Information Visualization Analysis of Patent Filing and Non-Patent Literature Publication Trends in the Field of Bioenergy

Celeste Corrado (clc354@drexel.edu) and Geoffrey Valentine (gtv123@drexel.edu)

College of Information Science and Technology, Drexel University
Professor Chaomei Chen
February 17, 2013

Abstract - (1) This paper describes using visualization tools to examine two realms of related research, patent literature and non-patent research literature, in the field of bioenergy in order to discern patterns that indicate relationships between the two realms. (2) The tools CiteSpaceII and Themescape are employed to create interactive visualizations (3) of data pulled from Web of Science and Thomson Innovation. (4) The visualizations are explored and analyzed. CiteSpaceII co-citation networks are used to identify research fronts and Themescape visualizations are used to explore the patent data set for analogous groups of documents. (5) The mean citee year of clusters identified in CiteSpaceII are compared to the mean filing years of the correlated patents. Patterns and trends are discussed and a method for predicting when a research front may be reduced to practice is considered.

1. Introduction

The field of bioenergy is dynamic and has undergone a great deal of transformation recently due to a societal focus on alternative and renewable energy solutions. We chose this field because we expect that an analysis of patent filings and non-patent bibliographic research structures will reveal interesting patterns of research and technology over time. We believe that the present-day focus on the field along with external factors driving research in this area, such the increasing industry demand for a viable economic model for bioenergy and the environmental imperative to reduce greenhouse gas emissions, will have a marked impact on the field. We believe that the combined approach to patent and research publication analysis we are proposing may be useful in many disciplines—not only bioenergy—to reveal and validate useful trends and patterns.

Our hypothesis is that research fronts will first surface in the form of research articles, and as concepts develop, become formalized through testing, and are ultimately reduced to practice, we will find patent filings that are analogous to the research front appearing at predictable time intervals later.

Our investigations into the patent and non-patent literature revealed a yearly increase in both patent filings and non-patent publications from 2005 to 2010 (please see Fig. 1). Patent filings have steadied since 2008 while non-patent publications have continued to increase. After 2010, patent filings have naturally dropped significantly due to the inherent delay between filing and publication. Non-patent publications, however, have continued to increase year to year up to 2011.

Our expectation was that analysis would reveal similar trends and patterns in the realms of both non-patent publications citation and patent filing analysis. In addition, we anticipated seeing evidence of a time delay when comparing the two realms as basic research transitions into a reduction to practice, and is eventually utilized by industry or otherwise commercialized.

This paper is organized into six sections including this introduction. Section 2 covers the tools used, the data sources, and the methods employed for this project. In section 3, the overall results of the visualizations are presented as Figures with brief descriptions. In section 4, we will identify and discuss some interesting clusters of research in common between the two realms, and some not in common, and analyze the temporal patterns we uncover. In addition we will describe how we validated our results using other tools. In the discussion section, 5, we will review and consider the implications of our findings. We will use our results to speculate on future directions in the field of bioenergy. Finally, we will offer some concluding remarks in section 6 and reflect on the potential utility of our methods of analysis for exploring other fields where the realms of patent and non-patent research interact.

2. Data, Tools, and Methods

2.1 Data

The patent data for this paper were collected from the commercial patent databases Thomson Innovation and PatBase and public databases available from the United States Patent Office's (USPTO) website. Non-patent data were collected from Web of Science. Searches were conducted using the same keyword query across all databases: (BIOFUEL* OR BIO FUEL* OR BIOENERG* OR BIO ENERG*). In addition, we consistently used a timeframe from 2000-2010 for patents and 2005-2012 for articles. We focused our time slice searching on the years 2005-2010, since both patent analysis and the Web of Science (WoS) bibliographic citation analysis revealed that this period contained the most publication/citation activity and patent activity. The rationale for the timeframe used resulted from an analysis of both patent filings and non-patent publications where we observed that the timeframe for a marked increase in both data sets occurred during the time period 2005-2010.

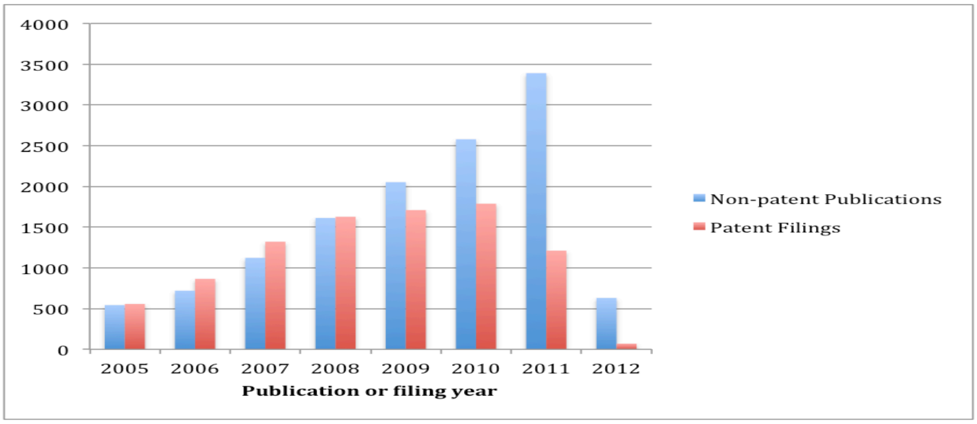

Figure 1 - Filing and publication years

2.2. Tools

The primary tools used to create the visualizations in this paper include CiteSpaceII [A] for exploring co-citation networks within the Web of Science [B] data set and Themescape [C] maps to render visualizations of the Innovation search results. For the validation steps we used VOSViewer [D] to render visualizations of non-patent data. We also employed Loet Leydesdorff's [6, E] patent citation network overlay program to generate network visualizations of USPTO granted patent data in VOSViewer. In addition, we used native visualization tools in the PatBase [F] system for further validation.

2.3 Methods

The query string noted above was used in all of our searches. Because we employed different sources for our data, some minor customization and/or variation was required for each source. Date ranges used in initial searches are different for patents to account for the fact that CiteSpaceII explores a co-citation network of articles that reaches back further than the date range of the initial Web of Science search. Our goal was to have a reasonably comparable timeframe within the different visualizations. We will briefly discuss below methods used for each of these three data sources and the visualizations we generated.

