

A DELINEATING PROCEDURE TO RETRIEVE RELEVANT PUBLICATION DATA IN RESEARCH AREAS: THE CASE OF NANOCELLULOSE¹

Douglas Henrique Milanez^{a,c,*}

Ed Noyons^{b,*}

Leandro Innocentini Lopes de Faria^{a,c}

^aFederal University of Sao Carlos, Materials Engineering Dept, Centre for Information Technology in Materials (NIT/Materiais), Washington Luis Highway, km 235, São Carlos – SP (Brazil)

^bLeiden University, Centre for Science and Technology Studies (CWTS), PO Box 905, 2300 AX Leiden (The Netherlands)

^cFederal University of Sao Carlos, Science Information Dept, Labmetrics, Washington Luis Highway, km 235, São Carlos – SP (Brazil)

*Corresponding authors:

Telephone: +55 16 3351 8551

Email address: douglas@nit.ufscar.br

Telephone: +31 71 527 3928

Email address: noyons@cwts.leidenuniv.nl

Abstract

Advances concerning publication-level classification system have been demonstrated striking results by dealing properly with emergent, complex and interdisciplinary research areas, such as nanotechnology and nanocellulose. However, less attention has been paid to propose a delineating method to retrieve relevant research areas on specific subjects. This study aims at proposing a procedure to delineate research areas addressed in case nanocellulose. We investigate how a bibliometric analysis could provide interesting insights into research about this sustainable nanomaterial. The research topics clustered by a Publication-level Classification System were used. The procedure involves an iterative process, which includes developing and cleaning a set of core publication regarding the subject and an analysis of clusters they are associated with. Nanocellulose was selected as the subject of study, but the methodology may be applied to any other research area or topic. A discussion about each step of the procedure is provided. The proposed delineation procedure enables us to retrieve relevant publications from research areas involving nanocellulose. Seventeen research topics were mapped and associated with current research challenges on nanocellulose.

Keywords: bibliometrics; research topics; science classification; nanotechnology.

MSC Classification 68P20

JEL Classification D83

Introduction

Bibliometrics has been used to monitor and quantitatively assess scientific fields within the context of science policy and research management (Moed, Glänzel, & Schmoch, 2004; Okubo, 1997; Raan, 2014; Zitt, 2015). Partly, it is a consequence of the increased use of Internet since the early 1990s and the development of information technologies. Together, they made a huge volume of scientific databases available. Meanwhile, scientific studies have become more complex and interdisciplinary, involving the exchange of knowledge between scientists from different disciplines in specific subject of research. Nanotechnology-focused research is a good example. Bibliometric indicators and tools are useful instruments to study and gain insight in science and, in particular, complex fields or research areas, c.f., van Raan (2004). Therefore, many studies on nanotechnology relied on bibliometric approaches (Hullmann & Meyer, 2003; Igami, 2008; Kostoff, Koytcheff, & Lau, 2009; Milanez, Faria, Amaral, Leiva, & Gregolin, 2014; Mogoutov & Kahane, 2007; Wang, Notten, & Surpatean, 2012). Nonetheless, the problem often is: how to delineate a field or research area; how to retrieve the relevant data; which publications to include and which not; what insights can be obtained from the set of publication retrieved.

¹ Earlier version presented at the 15th International Society of Scientometrics and Informetrics Conference in Istanbul (Turkey), Jun 29th – July 4th, 2015.

