A Dynamic Method to Identify New Fields of Scientific Research: an application to Bioenergy

Roney Fraga Souza[1,2], Sheila Cristina Ferreira Leite[1], Rosangela Ballini[2], José Maria Ferreira Jardim da Silveira[2]

[1]College of Economics, Federal University of Mato Grosso, Cuiabá-MT/Brazil
{roneyfraga, sfleite}@ufmt.br
[2]Institute of Economics, University of Campinas, Campinas-SP/Brazil
{ballini, jmsilv}@eco.unicamp.br

Resumo
A proposta do estudo é a detectar áreas emergente de pesquisa em bionergia usando um método baseado em medidas topológicas que se aplicam a redes de publicação científicas. O trabalho apresenta um esforço metodológico para tratar com dados baseados em trabalhos relacionados à bionergia utilizando palavras –chave sem a aplicação prévia de filtros. Os dados foram coletados do Instituto de Informação Científica (ISI) através do Banco de dados do Web of Science (WOS). Como resultado 244.307 artigos foram obtidos. Os resultados da pesquisa revelam que um agrupamento emergiu nos anos recentes, que todavia não é homogêneo do ponto de vista de seu conteúdo científico. Com isto, um novo processo de agrupamento foi realizado utilizando os trabalhos do agrupamento emergente (cluster). A análise mostra a obtenção de um sub-cluster que tem a propriedade de ser emergente e que está claramente identificado com bioenergia. Este sub-cluster emergente trouxe junto estudos de bioenergia e de biocombustíveis de terceira geração e temas de sustentabilidade da bioenergia, áreas de pesquisa de fronteira.

Palavras-chave: bionergia, agrupamentos, redes de citação, medidas topológicas

Abstract
The purpose of this study is to detect emerging areas of research on Bioenergy using a method based on topological measures applied to scientific publication networks. This paper presents a methodological effort to deal with a data base of papers related to bioenergy using key words without the definition of filters previously applied. The data was collected from the Institute for Scientific Information (ISI) through the Web of Science (WoS) database. As a result, 244,307 scientific articles were obtained. The results of the research reveals one cluster has emerged in recent years; however, it also shows that this is not homogenous. So, it was performed a new clusterization that considered only the papers from the emergent cluster. This analysis allows to achieve an emergent sub-cluster clearly identified with the themes of bioenergy. The emergent sub-cluster brings together studies from bioenergy and third-generation biofuel sustainability, two areas that also represent cutting-edge research.
Key words: bioenergy, clusters, citation networks, topological measures.
JEL : Q2;Q4;D85;O3.
AREA DA ANPEC: 9 Economia Industrial e da Tecnologia

1. Introduction

Energy Fields of Research are defined on the basis of a systematic application of the technical and scientific knowledge of agents who compete in a context of selective market processes, characterizing a science based area of economy (BELL; PAVITT, 1992). These processes take different trajectories and generate different patterns of technology diffusion within firms, between firms, among firms in the same industry, between industries and sectors in different countries, particularly USA and Brazil, the leading countries on ethanol production nowadays. (KAMM ET ALLI, 2008; BROWN; BROWN, 2013; HLPE, 2013).

This present study detects emerging areas of research using a method based on topological measures applied to scientific publication networks, which was originally proposed by Shibata et al. (2008), when the authors analyzed scientific articles on regenerative medicine. In another report by Shibata et al. (2011), the same method was used to analyze emerging areas of research on gallium nitride and in complex networks.

The tight connection between scientific and technological development and innovation justifies the importance of mapping the emerging areas of research of bioenergy (GLANZEL;SCHUBERT, 2004;PEREIRA; BAZI, 2009,). Efforts to detect emerging areas of research can be made using two alternative and complementary areas: a) experts in a given area of study; and b) computational techniques (KOSTOFF; SCHALLER, 2001). The use of specialist panels involves the difficulty that comes from the rapid growth of scientific knowledge, which causes the analysis of results to go through the lens of an analysis founded on a wider range of qualified information, a range that can also identify biases that are determined by the specialization of those who are interviewed. The other approach is based on computational techniques, which involve fewer problems that result from the increased amount of information, and which can complement a study performed with specialists.

The studies by Fontana et al (2009); Ribeiro et al (2010), Kraft, Quatraro, and Saviotti (2009), Saviotti, (2009) provide important clues for this investigation, but it was the observation of the results of the papers written by Dal-Poz et al (2012) and Silveira et al (2013) that motivated this study. They defined the area of bioenergy as an emerging area of research, but they based their opinion on distinct methodological procedures that were themselves based on patents. Their studies generated modest results in terms of the number of patents found with the use of filters (specifically in specific classes of patents). The methodological treatment based on Shibata et al (2008) and Shibata et al (2011) provided a viable alternative for avoiding any type of filter in the search process, and also permitted a more multi-faceted treatment of the citation networks. This treatment will become clearer later in the text.

The results of the research suggest a large growth in the number of studies in scientific areas that correspond to the key words that were used. The exploratory nature of the study means that a wide range of key words are used so that the area of bioenergy can be found and the areas of knowledge that it brings together can be better specified, without imposing any a priori restrictions that may exclude important papers. that with their removal, the networks would be destroyed (JACKSON, 2009). The methodology developed in this study is presented in Section 2. Section 3 presents the results obtained and the comments on these results. Section 5 presents the conclusions.

2. Methodology

2.1 Methodological Procedures

The methodology presented below is developed by the authors based on Shibata et al (2008) and Shibata et al. (2011).The following procedures are applied:
1) Collect data on citations from the Sciences Citation Index, compiled by the Institute for Scientific Information (ISI), which keeps databanks of citations that include thousands of academic journals and offer bibliographical database services. To perform this step, Web of Science (WoS) was used. It is an interface for the end user based on the web of citation databases from ISI; 2) Construct citation networks for each year, in which the articles are the nodes and the citations are the links, or connections, between the nodes; 3) Eliminate the isolate, it means, articles that do not cite other papers in the network, thus maintaining only the main components of the network. The goal of this step is to exclude papers that are not part of the network and which were part of the results because of the wide range of the terms used in the search; 4)Apply the Newman Algorithm (NEWMAN, 2004) to divide the network into groups and to identify each of the groups; 5) Visualize the citation network and determine the weight of each paper in the cluster and its coefficient of involvement in the clusters; extract the characteristic terms of each cluster using both a linguistic filter based on the abstracts of the papers and Natural Language Processing (NLP), (FRANTZI et al., 2000).