2.3.1 Non-patent Literature

Other than timeframe, topic, and prioritization by highest citation to lowest, there were no other filters applied to the WoS query. This bibliographic search resulted in 9,420 articles. The data sets were sorted by citation count, highest to lowest, and the first 1,000 results were extracted for analysis.

Two different tools were selected to visualize and map the bibliographic data set; CiteSpaceII [A] and VOSViewer [D]. Our rationale was to leverage at least two different information visualization solutions to compare and contrast our findings. Additionally, exploring both tools allowed us to evaluate the benefits and constraints of each when interpreting the results.

We analyzed the WoS data set in CiteSpaceII at n=top 50 for 1 year time slices. We used the Citation Burst as well as the Clustering and Labeling functions to achieve the visualizations [3, 4] in Figure 2. We chose to label clusters with title terms using LLR (log likelihood ratio test) algorithm because it tends to highlight the unique aspects of a cluster. This labeling algorithm resulted in terms that were more meaningful for comparison against the patent analysis than the other two methods, TF*IDF (term frequency–inverse document frequency) or MI (mutual information).

2.3.2 Patent Literature

For our primary patent data we searched Thomson Innovation [C], restricting the query to US patents or publications filed between 2000-2010. This search string pulled 7,801 patent documents. We used this data set to create an interactive Themescape map (Fig. 4) that we explored in order to compare and contrast the patent data with the clusters identified by CiteSpaceII.

Specifically, our first step was to employ time slicing, visualizing two-year time slices of patents filed in 2005-2006, 2007-2008, and 2009-2010 in an attempt to identify patterns over time. When this method proved to be less illuminating than we expected (Fig. 5), we compared CiteSpaceII clusters with the peaks identified on the Themescape map. For each CiteSpaceII cluster compared, we also deployed keyword visualizations on the Themescape map to help draw our eyes to the most relevant areas on the map. We then used the interactive tools in Themescape to analyze the relevant subsets of documents for the mean filing years, comparing this datum to the mean citee years reported by CiteSpaceII for analogous clusters.

2.3.3 Validation Methods

For our validation step, we used VOSViewer [D] to render a density map of the terms found in the Web of Science data set. In addition, we used PatBase visualization tools and Loet Leydesdorff's patent citation network overlay program [6, E] to generate alternative visualizations of the patent data. We compared the results to our previous findings.

3. Results

3.1 Non-patent CiteSpaceII Results- Overview

The CiteSpaceII cluster labeling function reveals the top 15 research clusters. The 15 clusters are organized primarily in three prominent groupings with at least two of these groups (Groupings A and C) linked with the third (Grouping B) appearing to have no linkages or relationships to the other key groupings.

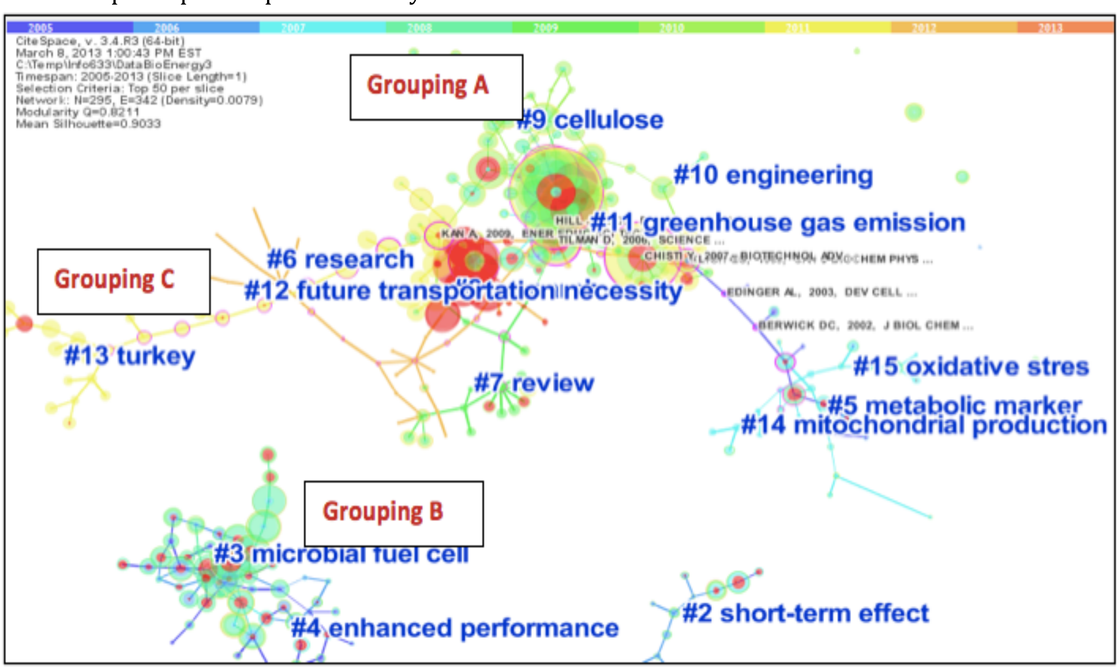

Figure 2 - Labeled clusters in CiteSpaceII

Tables 1 and 2 are excerpts from CiteSpaceII Automatically generated table of Major Clusters and Bursts.