In this sense, classification systems are an indispensable tool to study the structure and dynamics of scientific fields (Boyack, Klavans, & Börner, 2005; Glanzel & Schubert, 2003; Leydesdorff, Carley, & Rafols, 2013; Leydesdorff, de Moya-Anegón, & Guerrero-Bote, 2014; Ruiz-Castillo & Waltman, 2015; Y. Zhu & Yan, 2015; Waltman & van Eck, 2012; Zitt, 2015). They can simplify literature search and information retrieving procedures (Glanzel & Schubert, 2003; Ruiz-Castillo & Waltman, 2015; Waltman & van Eck, 2012). Zitt (2015) provides an interesting overview of current field delineation procedures and challenges of information retrieval. According to Glanzel and Schubert (2003), classification of science into a disciplinary structure can be as old as science and Ruiz-Castillo and Waltman (2015) highlighted that the assignment of individual publications or journals to research areas still remains an open question in Scientometrics. Currently, most of classification systems are based on journal assignment, such as the Web of Science and Scopus systems. The drawback of these journal-based classification systems is the fact they do not deal properly with multidisciplinary journals or interdisciplinary research (Neuhaus & Daniel, 2009; Ruiz-Castillo & Waltman, 2015; Waltman, van Eck, & Noyons, 2010; Waltman & van Eck, 2012). Some studies question the appropriateness of the database journal-based classification system, such as for the purpose of normalizing citation impact indicators (Ruiz-Castillo & Waltman, 2015), assignment of individual publications (Neuhaus & Daniel, 2009) or evaluating complex and emerging domains (Zitt, 2015).

The development of publication-level classification systems is currently a subject of research. Boyack et al. (2011) clustered a corpus of 2.15 million biomedical publications from Medline database (2004-2008) which generated coherent and concentrated cluster solution of text-based similarity approaches based on keywords extracted from titles and abstracts. They found their approach more precise than the Medical Subject Headings. Waltman and van Eck (2012) proposed a methodology to cluster a large-scale set of scientific publication indexed on Thomson Reuters' Web of Science database. Each publication was assigned to a single research area, which was organized in a three-level hierarchical structure. Their methodology took into account direct citation to cluster the publication and they labelled each research area with discriminative keywords extracted from titles and abstracts. Such publication-level classification systems will be used to gain insights on research areas involved in specific subjects.

In the present study, we intended to map research areas associated with nanocelluloses. These cellulose-based nanomaterials are sustainable and has a great potential for innovation (Isogai, 2013; Mariano, Kissi, & Dufresne, 2014; Milanez, Amaral, Faria, & Gregolin, 2013; Moon, Martini, Nairn, Simonsen, & Youngblood, 2011). Nanocelluloses have been a research area for many countries, including the major producers of cellulose worldwide, such as the USA, Canada, Finland, Sweden and Brazil (Milanez et al., 2013). Different disciplines are involved with nanocellulose research since its properties and behaviour have allowed applications as reinforcement agent in composite materials, packing material, optically transparent paper for electronic devices, texturizing agent in cosmetics and food, bio-artificial implants and bandages (Isogai, 2013; Klemm et al., 2011; Mariano et al., 2014; Moon et al., 2011; Siqueira, Bras, & Dufresne, 2010).

Nanocelluloses is a generic term referring to cellulose nanofibrils on the one hand and cellulose nanocrystals on the other (Dufresne, 2013; Klemm et al., 2011; Moon et al., 2011; Siqueira et al., 2010; TAPPI, 2011). The main difference between these two types of cellulose in nanoscale dimensions relies on the degree of crystallinity. Cellulose nanocrystals are basically shorter and rod-like crystalline cellulose, whereas cellulose nanofibrils are long chains of alternate amorphous and crystalline cellulose. Consequently, they differ on their mechanical and functional properties (Eichhorn et al., 2010; Mariano et al., 2014; Moon et al., 2011). Both types of nanocellulose can be obtained from renewable sources, including natural fibres, plants, pulp and forest and agricultural residues. In this case, mechanical process and chemical/enzymatic approaches are used to obtain cellulose nanofibrils and cellulose nanocrystals, respectively. Moreover, cellulose nanocrystals can be biosynthesized by bacteria, resulting in the also called bacterial cellulose (Klemm et al., 2011; Milanez et al., 2013; Moon et al., 2011).