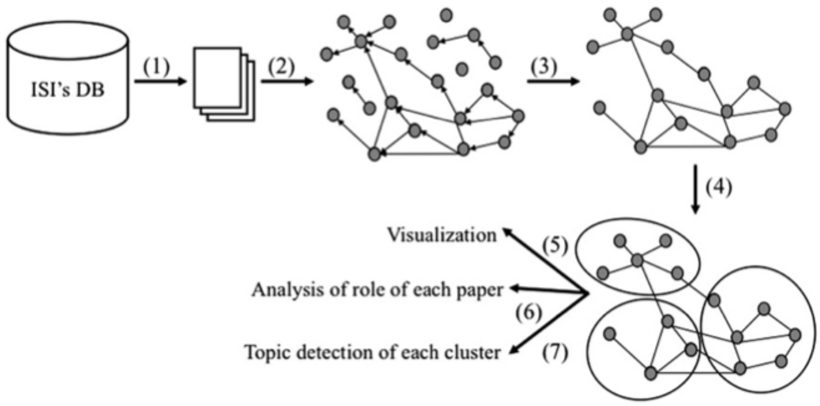

Figure 1 Methodology of the Paper:
Source: Shibata et al. (2011).

In step 5, the degree of importance of each paper (represented in the network by the nodes), is determined by the paper's relevance within the cluster (GUIMERA; AMARAL, 2005) represented by:

$$z_i = \frac{K_i - \overline{K_{s_i}}}{\sigma_{K_{s_i}}}$$

in which K_i is the number of connections of a node i relative to the other nodes in its cluster s_i, $\overline{K}_{s_i}$ is the mean of K over all of the nodes in s_i, and $\sigma_{K_{s_i}}$ the standard deviation of K in s_i,. Thus, the degree of relevance z_i of a node within a cluster is positive when the number of connections (degree) of the node has within the cluster is above the mean value.

To analyze a given term involvement in the clusters, which reveals the extent of the distribution of the connections of a given node i among the different clusters, the following equation must be used:

$$P_i = 1 - \sum_{s=1}^{N_M} \left(\frac{K_{is}}{K_i}\right)^2$$

in which K_{is} is the number of links of a given node i to the other nodes in the cluster s, and in which K_i is the total degree of the node i, or the number of links connected to the node i. A coefficient of involvement P_i close to one indicates that the connections are uniformly distributed among all of the clusters, while a coefficient close to zero indicates that all of the connections are within the node's own cluster, as shown by Guimera and Amaral (2005) and Shibata et al. (2011).The two indicators are used to define the position of the nodes in regions that define, for example, high relevance (increased weight) of the study (if it existed in an above-average number of citations) and an increased presence in a given cluster. The studies located in this region are indicators of knowledge that has an influence and is specialized. These factors are very important in cases with a large number of results in the searches performed using the key words.

To analyze the content of each cluster, Natural Language Processing (NPL) must be used. It allows for the detection of the research topic of each cluster. The terms are extracted from the abstracts of the articles using linguistic filters (FRANTZI et al., 2000). The filters extract the terms bound by elements that, when inserted in the sentence, make up a significant unit, thus maintaining dependent relations and order among the terms. Finally, it is important to note that the greatest challenges to compiling this information is in both the application of techniques of different areas of knowledge and the implementation of grouping methods, as well as of NLP.

2.2 Key Words and Data Sources

The data was collected from the Institute for Scientific Information (ISI) through the Web of Science (WoS) database. The collection was performed based on the following search criteria: (Bioconversion AND biomass) OR (Biofuel*) OR (Cellulos*) OR (Ethanol) OR (Lignocellulosic*) OR (Hemicellulose*) OR (Enzyme* AND Sugarcane*) OR (Fermentacion*) OR (Saccharification*) OR (Sugarcane* AND Delignification*) OR (Xylose*) OR (Hydrolisi* AND Enzyme*) OR (Hydrolisi* AND Cellulos*) OR (Hydrolisi* AND Ethanol). Papers from every year were accepted, but only the scientific articles were used. As a result, 244,307 papers were obtained.

The studies performed in the ISI via WoS can be saved 500 papers at a time; thus, 489 papers were obtained. The authors developed an algorithm in AppleScript language to automate the download of data. In order to bring together the files that were downloaded and to extract the data from the citation network, the Bibexcel software was used (PERSSON; DANELL; WIBORG; SCHENEIDER, 2009). To create the networks, graphs, and tables, the software known as R was used (R CORE TEAM, 2014), with the iGraph packages (CSARDI; NEPSUZ, 2006), ggplot2 packages (WICKHAM, 2009), plyr packages (WICHAM, 2011), Hmisc packages (FRANK; HARREL Jr, 2014) and tm packages (FEINERER; HORNIK , 2014).

3. Results

There has been significant growth in the number of papers obtained using the research criteria described in Step 1 of Section 2.1, particularly after 1990, as shown in Figure 1. Between 1945 and 2013, the volume of annual publications went from 55 papers per year to 18,406 papers per year. The amount of papers obtained using the research criteria described in Step 1 represents significant growth in the amount of publications over the years, as seen in Figure 2. Between 1945 and 2013, the volume of annual publications went from 55 papers per year to 18,406 papers per year.

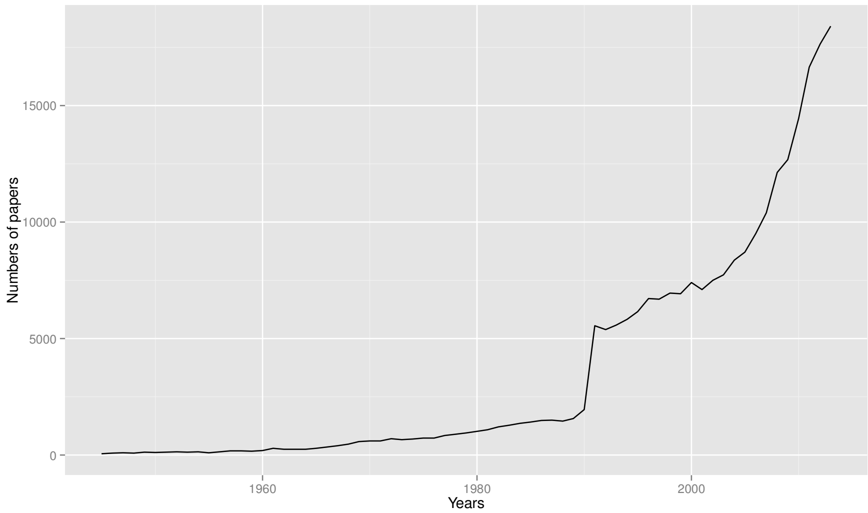

Figure 2 Evolution of the Amount of Scientific Articles Published Between 1945 and 2013

Source: research results

After Step 3 was performed, only the giant component of the network was maintained, and 193,555 papers remained. To perform Step 4 of applying an algorithm to divide the network into groups, the optimized version of the Newman algorithm (2004) was used (written by Clauset, Newman, and Moore, 2004). It is important to note that, in order to create a cumulative citation network, an article published in 1990 is considered part of the 1990 network, as well as the networks of all of the following years. Thus, a clustering process is performed for each year, and each article can change groups from one year to the next.