Table 1: Summary of the largest 6 clusters [A]

ClusterID	Size	Silhouette	Label (TFIDF)	Label (LLR)	Lbel (MI)	mean(Citee Year)
11	47	0.847	(13.15) lipid	greenhouse gas emission (43.16, 1.0E-4)	advanced biofuel	2004
8	38	0.945	(8.88) prof	research (90.3, 1.0E-4)	incentive	2007
3	34	0.948	(17.08) electricity generation	microbial fuel cell (115.7, 1.0E-4)	computational model	2004
9	28	0.873	(12.05) cellulose	cellulose (65.79, 1.0E-4)	bronsted acidic ionic liquid	2004
4	21	0.958	(11.62) enhanced performance	enhanced performance (47.41, 1.0E-4)	electrode surface	2000
7	20	0.895	(9.84) currency	review (56.42, 1.0E-4)	thermo-chemical conversion	2008

Table 2: Summary of top 6 ranked bursts and their related clusters [A]

bursts	references	cluster #
10.53	Demirbas B, 2010, ENER EDUC SCI TECH-A, V26, P37	8
9.91	Searchinger T, 2008, SCIENCE, V319, P1238	11
9.83	Demirbas AH, 2010, ENER EDUC SCI TECH-A, V26, P13	8
9.76	Balat H, 2010, ENER EDUC SCI TECH-A, V24, P85	8
9.57	Demirbas A, 2010, ENER EDUC SCI TECH-B, V2, P75	8
8.87	Demirbas MF, 2010, ENER EDUC SCI TECH-A, V25, P31	8

The timeline feature in CiteSpaceII depicts the top 15 clusters and the top highly cited articles. Again we note the aforementioned groupings. What is noteworthy is that the timeline view clearly reveals two "classic" research articles; Warburg (1956) and Bligh, E.G. (1959). Although each of these "classics" are linked to one of the two prominent groupings of clusters, Group A and Group B, again we do not observe any linkages between these two articles and their respective clusters.

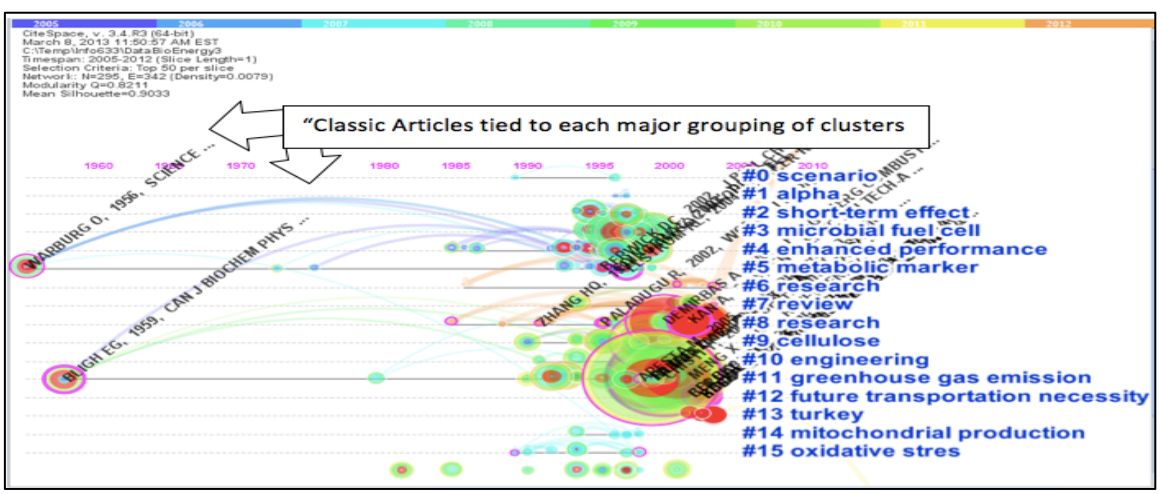

Figure 3 - Timeline view in CiteSpaceII

3.2 Patent Literature Themescape Results-Overview

Themescape's (proprietary) algorithms analyze patent text to create a topographical map that shows 'themes' within a document set. The themes are labeled using terms from the patent titles and abstracts. Figure 4 is a visualization of US patents and published applications filed from 2000-2010 matching the query anywhere in the full text of the patent. Contour lines appear around sets of documents that are related by theme. Color represents density - blue 'ocean' fewer documents; white 'snowcaps' most documents.

Figure 4 - Themescape map with Labels

Each document on the Themescape map may be represented as a dot and may be highlighted in colors to help guide the analysis. Below in Figure 5 is the time slice map, highlighting three time slices: patents filed in 2005 and 2006 highlighted in yellow; 2007 and 2008 highlighted in green; and 2009 and 2010 highlighted in red.

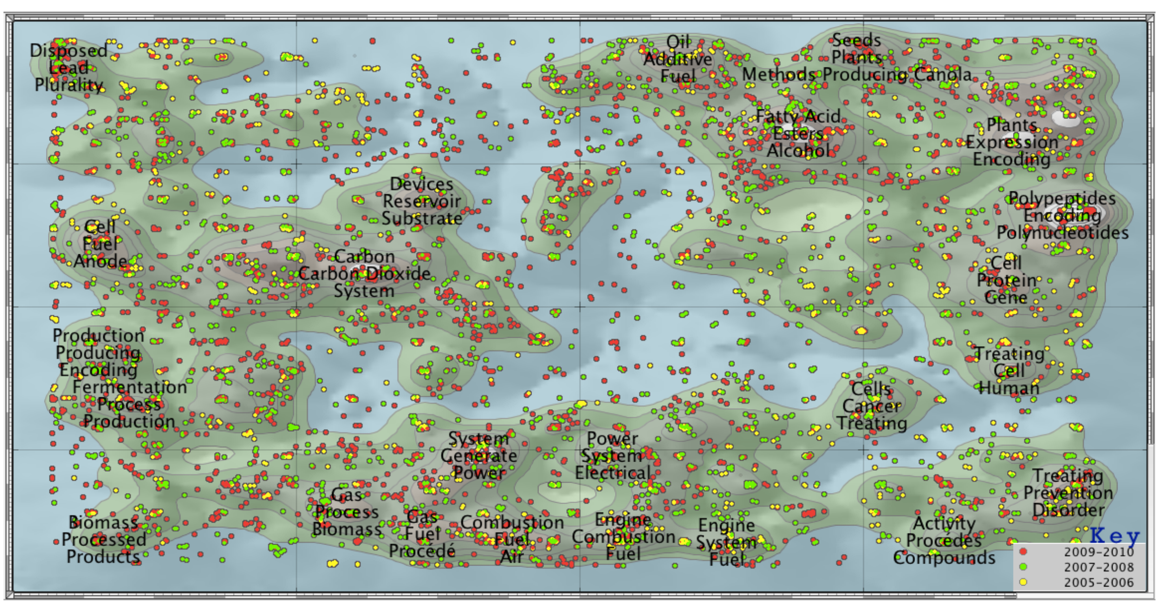

Figure 5 - Themescape map with time slices visualized

4. Analysis

4.1 Non-Patent Literature Analysis

Using the link walkthrough function in CiteSpaceII, we can learn more about the evolution of bioenergy research and its intellectual structure [3, 4]. From the 2005 link walkthrough (Fig. 6), we observe the base of the field's foundational knowledge beginning to coalesce with the two key foundational groupings, grouping A and B.

