publications. It appears that in terms of scientific collaboration the United States predominates, followed by Germany, France, United Kingdom and Sweden. Although China, Japan, India and Brazil present the largest number of the publications, they are less connected with other countries than others.

Table 2 shows the indicators of the International Collaborative Network (macro level) based on scientific publications for each country. The average centrality is 0.09466 (9.466 %). This value is considered such lower when analyzing the collaborative network, which indicates that no country has a relationship with all other countries, i.e. all share the condition of centrality, which is expected according to the cited literature. The profile of the collaborative network also shows that there is no single central country, but a range of countries that share this feature, facts that confirm the visual analysis.

The average density measures the proportion of bonds that occur between countries in relation to all possible links. A fully connected network is called a clique and has a specific gravity of 1. In this case, the density was found to be 0.093, which is low, but expected for this case. It is not uncommon to suppose that only a share of links will occur in this kind of network. Not all researchers will work together, but they can compete to publish a paper or some barriers like geographical, education, etc. can prevent the contact between authors.

Figure 3 represents the International Collaborative Network (micro level) for research institutions, universities and enterprises based on scientific publications. It appears that in terms of scientific collaboration USDA (United States) has predominance, followed by the University of California (United States) and USP (Brazil). The second group is made up of

[7] KeyWord Plus is a kind of automatic indexing used in the citation databases produced by ISI.

Table 1 Scientific publications in second-generation ethanol by country (national and international collaboration)

Country	Publications	(%) Total	National	(%)	Collaboration	(%)
United States	1559	22.44	1190	76.33	369	23.67
China	684	9.84	505	73.83	179	26.17
Brazil	347	4.99	257	74.06	90	25.94
Japan	332	4.78	249	75.00	83	25.00
India	299	4.30	250	83.61	49	16.39
Germany	290	4.17	169	58.28	121	41.72
Canada	275	3.96	178	64.73	97	35.27
United Kingdom	244	3.51	133	54.51	111	45.49
Spain	240	3.45	169	70.42	71	29.58
Sweden	231	3.32	116	50.22	115	49.78

Source: ISI (2012)