Figure 3 shows the result of the clusterization process, in which the clusters from 1995 to 2013 were maintained if they contained more than 500 papers. Circles represent clusters, and the size of each circle indicates the relative value of the number of papers in each cluster. The average age of each cluster indicates the average age of the papers that make up the cluster; thus, the lower the position of the circle in the coordinate plane, the newer the papers that make up the cluster, and the higher the value, the older the papers in the cluster are.

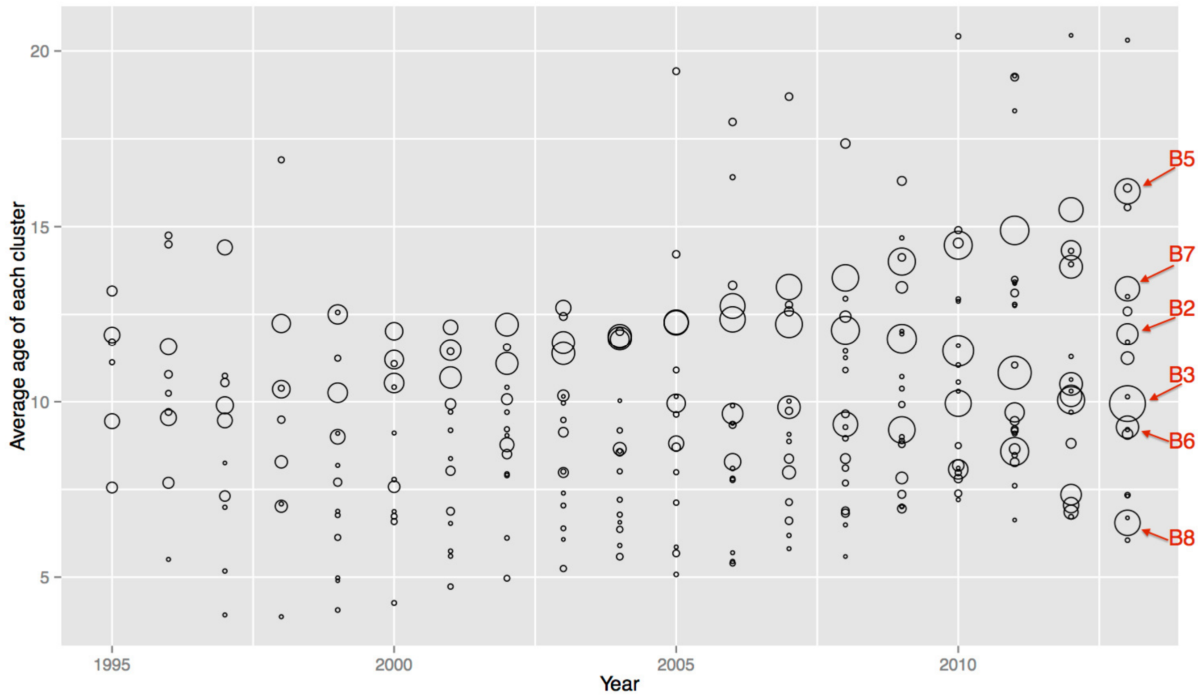

Figure 3 Average age of the papers that make up the clusters between 1995 and 2013
Source: research results

The B5 cluster is comprised of 24,725 papers. The average year of publication is 1997; thus, the average age of the papers in 2013 was 16. From 2005 on, an increase in the average age of the B5 cluster papers occurred; meanwhile, the average age of the papers in the B3 and B6 clusters decreased. An analysis of the trajectory of the B5 cluster indicates a high similarity among the papers included in the B5 cluster: that is to say, most of the papers in the B5 clusters in a given year would also be present in the B5 cluster in the following years. The composition of the B5 cluster indicates that the contributions of this cluster are based on established papers, or incremental innovations.

The B8 cluster is comprised of 25,083 papers. The average year of publication is 2006, and the average age of papers in 2013 is 7. The contributions of these papers that make up the B8 cluster tend to be based on more recent scientific articles, a finding which indicates the possibility for radical innovation. However, the point at which the cluster comes about is not clear, a factor which suggests that the cluster may be heterogeneous and may have papers from more than one area of research.

To visualize the network shown in Figure 4, the Gephi software was used (Bastian; Heymann; Jacomy, 2009) with the ForceAtlas 2 layout. Note that the B5 cluster (red) and the B7 cluster (yellow) present a high degree of connection that was already visible in 2002, a finding which indicates that they cover traditional areas of research. The B3 cluster (purple) is also visible in 2002; however, it presents a greater expansion than the B5 and B7 clusters. The B8 cluster (green), however, emerged more recently and can be clearly seen in 2013. It is also important to note that the papers in the B8 cluster are dispersed, which suggests low similarity among the citations of the papers that make up the group. This factor suggests that this cluster tends to be more heterogeneous than the others.

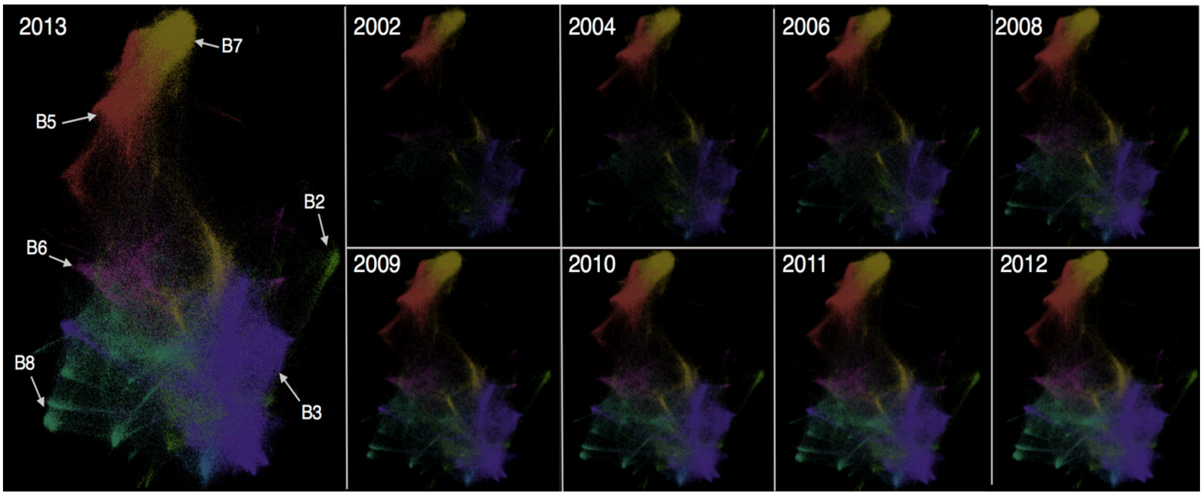

Figure 4 Network for 2013 with the aforementioned groups highlighted.