There are smaller groupings that are not as prominent

- Group A= containing primary clusters#9 (cellulose) and #11 (greenhouse gas emission)
- Group B= containing primary cluster #3 (microbial fuel cell for efficient electricity generation)

Note: these groups appear equally prominent but have no visible linkages. In the previously discussed timeline view (Figure 3), we also uncovered the "classic" research articles that each of these groups are linked to.

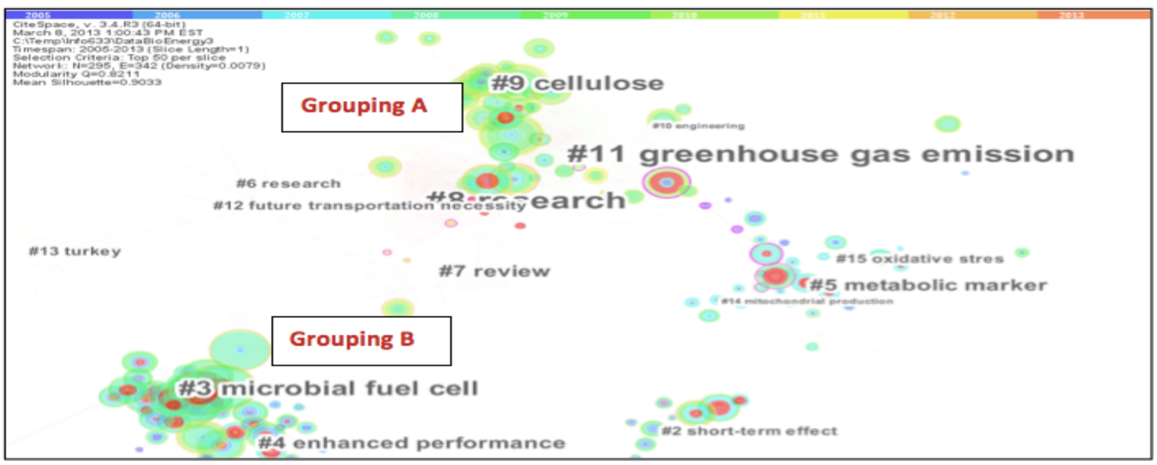

Figure 6 - Link Walkthrough in CiteSpaceII, 2005

In 2006 we see significant change within and around the Group A cluster with a burst of activity around the 'Farrell, AE (2006)' article at the center of cluster #11 (greenhouse gas emissions and advanced biofuel).

In 2008 (see Fig. 7) there is an acceleration in research linked to cluster #11 (greenhouse gas emission) with a burst of citation activity around the 'Chisti,Y (2007)' article in cluster #11. Group B's cluster #3 remains unchanged. We observe the emergence of a new citation burst and cluster #8 (research) adjacent to cluster #11 (greenhouse gas emissions and adv biofuels).

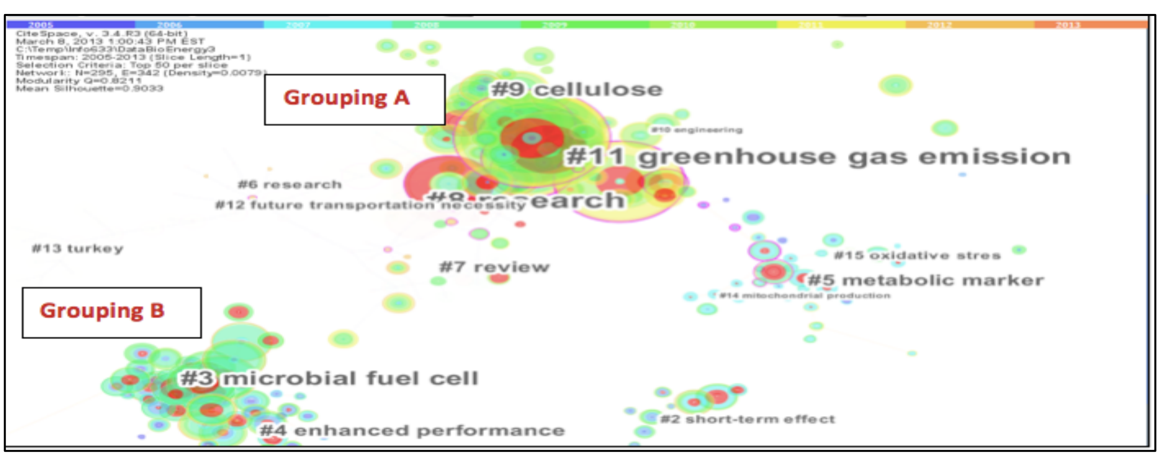

Figure 7 - Link walkthrough in CiteSPaceII, 2008

By 2010 we see several offshoots of research activity within Group A. We see two clusters labeled *research* developing. Cluster #8 is the more notable, showing an increase in citations and considerable "burst" activity. New smaller research fronts are emerging in the #13 cluster (turkey) (Figure 2). If we dig into burst article details we are likely to see that around this period there is a great deal of research activity being conducted by a Turkish researcher, 'Demirbas, A. (2010)' and cited by others. Lastly, we observe that clusters #6, #12, and #13 are linked to the foundational grouping A, which includes clusters #8, #11 and #6 (research) and is continuing to experience a burst of citation activity.

In 2012-13, we observe new linkages forming between cluster #7 (review) and #12 (future transportation necessity) and #6 (research). We observe a burst of citation activity in cluster #8 (research) that could indicate a future direction of research in this field.