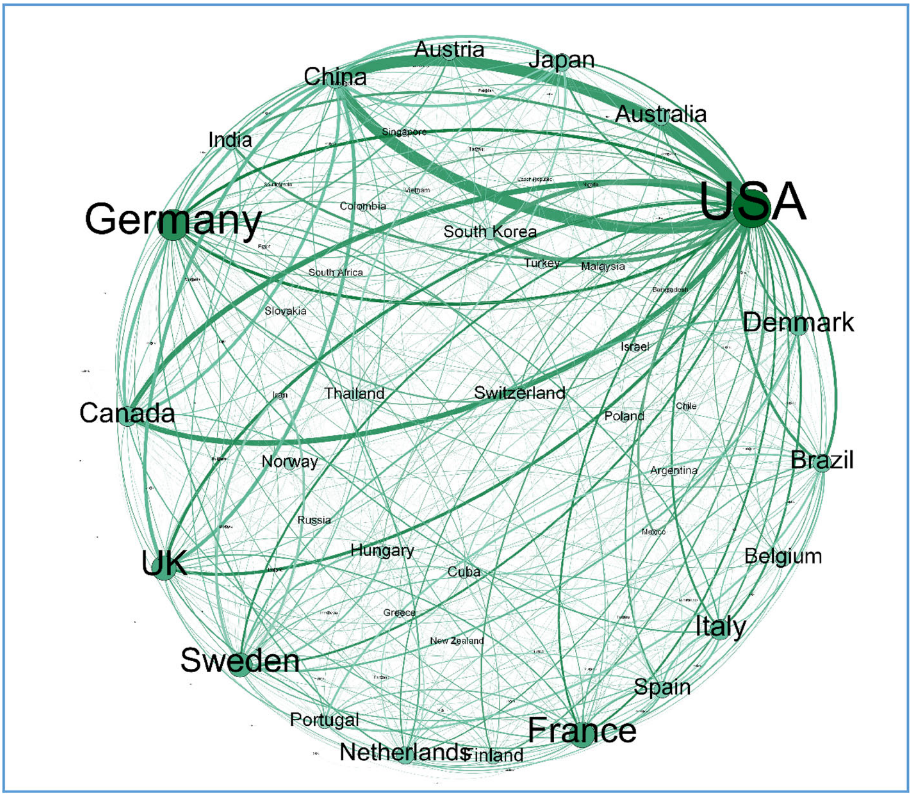

Fig. 2 International Collaborative Network of co-authorship between countries in second-generation ethanol

Table 2 Measure of International Collaborative Network of co-authorship between countries in second-generation ethanol

Index	Value
Average Centrality Degree	9.466
Average Density	0.093
Average Geodesic Distance	2.273

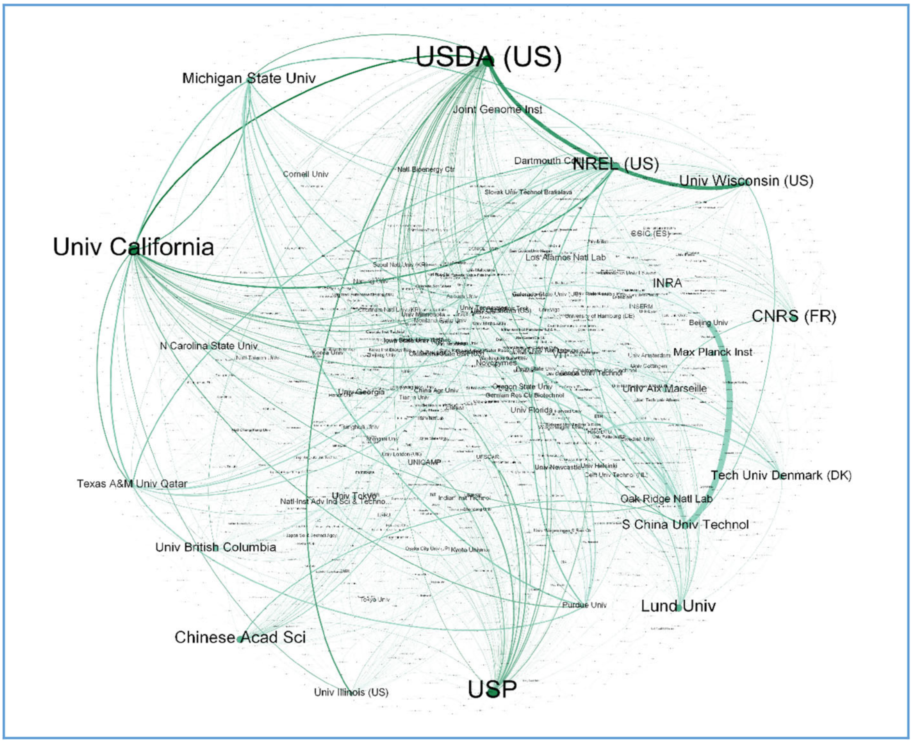

Fig. 3 International Collaborative Network of co-authorship between institutions in second-generation ethanol

NREL (United States), CNRS (the France), Lund University (Sweden) and the Chinese Academy of Sciences (China).

The North-American presence among the leading institutions is highlighted. It reveals the pro-active relationship between government and university, USDA (Government) and University of California (University). This combination is one of the indicators of success of an Innovation System. Brazil ranks second with the University of São Paulo, however, there is low government interaction (compared to the USDA relationship in the US) and the presence of other Brazilian universities in the top positions practically null (UNICAMP, another Brazilian University, occupies the 26th position in the ranking of scientific publications).

China's presence is directly connected to the partnerships that the United States are developing in second-generation ethanol with this country.