Source: Research Results

To determine whether the innovations are incremental or radical, the within-cluster degree (z_i) and the participation coefficient (P_i) were calculated for the 10 most frequently cited papers for the last 10 years. The idea is to know how well positioned the papers are in their own clusters and among clusters, given that vertices with the same role take on similar topological positions (GUIMERA; AMARAL, 2005). Areas of research with high z and P values suggest that the innovations are incremental, while a high z value and a low P value occurs in cases of radical innovations (SHIBATA et al., 2011).

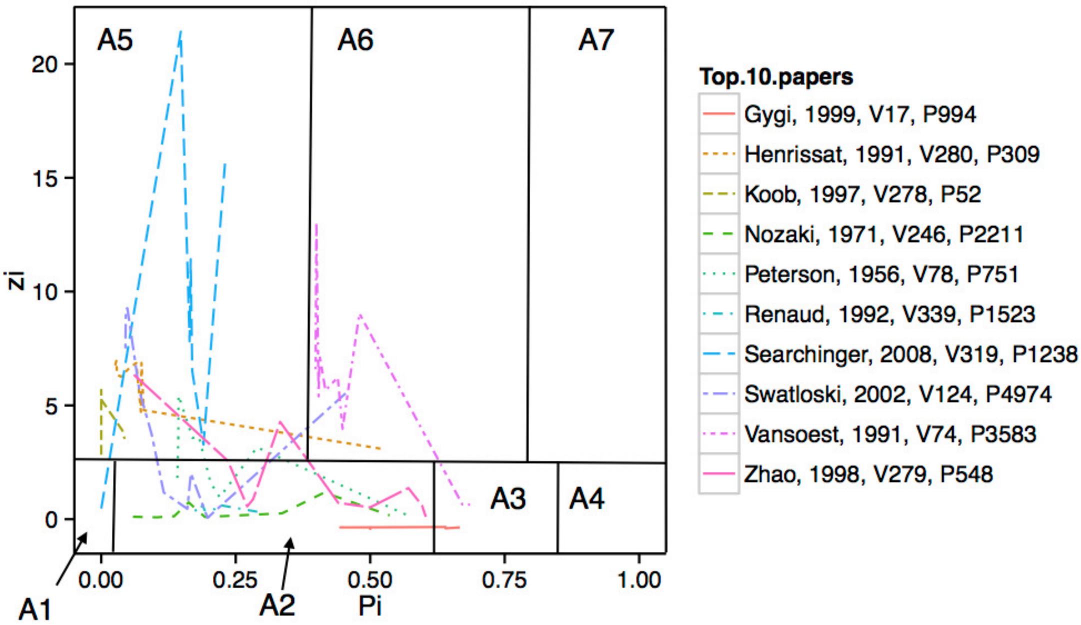

Figure 5 Role of each node in the topology.

Source: Research Results

The classification established by Guimera and Amaral (2005) suggests that the nodes with z<2.5 are hubs, while those with z<2.5 are considered non-hubs, as seen in Figure 5. The non-hub nodes are divided into four categories: (A1) outermost nodes, or nodes with most of their links within their cluster; (A2) peripheral nodes, or nodes with many links within their cluster; (A3) non-hub connecting nodes, or those with an increased proportion of links with other clusters; and (A4) non-hub kinless nodes, or nodes with edges that are homogeneously distributed among all clusters.

Figure 5 shows that many of the nodes are distributed among (A5), (A6), (A2) and (A3), and an explicit innovation pattern could not be identified. Because the amount of papers found in the search was higher (a factor which indicates that the field of study is wide), it is natural for space for incremental and radical innovation to occur simultaneously within the area of Bioenergy research.

The next point to clarify is the content researched in the papers in each cluster, which is investigated through analyses of the abstract of each paper, presented in the Table 1, below.

Table 1 A Summury of the Contento f Each Cluster

Cluster id	Number of papers	Average publication year	Terms frequently cited
B3	49149	2003	amino acid, cellulose membrane, cellulose acetate, ldl aphaeresis, affinity membrane, dextran sulfate, molecular weight, adsorption capacity, cellulose sulfate, bovine serum albumin, cellulose bead, bead cellulose, affinity chromatography, serum albumin, human immunodeficiency virus, low-density lipoprotein, dextran sulfate cellulose, composite fibre, protein kinase, molecular mass.
B5	24725	1997	xylose isomerase, amino acid, metal ion, active site, crystal structure, glucose isomerase, d-xylose isomerase, escherichia coli, enzyme activity, organic solvent, molecular mass, space group, water molecule, oxygen atom, molecular weight, optimum ph, hydrogen bond, amino acid sequence, specific activity, gel filtration.
B7	22898	2000	chiral stationary phase, mobile phase, stationary phase, chiral recognition, liquid chromatography, carboxymethyl cellulose, high-performance liquid chromatography, molecular weight, cellulose tris, metal ion, silica gel, chiral recognition ability, chiralpak ad, electron microscopy, acrylic acid, chiralcel od, chiral separation, cellulose derivative, chiral selector, sodium carboxymethyl cellulose.
B8	25083	2006	alcohol consumption, chronic ethanol, liver disease, lipid peroxidation, nucleus accumbens, alcohol intake, alcoholic liver disease, ethanol exposure, ethanol consumption, liver injury, nmda receptor, alcoholic liver, ethanol intake, alcohol dehydrogenase, oxidative stress, receptor antagonist, ethanol withdrawal, alcohol exposure, alcohol abuse, central nervous system.

Source: Results of the Research

For data on which papers are most frequently cited in each cluster, see Table 2.