If we zoom into the CiteSpaceII cluster labeled *visualization* (Figure 8) to the key prominent clusters, we can see both the highest ranked clusters and their respective burst activity. We see the relative timeframes in which the bursts occur. Cluster #11 (greenhouse gas emission) is the more foundational cluster of the grouping and #8 (research) has the most recent burst activity. In the discussion section, we will elaborate on the significance of cluster #8 (research) and the topic of *microalgae*, which emerges from the burst articles (Table 2) as a consistent topic in the cluster. This observation has potential implications for the future direction of bioenergy.

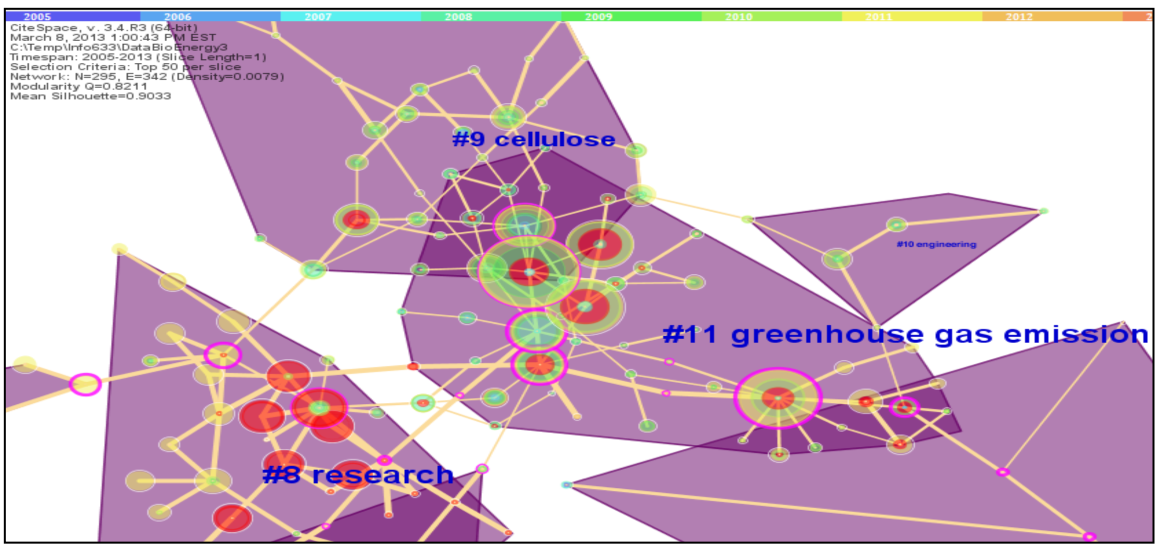

Figure 8 - Zoomed in view of cluster #11

4.2 Comparison of Patent Literature and Non-Patent Literature Visualizations

A comparison of the patent data as visualized by Themescape to the top six clusters identified in CiteSpaceII reveals no precise alignments. However, we will attempt to draw some comparisons between five of these clusters and the peaks revealed by Themescape.

The largest cluster identified by CiteSpaceII is #11, labeled *greenhouse gas emission* by the LLR algorithm. This appears analogous to the peak labeled *carbon dioxide system* on the Themescape map (see Fig. 9). This set of 79 patent documents is largely comprised of methods for mitigating CO2 emissions, a topic similar to that in the *greenhouse gas emission* cluster of non-patent documents. We can reasonably interpret the mean 'citee year' of 2004 reported by CiteSpaceII (see table 1) for cluster #11 as the mean year of the research base. This analogous grouping of patent filings has a mean patent filing date of 2008. This may be evidence of a natural delay between the former research front of 2004 (now, in 2013, interpreted as the research base) and the reduction to practice that a patent filing can represent. In other words, research from around 2004 begins to be translated to patent filings in 2008.

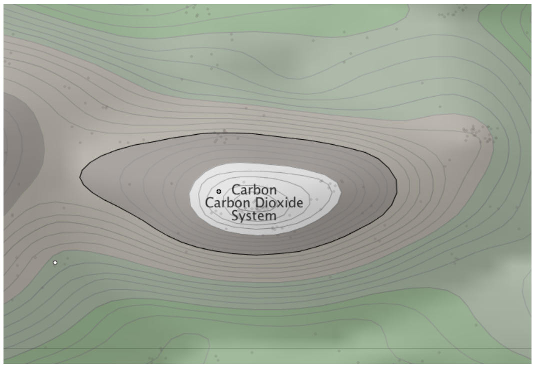

Figure 9 - Themescape 'CO2 system' peak

CiteSpaceII's second largest cluster, #8, *research*, is not a good one for comparison simply because the label terms are too broad to bring anything useful up in the patent search.

The next largest cluster, # 3, labeled *microbial fuel cell* by the LLR algorithm, appears to be analogous to the Themescape map peak labeled *cell fuel anode*. Fig. 10 is a zoomed view of the peak. In this image the red dots highlight patent documents that contain any of the terms 'microbial' or 'fuel' or 'cell'. The average filing year for this cluster of patents is 2007. The average citee year for the analogous CiteSpaceII cluster is 2004. We note a similar disparity here between the mean citee and mean filing years.

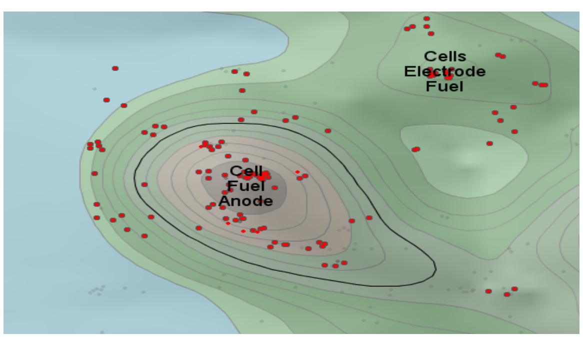

Figure 10 - Themescape 'cell fuel anode' peak

The fourth largest cluster in CiteSpaceII is #9, labeled *cellulose* by the LLR algorithm, appears to correspond to a broader area on the Themescape map, not one specific peak (see Fig. 11). In this image, the red dots identify patent documents using the keyword 'cellulose', which is a fairly broad term in the field. There are 298 patent documents in the group with an average filing year of 2008. The corresponding CiteSpaceII cluster has a mean citee year of 2004. Note a similar delay between the mean publication and filing years.