Table 3 Measure of International Collaborative Network of co-authorship between institutions in second-generation ethanol

Index	Value
Average Centrality Degree	3.776
Average Density	0.002
Average Geodesic Distance	4.512

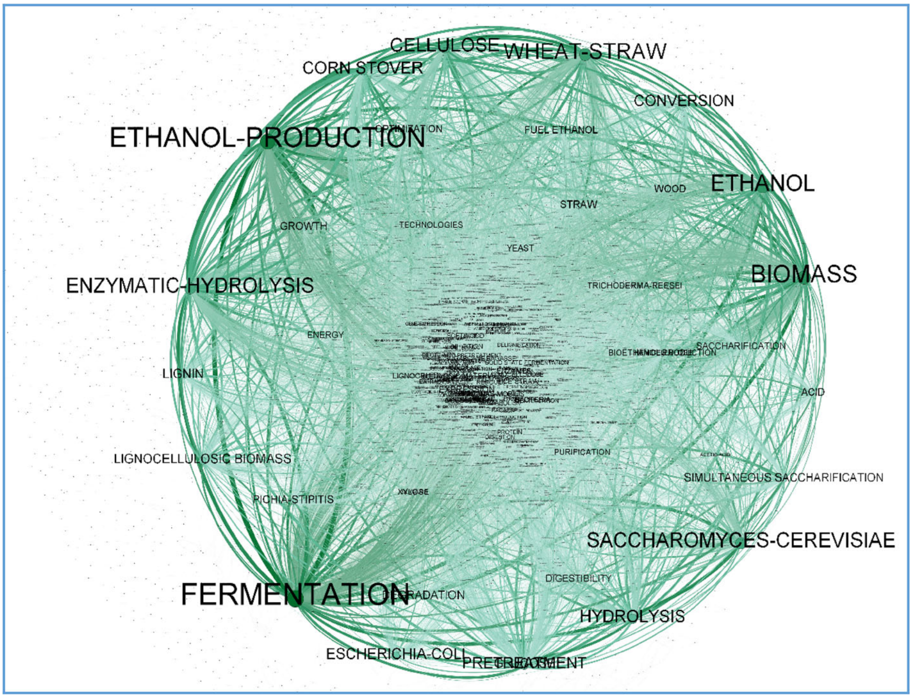

Fig. 4 Collaborative Network of KeyWord Plus from research database

Table 3 shows the indicators of the International Collaborative Network for institutions based on scientific publications. The average centrality is 0.03776 (3.776 %). This value is considered low because for the network of publications it is not expected that all institutions have a relationship with all others, i.e., all being central, as indicated by literature.

In this case, the density has proved to be 0.002 or lower, as expected. The relatively high number of institutions shows that there is a low-share of centralized institutions that are preferential or more prominent when compared to others.

Figure 4 represents the Collaborative Network for KeyWords Plus based on scientific publications that were emerged from the search using the research parameter. It appears that in terms of areas of research in second-generation ethanol, the most common ones are: Fermentation, Ethanol, Ethanol Production and Biomass (all areas except fermentation were expected by trimming the sample). The second group consists of Enzymatic Hydrolysis, *Saccharomyces-Cerevisiae* (fungus), Hydrolysis, Cellulose, Wheat Straw, Escherichia-Coli (bacteria), Corn Straw and Pretreatment.

This information corroborates the recent considerations about the commercial viability of second-generation ethanol. The main bottleneck nowadays is the fermentation of glucose and xylose obtained in the enzymatic process. There is also a debate on the best hydrolysis process, but results indicate that there may be a converging technology path for enzymatic hydrolysis.

The *Escherichia-Coli* is related to the enzymatic hydrolysis process. This bacterium (bacillus positive lactase) is present in the gut of humans and animals produces a sugar fermentation enzyme, and catalyzes the hydrolysis of lactose to glucose and galactose.

Other considerations shall be made with a focus on the use of some materials, such as wheat straw and corn processes for obtaining second-generation ethanol. As sugarcane is predominantly produced in Brazil, the methods chosen by the other countries are not specific to any raw material, but can generate Brazilian technological dependence on countries that develop and improve the most efficient routes.