Table 2. Brief Description of the Content of the Top 10 papers in each cluster of the Network

Cluster id	Top 10 papers	Times cited
B3	Vansoest, 1991. Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. Journal of Dairy Science, V74, P3583.	7004
	Henrissat, 1991. A Classification of glycosyl hydrolases based on amino-acid-sequence similarities. Biochemical Journal, V280, P309.	1838

	Swatloski, 2002. Dissolution of cellose with ionic liquids. Journal of the American Chemical Society, V124, P4974.	1470
	Farrell, 2006. Ethanol can contribute to energy and environmental goals. Science, V311, P506.	1033
	Dehal, 2002. The draft genome of Ciona intestinalis: Insights into chordate and vertebrate origins. Science, V298, P2157.	956
	Muhandiram, 1994. Gradient-enhanced triple-resonance 3-dimensional nmr experiments with improved sensitivity. Journal of Magnetic Resonance Series B, V103, P203.	664
	Verduyn, 1992. Effect of benzoic-acid on metabolic fluxes in yeasts - a continuous-culture study on the regulation of respiration and alcoholic fermentation.Yeast, V8, P501.	585
	Nishiyama, 2002. Crystal structure and hydrogen-bonding system in cellulose 1 beta from synchrotron X-ray and neutron fiber diffraction. Journal of the American Chemical Society, V124, P9074.	584
	Vansoest, 1968. Determination of lignin and cellulose in acid-detergent fiber with permanganate. Journal of the Association of Official Analytical Chemists, V51, P780.	563
	Reese, 1950. The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. Journal of Bacteriology, V59, P485.	527
B5	Gygi, 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature Biotechnology, V17, P994.	3117
	Renaud, 1992. Wine, alcohol, platelets, and the french paradox for coronary heart-disease. Lancet, V339, P1523.	1990
	Guengerich, 1991. Role of human cytochrome-p-450-iie1 in the oxidation of many low-molecular-weight cancer suspects. Chemical Research in Toxicology, V4, P168.	1075
	Lovinger, 1989. Ethanol inhibits nmda-activated ion current in hippocampal-neurons. Science, V243, P1721.	1062
	Browning, 2004. Hepatology, V40, P1387.	1031
	Ikonomidou, 1999. Blockade of NMDA receptors and apoptotic neurodegeneration in the developing brain. Science, V283, P70.	900
	Lieber, 1970. Hepatic microsomal ethanol-oxidizing system - in-vitro characteristics and adaptive properties in-vivo. Journal of Biological Chemistry, V245, P2505.	748
	Ikonomidou, 2000. Ethanol-induced apoptotic neurodegeneration and fetal alcohol syndrome. Science, V287, P1056.	688
	Decarli, 1967. Fatty liver in rat after prolonged intake of ethanol with a nutritionally adequate new liquid diet. Journal of Nutrition, V91, P331.	647
	Paceasciak, 1995. Fatty liver in rat after prolonged intake of ethanol with a nutritionally adequate new liquid diet. Clinica Chimica Acta, V235, P207.	621
B7	Koob, 1997. Drug abuse: Hedonic homeostatic dysregulation. Science,V278, P52.	1145
	Volpicelli, 1992. Naltrexone in the treatment of alcohol dependence. Archives of General Psychiatry, V49, P876.	1131
	Schaerenwiemers, 1993. Histochemistry, V100, P431.	939
	Crabbe, 1999. Science, V284, P1670.	808
	Komuro, 1993. Modulation of neuronal migration by nmda receptors. Science, V260, P95.	740
	Rudolph, 1999. Benzodiazepine actions mediated by specific gamma-aminobutyric acid(A) receptor subtypes. Nature, V401, P796.	691
	Imperato, 1986. Preferential stimulation of dopamine release in the nucleus-accumbens of freely moving rats by ethanol. Journal of Pharmacology and Experimental Therapeutics, V239, P219.	628
	Timpl, 1998, Nature Genetics, V19, P162.	591
	Majchrowicz, 1975, Psychopharmacologia, V43, P245.	556
	Suzdak, 1986, Proceedings of the National Academy of Sciences of the United States of America, V83, P4071.	536
B8	Zhao, 1998. Science, V279, P548.	5960

Searchinger, 2008. Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. Science, V319, P1238.	1207
Tian, 2007. Synthesis of tetrahexahedral platinum nanocrystals with high-index facets and high electro-oxidation activity. Science, V316, P732.	1106
Fargione, 2008. Land clearing and the biofuel carbon debt. Science, V319, P1235.	1025
Wan, 2004. Fabrication and ethanol sensing characteristics of ZnO nanowire gas sensors. Applied Physics Letters, V84, P3654.	973
Comini, 2002. Stable and highly sensitive gas sensors based on semiconducting oxide nanobelts. Applied Physics Letters, V81, P1869.	874
Hill, 2006, Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proceedings of the National Academy of Sciences of the United States of America, V103, P11206.	797
Yamazoe, 1991. New approaches for improving semiconductor gas sensors. Sensors and Actuators B-Chemical, V5, P7.	751
Williams, 2008. TiO2-graphene nanocomposites. UV-assisted photocatalytic reduction of graphene oxide. ACS Nano, V2, P1487.	744
Cambardella, 1992. Particulate soil organic-matter changes across a grassland cultivation sequence. Soil Science Society of America Journal, V56, P777.	722

Source: Research Results

As seen in Figure 3, above, the emerging cluster is the B8 cluster, with 25,083 papers, the average year of publication of which is 2006. Figure 3 shows that the B8 cluster has emerged in recent years; however, it also shows that this is not homogenous: the papers that make up this cluster may belong to subareas of knowledge within studies on Bioenergy.

To perform a better analysis of the content of the B8 cluster, a new clusterization was performed that considered only the papers from the B8 cluster. It can be seen in Figure 6. It is important to note the presence of four large groups: G1, with 4,756 papers; G2 with 3,212 papers; G3 with 1,969 papers; and G4 with 4,769 papers. As was mentioned previously, the network created by the B8 cluster presents a low degree of similarity, since the citations are more intra-group than inter-group, what is an interesting result.