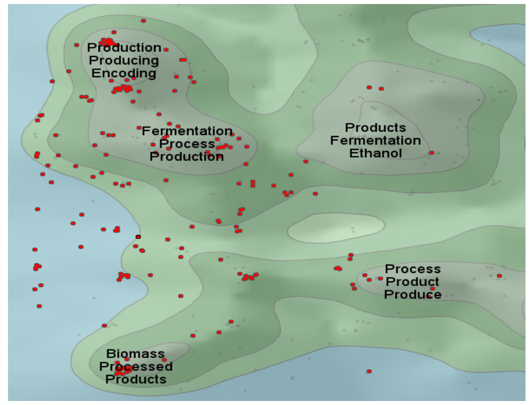

Figure 11 - 'cellulose' peaks

For the fifth largest cluster, #4, we will use the label generated by the MI algorithm in CiteSpaceII, *electrode surface,* because the LLR and TFI-DF algorithms are not specific enough to be helpful for comparison with our patent data. This is another label that appears to correspond to a larger area of the Themescape map, appearing in the peaks labeled *cells electrode fuel*, *cell fuel ano*de, *fluid disposed fuel*, and others. In Fig. 12, the red dots highlight patent documents having both terms 'electrode' and 'surface' in the text. The average filing date for these patent documents is 2006. The mean citee year for the analogous non-patent literature per CiteSpaceII is 2000. Here again, we see a difference between the mean publication and filing years in the two realms.

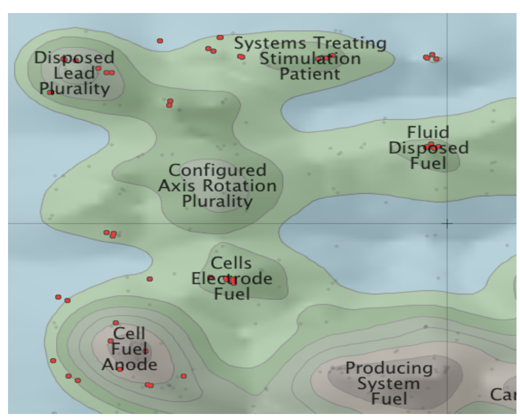

Figure 12 - highlighting 'electrode surface' patents

The sixth largest cluster, #7, labeled as *thermo-chemical conversion* using the MI labeling algorithm in CiteSpaceII, does not have any analogous peaks in the Themescape map. There are just ten patent documents that correspond to the keywords 'thermo-chemical' or 'themochemical' with a mean filing date of 2009; seven out of ten were filed in 2010. This is notable because the mean citee year for this cluster in CiteSpaceII is 2008. Here we have what appears to be a newer research from in the non-patent literature. We may be able to predict a peak of patent filings in 2012-2013 once patents filed in those years begin publishing in earnest.

There are some outlier peaks in the Themescape map. Notably, there are several peaks that pertain to cancer treatments. A deeper look at these indicates that the term 'bioenergetic' is a term of art in cancer research. Lastly, additional false peaks are generated from citations in medical related patents to the articles from the Journal of Bioenergetics and Biomembranes. Our broad keyword search picked these up. Further investigations into this field should take possible false hits like these into account.

4.3 Validation

In order to validate our findings, we generated visualizations using other tools. For example, Figure 13 is a heatmap visualization generated by VOSViewer of terms found in the Web of Science data set. This map helps to illustrate the connection between the patent data set and the research literature data set. The image reveals keyword clusters in common with those we found when exploring the Themescape map (Fig. 4), including, for example, 'anode', 'electrode', 'fuel', 'cell', and 'oil'. Also of interest, this visualization shows the 'bioenergetic' outlier that we discovered in the patent search results.

The VOSViewer visualization also reinforces the research clusters identified by CiteSpaceII, for instance, 'emission' , 'cellulose', and 'microbial fuel cell' are comparable to the clusters labeled #11, #9, and #3 respectively as identified by CiteSpace (see Fig. 2).

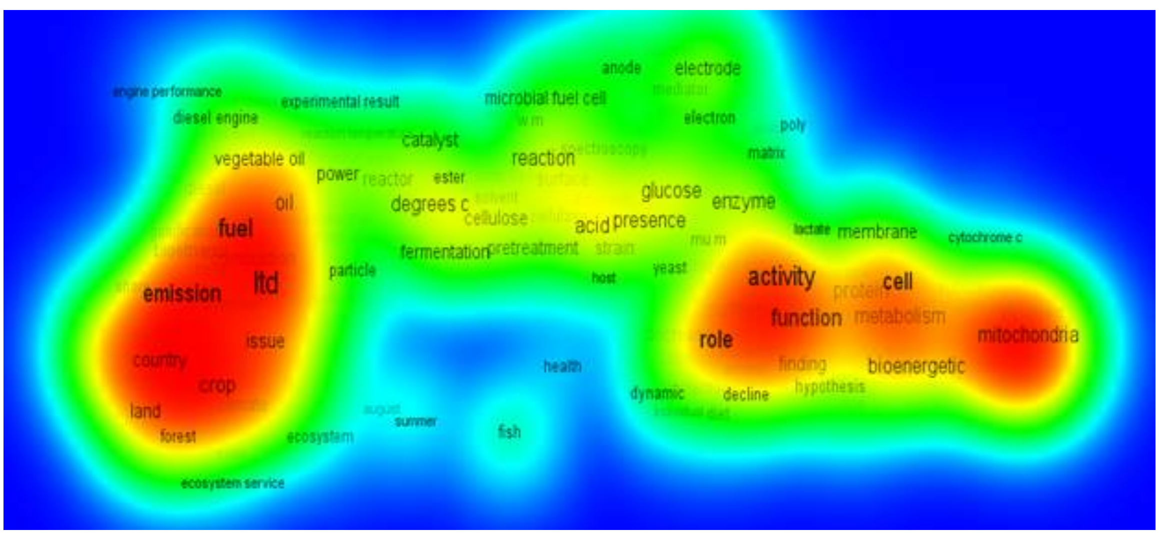

Figure 13 - Text Corpus desity map of citation results VOSViewer

We also used Loet Leydesdorff's patent citation network overlay program to generate a similar heat map of US granted patents filed in 2007 and 2008 (Fig. 14). The labeling terms are generated by international patent classification (IPC) code descriptions. We can draw comparisons between the dense cluster 'fuel not otherwise provided' and the peaks labeled with the term 'fuel' in the Themescape map. Also, the cluster 'internal-combustion piston engine' may be correlated to the peak named 'combustion fuel air'.