Table 4 presents the indicators of the Collaborative Network for KeyWord Plus based on scientific publications from de database generated after search in ISI WoS with the research parameter described in the section "Data collection"

The average centrality was 0.08192 (8.192 %). This value is considered average for this network. This reflects that it is not expected that all areas of knowledge, spread among various subjects, linked to all the others, converges to one central theme, as ethanol.

The average density measures the proportion of links that occur between countries in relation to all possible links. In this case the density was found to be 0.002, or lower, as expected. The relatively high number of KeyWord Plus indicates that researches in this sphere are not centralized.

The geodesic distance to the value of 3.066 shows that the research areas are on average separated by KeyWords Plus 3, i.e., there is a close proximity of research areas, which are relevant. In a different case, a large distance shows that there is no convergence of efforts in second-generation ethanol.

Figure 5 presents the Collaborative Network for authors through the co-occurrence of these authors in scientific publications. It appears that in terms of scientific collaboration, the following authors are important:

1. Lee Rybeck Lynd—Associate Professor of Biological Sciences—Thayer School of Engineering at Dartmouth—United States;
2. Nathan Mosier—Associate Professor in Agricultural and Biological Engineering—Purdue University—United States;
3. Charles Wyman—Professor of Chemical and Environmental Engineering—University of California Riverside—United States; and
4. Sun Ye—Professor, Department of Biological and Agricultural Engineering, North Carolina State University—United States.

Thus, it can be deduced that state-of-art in second-generation ethanol comes exclusively from North-American authors.

Table 4 Measure of Collaborative Network of KeyWord Plus from research database

Index	Value
Average Centrality Degree	8.192
Average Density	0.002
Average Geodesic Distance	3.066

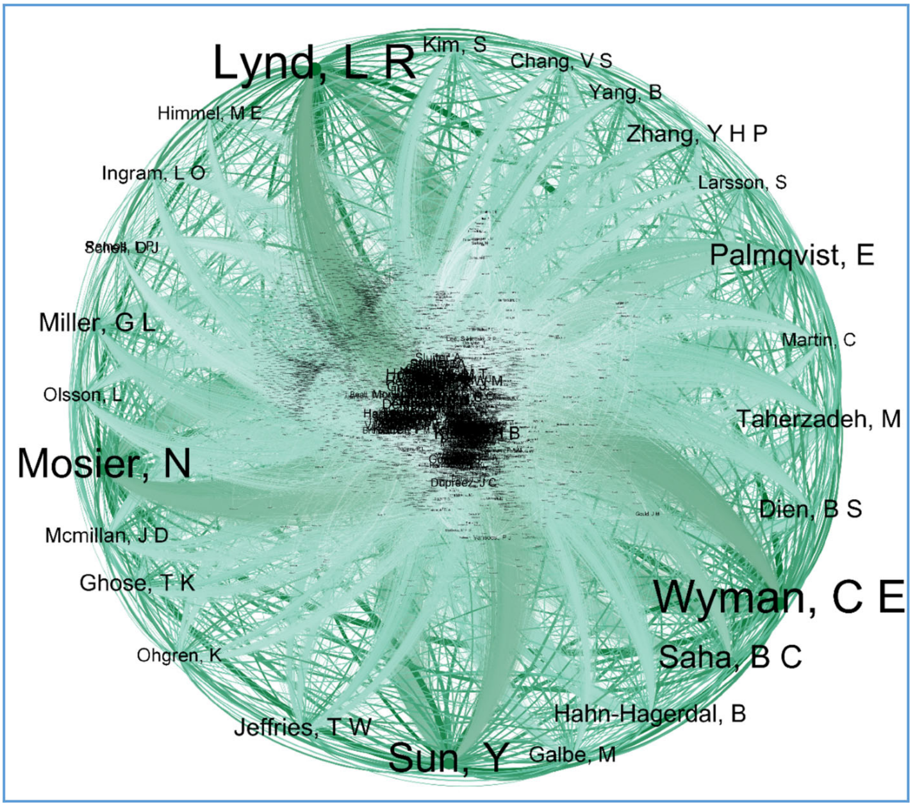

Fig. 5 Collaborative Network of co-authorship from citations in second-generation ethanol