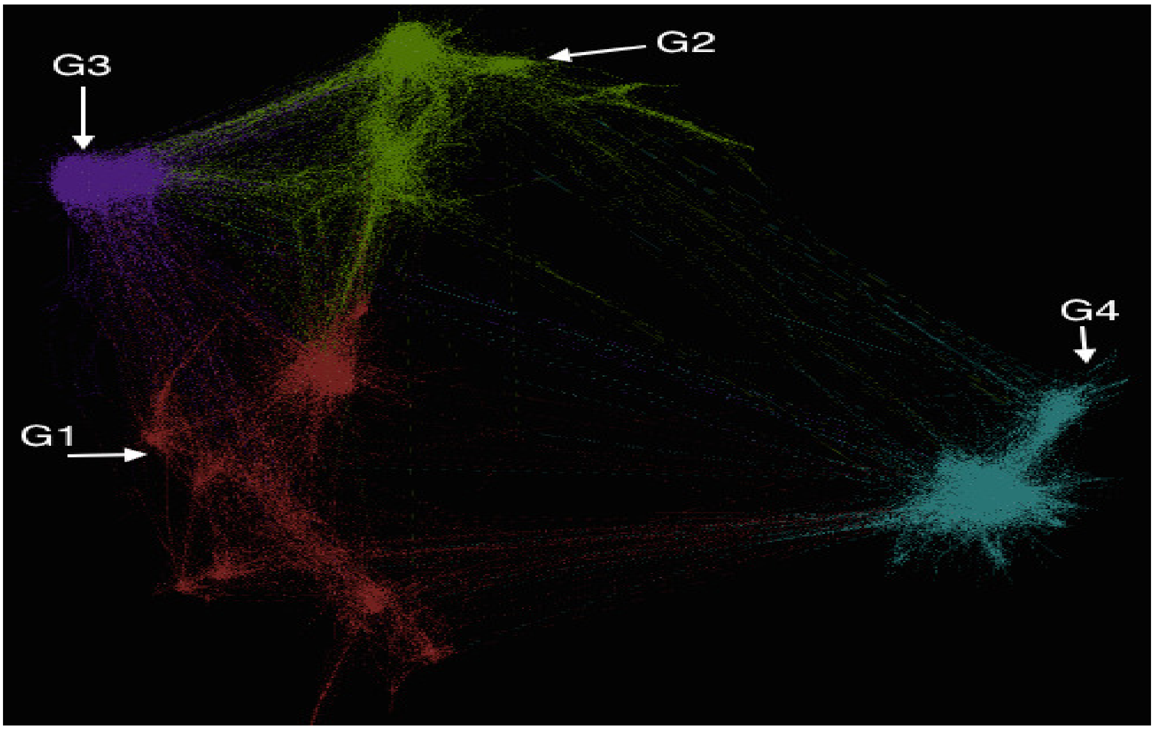

Figure 6 Sub-clusters of Cluster 8

To determine the content of each emerging sub-cluster, which are highlighted in Figure 6, the multi-word terms of the abstracts of the papers from the main clusters were analyzed. These main clusters represent a sub-division of the B8 cluster, as seen in Table 3.

Table 3 Information on each cluster. (TC =number of times cited or number of citation of the paper.)

Cluster id	Number of papers	Average publication year	Terms frequently cited
G1	4756	2006	electron microscopy, x-ray diffraction, transmission electron microscopy, gas sensor, surface area, thin film, transmission electron, catalytic activity, acetic acid, room temperature.
G2	3212	2005	biofuel cell, fuel cell, electron transfer, glucose oxidase, microbial fuel cell, power density, alcohol dehydrogenase, current density, maximum power density, carbon electrode.
G3	1968	2008	fuel cell, ethanol oxidation, direct ethanol fuel cell, transmission electron microscopy, cyclic voltammetry, electron microscopy, ethanol fuel cell, catalytic activity, electrocatalytic activity, transmission electron.
G4	4769	2010	fossil fuel, greenhouse gas, life cycle, biofuel production, greenhouse gas emission, ethanol production, climate change, ghg emission, gas emission, renewable energy, life cycle assessment, energy crop, environmental impact.

Source: Research Results

To aid in the analysis of the content of each cluster, see Table 4 for the most frequently cited papers in each sub- cluster of cluster B8.

Tabela 4 A Brief Description of the Contento f the Top 10 Papers in each sub-cluster of cluster B8

Cluster id	Top 10 papers	TC
G1	Wan, 2004. Fabrication and ethanol sensing characteristics of ZnO nanowire gas sensors. Applied Physics Letters, V84, P3654.	973
	Comini, 2002. Stable and highly sensitive gas sensors based on semiconducting oxide nanobelts. Applied Physics Letters, V81, P1869.	874
	Yamazoe, 1991. New approaches for improving semiconductor gas sensors. Sensors and Actuators B-Chemical, V5, P7.	751
	Chen, 2005. Advanced Materials, V17, P582.	661
	Meulenkamp, 1998. Synthesis and growth of ZnO nanoparticles. Journal of Physical Chemistry B, V102, P5566.	628
	Zhang, 2002. Control of ZnO morphology via a simple solution route. Chemistry of Materials, V14, P4172.	509
	Tian, 2005. Journal of the American Chemical Society, V127, P7632	442
	Lonergan, 1996. Array-based vapor sensing using chemically sensitive, carbon black-polymer resistors. Chemistry of Materials, V8, P2298.	394
	Greenler, 1962. Infrared study of adsorption of methanol and ethanol on aluminum oxide. Journal of Chemical Physics, V37, P2094.	350
	Dicosimo, 1998. Structure and surface and catalytic properties of Mg-Al basic oxides. Journal of Catalysis, V178, P499.	306
G2	Wang, 2003. Carbon nanotube/teflon composite electrochemical sensors and biosensors. Analytical Chemistry, V75, P2075.	581
	Zhou, 2009. Electrochemical Sensing and Biosensing Platform Based on Chemically Reduced Graphene Oxide. Analytical Chemistry, V81, P5603.	512
	Chaudhuri, 2003. Electricity generation by direct oxidation of glucose in mediatorless microbial fuel cells. Nature Biotechnology, V21, P1229.	509
	Rabaey, 2004. Applied and Environmental Microbiology, V70, P5373.	434
	Zhang, 2004. Analytical Chemistry, V76, P5045.	429
	Caccavo, 1994. Geobacter sulfurreducens sp-nov, a hydrogen-oxidizing and acetate-oxidizing dissimilatory metal-reducing microorganism. Applied and Environmental Microbiology, V60, P3752.	347
	Heller, 2004. Miniature biofuel cells. Physical Chemistry Chemical Physics, V6, P209.	346
	Gil, 2003. Biosensors & Bioelectronics, V18, P327.	311
	Rabaey, 2003. A microbial fuel cell capable of converting glucose to electricity at high rate and efficiency. Biotechnology Letters, V25, P1531.	296
	Cai, 2004, Analytical Biochemistry, V332, P75	285
G3	Tian, 2007. Synthesis of tetrahexahedral platinum nanocrystals with high-index facets and high electro-oxidation activity. Science, V316, P732.	1106
	Lamy, 2002. Journal of Power Sources, V105, P283.	476
	Teranishi, 1998. Chemistry of Materials, V10, P594.	470
	Teranishi, 1999. Size control of monodispersed Pt nanoparticles and their 2D organization by electrophoretic deposition. Journal of Physical Chemistry B, V103, P3818.	385
	Lamy, 2001. Electrocatalytic oxidation of aliphatic alcohols: Application to the direct alcohol fuel cell (DAFC). Journal of Applied Electrochemistry, V31, P799.	320
	Zhou, 2003. Pt based anode catalysts for direct ethanol fuel cells. Applied Catalysis B-Environmental, V46, P273.	318
	Lamy, 2004. Electrochimica Acta, V49, P3901.	292
	Vigier, 2004. On the mechanism of ethanol electro-oxidation on Pt and PtSn catalysts: electrochemical and in situ IR reflectance spectroscopy studies. Journal of Electroanalytical Chemistry, V563, P81.	244
	Toshima, 1993. Polymer-protected palladium-platinum bimetallic clusters - preparation, catalytic properties and structural considerations. Journal Of The Chemical Society-Faraday Transactions, V89, P2537.	233
	Wang, 1995. Evaluation of ethanol, 1-propanol, and 5-propanol in a direct oxidation	229