We may draw a parallel between the cluster 'micro-organisms or enzymes' and the CiteSpaceII cluster #3, 'microbial fuel cell'. Additionally, CiteSpaceII cluster #9, 'cellulose' may be comparable to the label 'production of cellulose by rem' on the VOSViewer visualization.

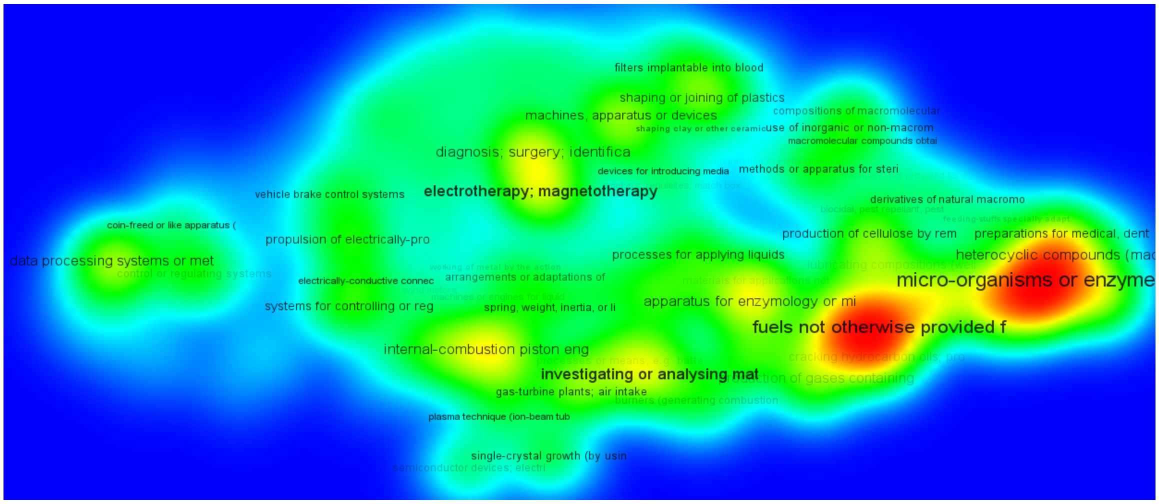

Figure 14 - Map of IPC codes, VOSViewer

We also generated visualizations using PatBase that illustrated similar correlations between the two data sets. PatBase uses a combination of keywords from patent titles and abstracts and IPC code descriptions to generate concentric ring view of patent data that illustrates term frequency and proximity. For example, terms such as 'bio fuel', 'combustion engines' and 'micro-organisms' are highlighted, which are also comparable to our findings in CiteSpaceII and Themescape visualizations.

5. Discussion

Our primary research objective was to explore and identify characteristics of patent filing and research publication trends in the field of bioenergy in order to discern similarities and differences, and to identify interesting relationships or links between our patent and research analysis.

Defining a correlation between research and patent fronts is challenging given the fact that the process of filing patents is very different from the process of submitting scientific papers. Nonetheless, both of these processes are, at some point, derived from some form of research and development activity with novelty being one of the criteria for acceptance into their respective structures [7]. In addition, both of these intellectual structures are subject to rigorous review and each represent a base of knowledge that are likely to have themes in common that can be extrapolated and used in complementary analysis. One of the key questions we have attempted to answer in this paper is whether there is a timing difference between these two structures; specifically, is there a consistent temporal delay between the publication of research articles and the filing of analogous patents? The challenge is the fact that scientific research and commercial technology are often intertwined and not easy to decouple [7].

5.1 Fundamental similarities and differences between patent and non-patent publication analysis

Scientific research is collaborative by nature, typically involving multiple authors. It exists within a culture of ethics whereby the related work of others is voluntarily cited and research papers are subject to peer review and scrutiny that reinforces the validity of the content [7]. Patents are intellectual property, and, as such, are less openly collaborative. Inventors are often backed or funded by commercial interests. Patent applications are reviewed and challenged by a patent examiner and both the inventor and the patent examiner provide citations to related patent and non-patent prior art.

Another interesting difference is that the scientific research community has the freedom to write on topics, ideas, methodologies, policies, etc. that may not translate into a reduction to practice through the patent system. Non-patent bibliographic structures and insights derived from those structures have been studied quite a bit over the last two decades. However, co-inventive and patent activity is not as well studied. [6] Relationships between these two structures have been studied in the past but remain less well understood.

5.1.1 Patent Filing Year vs. Citee Year as one method of comparison

We compared the mean patent filing year against the mean citee year for the analogous CiteSpaceII clusters and patent peaks and noticed a 4 to 6 year delay between the research fronts and patent filings (see Table 3). Research fronts were noted in CiteSpaceII, on average, 4 to 6 years earlier then when they surfaced in the patent filing database. This correlation could form the basis of a combined patent/research analysis methodology that is repeatable.

Table 3. Comparison of mean cite year and patent filing year

	Mean Citee Year	Mean Filing Year
Greenhouse Gas	2004	2008
Microbial Fuel Cell	2004	2007
Cellulose	2000	2008
Electrode	2000	2006
Thermo-chemical	2008	2009

5.2 Bioenergy Patterns & Trends

Using the results of our bioenergy patent and non-patent publication analysis, we uncovered both similarities and differences when associating clusters of research with clusters of patent activities. We observed research patterns in the intellectual structure of the vast field of bioenergy using CiteSpaceII's link walkthrough feature and Themescape's time slicing and highlighting features. This effort yielded a rational basis for the research trails evolving over the past 8 years while providing indicators of current research fronts that are likely to carry forward into the future.