Table 5 Measure of Collaborative Network of co-authorship from citations in second-generation ethanol

Index	Value
Average Centrality Degree	11.728
Average Density	0.002
Average Geodesic Distance	2.998

Table 5 shows the indicators of the Collaborative Network for authors based on co-occurrence in scientific publications. The average centrality was 0.11728 (11.728 %). This value is considered low for network publications because is not expected that all authors have collaborated with all others, i.e., all are central. It does however appear that some are in a more prominent position than others. These authors are linked directly to the process of converting biomass, specifically hydrolysis and pretreatment. These facts indicate the presence of preferential attachment, i.e., more experienced authors in the area with a high number of citations.

Conclusion and discussion

The commercial production of lignocellulosic ethanol as well as new products from this technological conversion will become a reality in the coming years. This will not only be motivated by the need for a replacement for petroleum and its derivatives, but also by the productive capacity and opportunities that the second-generation ethanol will bring to the countries with benefits resulting from such technology. However, leadership and productive capacities needed for insertion in this market require that countries to increase their competitiveness and deepen investments in innovation to address such advents.

This article aims to present the methodology of knowledge-based networks, a method that uses both a bibliometric and scientometric approach from Innovation Systems perspective and with the scope of the Social Network Analysis—Collaborative Networks—applied to the case of bioenergy, in particular, for the production of second-generation ethanol.

It was found that there is a well-defined technology, and that a commercial scale begins to take its first steps for some technological routes. However, the graphical analysis of the networks performed in this research, as well as the indicators described, reveal the role of the insertion in the network. It is unlikely that a country that is outside the network will be able to obtain improved results and feasibility before the world leaders in the research.

Even though the centrality indicator of network does not conclude that there is a leader in the network, the United States are apparently moving in this direction. The North-American programs such as RFS2 (Renewable Fuel Standard Program) from EPA (United States Environmental Protection Agency) require a large portion of advanced fuel that could be placed by lignocellulosic ethanol, even regarding the current reconsiderations undertaken by President Obama in 2014.

China proves a great ally of the United States in terms of the production of technology to produce lignocellulosic ethanol. Brazil however, does not appear well placed in the network, despite being the second largest producer of first generation ethanol in the world.

The set of indicators calculated has highlighted the major public and private institutions involved in the emerging areas, with the more relevant actors forming partnerships. These facts are directly related to the theory of Innovation Systems. It shows the development of Universities in the triple University-Government-Enterprise.

For a IS in second-generation ethanol to reach a high degree of development it is necessary that a stock of knowledge is transformed into patents. After this, innovations will be destined to the market.

Currently, the main research efforts of research emerge from Universities (Government and Enterprises have a lower role in this scenario) who transfer knowledge to other spheres.

The United States showed an important position in comparison to other countries in the development of knowledge in second-generation ethanol. They also have government programs that focus on lowering carbon emissions, using advanced fuels, partnerships with universities and more. Similarly, they have a higher stock of capital than other countries, as well as an important patents agency (USPTO).

These facts lead to the conclusion that the IS of United States based on second-generation ethanol is more highly developed when compared to other countries. In addition, to attain a state-of-art in lignocellulosic process, it is very important to collaborate with the US.

It can be concluded that the set of indicators was effective in the first approach employed in this article to identify the formation of scientific collaboration networks from the perspective of Innovation Systems.

To reach a complete understanding of the IS based on lignocellulosic ethanol it is also necessary to add the analyses of the innovation networks based on patents and the presence/absence of the Government in the interactions between the actors.