	polymer-electrolyte fuel cell - A real-time mass spectrometry study. Journal of The Electrochemical Society, V142, P4218.	
G4	Searchinger, 2008. Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. Science, V319, P1238.	1207
	Fargione, 2008. Land clearing and the biofuel carbon debt. Science, V319, P1235	1025
	Hill, 2006. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proceedings of the National Academy Of Sciences Of The United States Of America, V103, P11206.	797
	Cambardella, 1992. Particulate soil organic-matter changes across a grassland cultivation sequence. Soil Science Society of America Journal, V56, P777.	722
	Tilman, 2006. Carbon-negative biofuels from low-input high-diversity grassland biomass. Science, V314, P1598.	594
	Jespersen, 1987. Measurements of chlorophyll-a from phytoplankton using ethanol as extraction solvent. Archiv Fur Hydrobiologie, V109, P445.	375
	Schenk, 2008. Second Generation Biofuels: High-Efficiency Microalgae for Biodiesel Production. Bioenergy Research, V1, P20.	364
	Tilman, 2006. Biodiversity and ecosystem stability in a decade-long grassland experimente. Nature, V441, P629.	353
	Schmer, 2008. Net energy of cellulosic ethanol from switchgrass. Proceedings Of The National Academy Of Sciences Of The United States Of America, V105, P464.	352
	Crutzen, 2008. N2O release from agro-biofuel production negates global warming reduction by replacing fossil fuels. Atmospheric Chemistry And Physics, V8, P389.	310

Source: Research Resuts

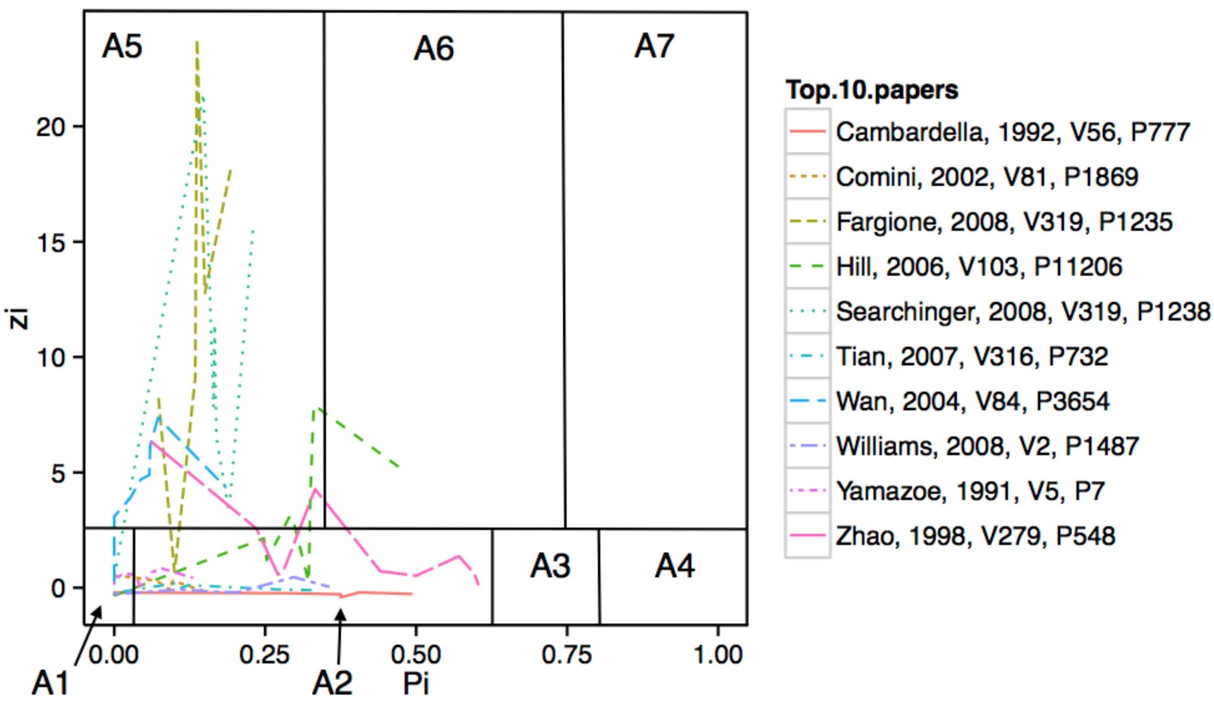

Figure 7. Role of each node in the topology from cluster B8

The analysis of the sub- clusters and of Cluster 8 reveals significant differences among them, as well as the cluster's adherence to areas on the front lines of bioenergy research. G1 strongly involves advances in the enabling technologies of processes, which are supported in radical innovations and connected to nanotechnology. G2 and G3 are close and involve developments in

the area of fuel cells, or, more specifically, motors, research which is also supported by nanotechnology. G4 brings together studies from bioenergy and third-generation biofuel sustainability, two areas that also represent cutting-edge research.
As was expected, the involvement values for the 10 most frequently cited papers over time (1945-2013) are lower than the values for the 10 most frequently cited papers in the complete network. This difference suggests greater involvement in the subclusters than in the B8 cluster. There are a greater number of points in the A5, A1, and A2 areas, which, according to Shibata et al (2011), indicates emerging technologies. The study by Zhao (1998), which is represented in Figure 7 by the pink dotted line, begins with a higher-than-average frequency of citation. The frequency then falls, goes up again, and falls again, thus coming close to the average citation frequency. However, it is spread throughout many clusters, which suggests that the topic of the paper is related to enabling technologies. Meanwhile, the paper by Seachinger (2008) reveals the capacity of the methodology developed for identifying papers that are cited in a specific cluster. Though they are frequently cited, they do not overlap into other clusters because the topic of study (in this case, bioenergy) is so specific.