5.3 Information Visualization Software Observations

We explored several information visualization software solutions for both the patent and non-patent analysis. We found that no single visualization software provided the functionality to do a complete and consistent analysis for either the patent or bibliographic analysis. Through experimentation and iteration, we selected the best of each software solution for our purposes and pieced together the correlation between patents and non patent publications. When compared to VOSViewer, CiteSpaceII offered far more robust functionality . VOSViewer would have been more useful if it combined both citation and text corpus views and there was a way to perform temporal analysis. For patent analysis, the Themescape software allowed for temporal analysis and associations of patent groupings. Although PatBase had time slicing capability and we found the tool to be visually interesting, it did not add sufficient content value to be included our analysis.

5.4 Indications & Future Directions

Uncovering correlations between these two intellectual structures may provide a method of predicting which research fronts will yield significant clusters of patents and when that might occur. We believe we achieved the research objectives we set out to accomplish.

The results from this analysis could be beneficial to bioenergy stakeholders interested in developing a comprehensive understanding of the recent evolution of the field of bioenergy as well as anyone interested in predicting which research front concepts may be later reduced to practice, e.g., investors or companies exploring new technologies to exploit.

Using our results to look ahead, we predict that from the evolving eight year research trail and the recent burst activity observed in CiteSpaceII [1, 2], research in microalgae as a viable feedstock solution will likely continue well into the future. We observed a pattern of progressive research leading up to the current research front in microalgae. We observe that the intellectual base has evolved from unsustainable bioenergy feedstocks such as ethanol and first generation combustion processing solutions to an awareness of how these solutions contribue to greenhouse gas emissions and have negatively impacted our ecosystem including food supply, water, land use, etc. These associations are best observed in the text corpus view of VOSViewer (fig. 13). Around the 2008 timeframe, we see a burst of research activity focused on greenhouse gas emissions (cluster #11) and their impact on the ecosystem. There is further support of this observation in CiteSpaceII's Burst Report [Table 2] indicating 'Searchinger, T. (2008)' as the second ranked item (9.91) by bursts and whose article content directly relates to this topic.

More recent research bursts are focused on micro-algae as a potential new feedstock, which is considered both environmentally sustainable and has fewer negative effects on human health and well being. We see detailed evidence of the micro-algae focus through CiteSpaceII's automatically generated burst report where 'Demirbas, B. (2010)' is the top ranked item by bursts (10.53) and the topic of his paper as 'microalgae as a feedstock for biodiesel'. When we look at the topics that are contained within each of CiteSpaceII's top burst and centrality reports, which are indicators of continuing growth/activity, we see that micro-algae tops both lists.

Considering our information visualization findings and observations and the aforementioned temporal correlation, we might expect to see micro-algae solutions reduced to practice in the form of patents around the 2016-18 timeframe, or 4-6 years after the research front.

In addition, we see a recent burst of citation activity around bioenergy policy and education. This topic is likely a manifestation of extrinsic drivers; it an additional important contribution to the 8-year evolution of the growing bioenergy knowledge base. Although this type of research activity may not be reduced into practice, it is very helpful in uncovering concepts, ideas, and philosophies that are important contributions to the bioenergy body of knowledge. Often these philosophical or policy papers may be the catalyst or missing link for enabling the reduction of bioenergy solutions to practice. This is precisely the reason that non- patent publication analysis is both beneficial and complementary to patent analysis.

6. Conclusion

We set out in this paper to find a relationship between the research front in the field of bioenergy as revealed by visualizations of co-citation networks among non-patent research articles and the reduction to practice as revealed by visualzations of related patent concepts. When we examined analogous clusters of research publications and patents we observed a four to six year time delay between the average citee year of the research publications and the average filing year of the patents. We believe that a logical explanation for this is that research article clusters may define the temporal boundaries of research fronts while analogous patent clusters may define the temporal boundaries of the related concepts being reduced to practice.

Because our investigation was limited to a single field, bioenergy, and a relatively small timeframe, further investigations into other fields and of larger timeframes are needed to confirm whether the time delay we observed is unique to our data set, or if this method of analysis in other fields will reveal comparable relationships. Similar analysis of other knowledge domains may also show differences in timing from field to field. Additionally, some fields, such as mechanical engineering, have little to no reseach literature to analyze.

If our findings are confirmed on a broader scale, this technique may be useful for making predictions about future patent trends in intellectual domains where research literature forms the basis for future reductions to practice.

References

1. Chen, C., Zhigang, H., Shengho, L., Hung, T. (2012) Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace. *Informa Healthcare,* 1-14.

2. Chen, C., (2012) Predictive effects of structural variation on citation counts. *Journal of the American Society for information sciences and technology, 63(3),* 431-449.

3. Chen, C. (2006) CiteSpaceII II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, **57**(3), 359-377.

4. Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010) The structure and dynamics of co-citation clusters: A multiple-perspective co-citation analysis. Journal of the American Society for Information Science and Technology, 61(7), 1386-1409. 10.1002/asi.21309

5. Cobo, M.J., Lopez-Herrera, A.G., Herrera-Viedma, E., Herrera, and F. (2011) The science mapping software tools: review, analysis, and cooperative study among tools, *JASIST,* 62(7), 1382-1402.

6. Leydesdorff, L. (2012), Interactive overlay maps for US patent (USPTO) data based on International Patent Classifications (IPC). *Scientometrics (in press).* Retrieved from http://arxiv.org/abs/1210.6456v2.

7. Myer, M., Bhattacharya, S., (2004) Commonalities and differences between scholarly and technical collaboration. *Scientometrics. 61(5), 443-456.*

Software

A. CiteSpaceII: http://cluster.cis.drexel.edu/~cchen/citespace/
B. The Citation Map function in the ISI Web of Science: http://wokinfo.com/
C. Thomson Innovation 'Analyst' with Themescape: http://info.thomsoninnovation.com/
D. VOSViewer: http://www.vosviewer.com/
E. Loet Leydesdorff Interactive overlay maps for US patent (USPTO) data
   - http://www.leydesdorff.net/overlaytoolkit/
   - http://www.uspto.gov/patents/process/search/

F. PatBase: http://www.patbase.com/