References

Antonelli, C. (2003). *The economics of innovation, new technologies and structural change (studies in global competition)*. Abingdon: Routledge.

Babcock, B. A., & Pouliot, S. (2013). *The economic role of RIN prices CARD policy briefs* (Vol. 13, p. 4). Ames: Iowa State University.

Bonaccorsi, A., & Thoma, G. (2007). Institutional complementarity and inventive performance in nano science and technology. *Research Policy, 36*(6), 813–831. doi:10.1016/j.respol.2007.02.009.

Brown, R. C., & Brown, T. R. (2012). *Why are we producing biofuels*? Ames, Iowa: Brownia LLC.

Carlsson, B., & Stankiewicz, R. (1991). On the nature, function and composition of technological systems. *Journal of Evolutionary Economics, 1*(2), 93–118. doi:10.1007/BF01224915.

Croft, W. B. (2000). Combining Approaches to Information Retrieval. In *Advances in information retrieval: Recent research from the center for intelligent information* (pp. 1–36). Kluwer Academic Publishers.

Dal Poz, M. E. S. (2006). *Biotechnology innovation networks: Genomics and intellectual property rights*. Campinas: Universidade Estadual de Campinas.

Edquist, C. (2001). *The systems of innovation approach and innovation policy: An account of the state of the art*. Paper presented at the druid conference, Aalborg.

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: From National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. *Research Policy, 29*(2), 109–123. doi:10.1016/S0048-7333(99)00055-4.

Foray, D., & Lundvall, B. A. (1996). *The knowledge-based economy: From the economics of knowledge to the learning economy*. Paper presented at the unemployment and growth in the knowledge-based economy, Paris.

Freeman, C. (1995). The 'National System of Innovation' in historical perspective. *Cambridge Journal of Economics, 19*(1), 5–24.

Freeman, L. C. (2004). *The development of social network analysis: A study in the sociology of science*. North Charleston: BookSurge.

Freeman, C., & Soete, L. (2008). *A economia da inovação industrial* (A. L. S. d. Campos, & J. O. P. d. Costa, Trans., 1 ed., Coleção Clássicos da Inovação). Campinas: Editora da UNICAMP.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience, 21*(11), 1129–1164. doi:10.1002/spe.4380211102.

Geisler, E. (2000). *The metrics of science and technology*. Westport: Quorum Books.

Glänzel, W., & Schubert, A. (2005). Analysing scientific networks through co-authorship. In H. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 257–276). Netherlands: Springer.

Han, E.-H. S., & Karypis, G. (2000). *Centroid-based document classification: Analysis and experimental results*. Paper presented at the proceedings of the fourth European conference on the principles of data mining and knowledge discovery (PKDD), Lyon.

HLPE. (2013). *Biofuels and food security*. A report by the high level panel of experts on food security and nutrition of the committee on World food security. Rome: FAO.

ISI. (2012). Web of science. http://thomsonreuters.com/web-of-science/. Accessed October 24, 2012.

Kajikawa, Y., & Takeda, Y. (2008). Structure of research on biomass and bio-fuels: A citation-based approach. *Technological Forecasting and Social Change, 75*(9), 1349–1359. doi:10.1016/j.techfore.2008.04.007.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1–18. doi:10.1016/S0048-7333(96)00917-1.

Konur, O. (2011). The scientometric evaluation of the research on the algae and bio-energy. *Applied Energy, 88*(10), 3532–3540. doi:10.1016/j.apenergy.2010.12.059.