4. Conclusions

This paper presents a methodological effort to deal with a data base of papers related to bioenergy using key words without the definition of filters previously applied. The methodology allows to a dynamic view of the emergence of the clusters, what is very difficult to perform with classical methodologies to obtain timelines (NOOY,MVAR.BATTLE, 2006; SILVEIRA et ALLI, 2013)).
It is also able to deal with a big number of papers and huge networks. The results in Table 1 are quite intriguing: the main key words identified by lexicographic analysis had related the cluster B8, the emergent one, with human health, not with bioenergy. The subsequent analysis of the content of the paper reveals the importance of papers directly related to bioenergy, which motivated the application of the steps of the methodology to the cluster 8.
The results allows to achieve a sub-cluster G4 clearly identified with the themes of bioenergy, an emergent cluster that is not strongly connected with the other 3 sub-clusters. Nanotechnology applied to sensors and engines is the main subject of the other three sub-clusters, indicating a future convergence of technologies, which reinforces the original hypothesis of the paper about the emergent nature of bioenergy.

References

BASTIAN M., HEYMANN S., JACOMY M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

BELL, M.; PAVITT, K. (1992) National capacities for tecnological accumulation: evidence and implications for developing countries, World Bank's Annual Conference on Development Economics, Washington, D.C., USA, 30 April & 1 May 1992.

BROWN,R C.. ;BROWN,T. (2012) Why are we producing Biofuels? Paperback, 341p.

CSARDI G, NEPUSZ T. The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. http://igraph.org

CLAUSET, Aaron; NEWMAN, Mark EJ; MOORE, Cristopher. Finding community structure in very large networks. Physical review E, v. 70, n. 6, p. 066111, 2004.

DAL POZ, M. E., SILVEIRA, J. M. J., MASAGO, F. K. (2012). Innovation Networks: Emerging Technological Trajectories on Ethanol Fermentation Processes, *International Consortium on Bioeconomy 2012.*

FRANK e HARRELL JR, with contributions sfrom Charles Dupont and many others. (2014). Hmisc: Harrell Miscellaneous. R package version 3.14-4. http://CRAN.R-project.org/package=Hmisc

FONTANA, R. NOVOLARI, A; VERSPAGEN, B. Mapping technological trajectories as patent citation networks. Economics of Innovation and New Technology Volume 18, Issue 4, 2009

FRANTZI, K.; ANANIADOU, S.; MIMA, H. Automatic recognition of multi-word terms:. the C-value/NC-value method. International Journal on Digital Libraries, v. 3, n. 2, p. 115–130, 2000. Acesso em: 27/1/2014.

GLÄNZEL, W.; SCHUBERT, A. (2004). Analyzing scientific networks through co-authorship. In: MOED, H. F.; GLÄNZEL, W.; SCHMOCH, U. Handbook of quantitative science and technology research. Netherlands: Kluwer Academic, 2004. p. 257-276.

GUIMERA, R.; AMARAL, L. A. N. Functional cartography of complex metabolic networks. Nature, v. 433, n. 7028, p. 895–900, 2005. Acesso em: 27/1/2014.

H. WICKHAM. ggplot2: elegant graphics for data analysis. Springer New York, 2009.

HADLEY WICKHAM (2011). The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software, 40(1), 1-29. URL http://www.jstatsoft.org/v40/i01/.

HLPE. (2013). Investing in smallholder agriculture for food and nutrition security. A report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security, Rome.

INGO FEINERER AND KURT HORNIK (2014). tm: Text Mining Package. R package version 0.6. http://CRAN.R-project.org/package=tm

JACKSON, M. (2008). Social and Economic Networks. 1a ed. Princeton University Press.420p.

KAMM, B., KAMM, M., SCHMIDT, M., HIRTH, T. and SCHULZE, M. (2008) Lignocellulose-based Chemical Products and Product Family Trees, in Biorefineries-Industrial Processes and Products: Status Quo and Future Directions (eds B. Kamm, P. R. Gruber and M. Kamm), Wiley-VCH Verlag GmbH, Weinheim, Germany. doi: 10.1002/9783527619849.ch20.

KRAFFT, J., QUATRARO, F , SAVIOTTI, P. (2010). The Evolution of Knowledge Base in Knowledge Intensive Sectors: Social Network Analysis of Biotechnology, *Bureau of Research in Innovation, Complexity and Knowledge*, *Collegio Carlo Alberto Working paper* 09/2009.

KOSTOFF, R.N., SCHALLER, R.R., 2001. Science and technology roadmaps. IEEE Transactions on Engineering Management 48, 132–143.

NEWMAN, M. E. Fast algorithm for detecting community structure in networks. Physical review E, v. 69, n. 6, p. 066133, 2004. Acesso em: 28/5/2013.

NOOY, W.;MRVAR, A.;BATAGELJ (2005). Exploratory Social Network Analysis with Pajek paperback, Cambridge University Press, 334p.

PEREIRA C.A.; BAZI, R.E.R.(2009) Flow and social relationships of knowledge in Scince, Technology and Innovation: a patento-metric study of Unicamp's technological production. Scientometrics, 81, 61-72.

PERSSON, O., R. DANELL, J. WIBORG SCHNEIDER. 2009. How to use Bibexcel for various types of bibliometric analysis. In Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday, ed. F. Åström, R. Danell, B. Larsen, J. Schneider, p 9–24. Leuven, Belgium: International Society for Scientometrics and Informetrics.

RIBEIRO, L. C.(2010) ; Ruiz, R. M. ; Bernardes, A.; Albuquerque, E. M. Matrices of science and technology interactions and patterns of structured growth: implications for development. Scientometrics, Vol. 83.

R CORE TEAM (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

SAVIOTTI, P. P. (2009). “Knowledge networks: Structure and dynamics”. In Innovation networks: Developing an integrated approach, Edited by: Pyka, A. and Scharnorst, A. 19–42. Heidelberg: Springer Verlag.

SHIBATA, N.; KAJIKAWA, Y.; TAKEDA, Y.; MATSUSHIMA, K. Detecting emerging research fronts based on topological measures in citation networks of scientific publications. Technovation, v. 28, n. 11, p. 758–775, 2008. Acesso em: 27/5/2013.

SHIBATA, N.; KAJIKAWA, Y.; TAKEDA, Y.; SAKATA, I.; MATSUSHIMA, K. Detecting emerging research fronts in regenerative medicine by the citation network analysis of scientific publications. Technological Forecasting and Social Change, v. 78, n. 2, p. 274–282, 2011. Acesso em: 10/5/2013.

SILVEIRA, J. M. F. J; DAL-POZ, M. E.; LUNA, I. H.; FERRARI, V. E; MASAGO, F. (2013) Análise de Redes de patentes em Biotecnologia Agrícola: apropriabilidade e destruição adaptativa. Anais do XLI Encontro Nacional de Economia, ANPEC