Konur, O. (2012). The scientometric evaluation of the research on the production of bioenergy from biomass. *Biomass and Bioenergy, 47*, 504–515. doi:10.1016/j.biombioe.2012.09.047.
Larkey, L. S., & Croft, W. B. (1996). *Combining classifiers in text categorization.* Paper presented at the proceedings of the 19th annual international ACM SIGIR conference on research and development in information retrieval, Zurich, Switzerland.
Laurens, P., Zitt, M., & Bassecoulard, E. (2010). Delineation of the genomics field by hybrid citation-lexical methods: Interaction with experts and validation process. *Scientometrics, 82*(3), 647–662. doi:10.1007/s11192-010-0177-9.
Leclerc, M., Okubo, Y., Frigoletto, L., & Miquel, J.-F. (1992). Scientific co-operation between Canada and the European Community. *Science and Public Policy, 19*(1), 15–24. doi:10.1093/spp/19.1.15.
Lee, J. (1997). Biological conversion of lignocellulosic biomass to ethanol. *Journal of Biotechnology, 56*(1), 1–24. doi:10.1016/S0168-1656(97)00073-4.
Lewis, D. D., & Hayes, P. J. (1994). Guest editorial—special issue on text categorization. *ACM Transactions on Information Systems, 12*(3), 231.
Leydesdorff, L. (2001). *The challenge of scientometrics: The development, measurement and self-organization of scientific communications* (2nd ed.). USA: Universal Publishers/uPUBLISH.com.
Liu, W., Gu, M., Hu, G., Li, C., Liao, H., Tang, L., et al. (2014). Profile of developments in biomass-based bioenergy research: A 20-year perspective. *Scientometrics, 99*(2), 507–521. doi:10.1007/s11192-013-1152-z.
Morone, P., & Taylor, R. (2010). *Knowledge diffusion and innovation: Modelling complex entrepreneurial behaviours*. Cheltenham: Edward Elgar Publishing Limited.
Mytelka, L., & Farinelli, F. (2000). *Local clusters, innovation systems and sustained competitiveness.* Paper presented at the meeting on local productive clusters and innovation systems in Brazil: New industrial and technological policies for their development, Rio de Janeiro.
Patel, P., & Pavitt, K. (1994). Uneven (and divergent) technological accumulation among advanced countries: Evidence and a framework of explanation. *Industrial and Corporate Change, 3*(3), 759–787. doi:10.1093/icc/3.3.759.
Rabelo, S. C. (2010). *Evaluation and optimization of pretreatments and enzymatic hydrolysis of the sugarcane bagasse for second generation ethanol production*. Campinas: Universidade Estadual de Campinas.
Rausser, G., & Papineau, M. (2008). *Managing R&D risk in renewable energy.* Paper presented at the transition to a bio economy conferences, risk, infrastructure and industry evolution conference, Berkeley, California.
Saviotti, P. P. (2009). Knowledge networks: Structure and dynamics. In A. Scharnhorst & A. Pyka (Eds.), *Innovation networks: New approaches in modelling and analyzing (Understanding Complex Systems)* (p. 330). Heidelberg: Springer.
Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2008). Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation, 28*(11), 758–775. doi:10.1016/j.technovation.2008.03.009.
Souza, L. G. A. (2013). *Redes de inovação em etanol de segunda geração*. Piracicaba: Universidade de São Paulo.
Sun, Y., & Cheng, J. (2002). Hydrolysis of lignocellulosic materials for ethanol production: A review. *Bioresource technology, 83*(1), 1–11. doi:10.1016/S0960-8524(01)00212-7.
Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy, 34*(10), 1608–1618. doi:10.1016/j.respol.2005.08.002.
Wielen, L. V. D., & Breugel, J. V. (2014). Engineering the BBE: Drop-in or drop-out?. Ravello. Disposable on:http://www.economia.uniroma2.it/icabr-conference/Public/17/file/PPT2013/Auditorium/20-06-2013/5.00-6.00%20Plenary%20IV%20Santaniello%20Lecture/van%20der%20Wielen.pptx.
Willems, P. (2015). *Strategic direction in socio-economic issues*. Berkeley: Energy Biosciences Institute. Disposable on:http://www-app.igb.illinois.edu/conference/strategicdirections/speakers/paul-willems.ppt.
Zitt, M., & Bassecoulard, E. (2006). Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information Processing and Management, 42*(6), 1513–1531. doi:10.1016/j.ipm.2006.03.016.