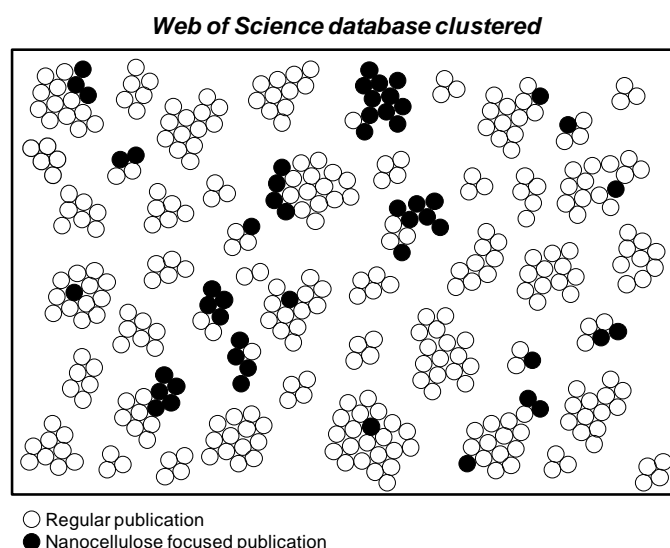
Checking the research topics associated with nanocelluloses can provide insights into current technical challenges concerning this nanomaterial, such as increasing the scale of production minimizing costs, characterization of sources and mechanical properties. Surface modifications to reduce moisture adsorption and improve the adhesion between the nanomaterial and the polymeric matrix, thermal degradation, and biocompatibility with living tissues has also been target of research (Gardner, Opo, Oporto, Mills, & Samir, 2008; Isogai, 2013; Klemm et al., 2011; Mariano et al., 2014; Milanez et al., 2013; Moon et al., 2011; Siqueira et al., 2010).

This study aims at proposing a delineation procedure to retrieve relevant research areas addressed to a specific topic. Nanocellulose was selected as a case, but it may be used for other subjects, of course. The approach involves research areas identified in the CWTS Web of Science Publication-level Classification System, a 2014 update of the version introduced by Waltman & van Eck (2012). This paper is structured as follows. In the next section, we describe the overall delineating procedure and its general issues. Next, we discuss details concerning specific parts and tasks. We present and discuss results in Section 3 and finally in Section 4 we draw our conclusions.

Methodology

Overall delineation procedure

To delineate the field, i.e., to collect a relevant set of publications to represent nanocellulose, clusters (research areas) were selected from the CWTS publication level classification system. By this method, papers that were not easily picked up by keyword or journal based search strategies could be identified. Figure 1 presents a schematic representation of the distribution of the clustered Web of Science publications according to CWTS Publication-level classification system (Waltman & van Eck, 2012). Predefined nanocellulose publications are indicated as black circles and the first step is retrieving all research area that contains at least one of them.



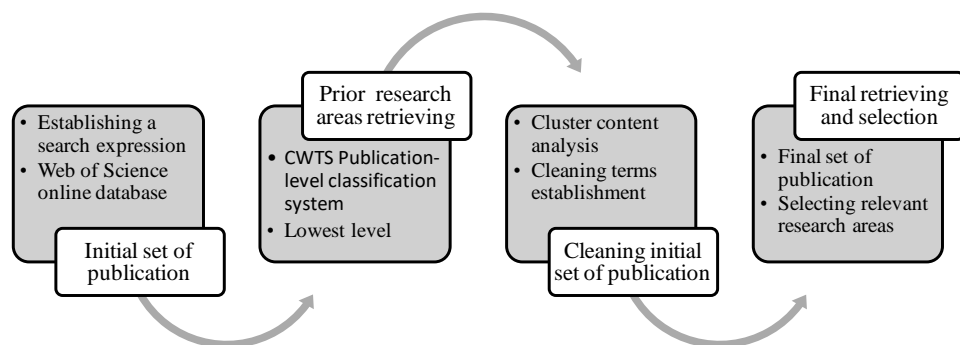
Source: authors.

Fig. 1 Schematic representation of Web of Science publications clustered according to the CWTS Publication-level Classification System. The black nodes represent the publications focused on nanocellulose.

Figure 2 depicts the proposed procedure as an iterative process which can be described in four main steps:

1. *Determine an initial set of publication concerning the theme of interest.* In this first step, a set of publication which well represents the theme of interest (nanocellulose) is retrieved via the online Web of Science database, using a straightforward search strategy. This set of publication is a starting set and will be refined as well as expanded through the next steps;
2. *Prior retrieval of nanocellulose research areas.* The second step involves locating the research areas (publication clusters) with at least one publication from the initial set of nanocellulose. The bottom level of the classification scheme was used in this study (Waltman & van Eck, 2012);
3. *Analysis of retrieved research area and cleaning of the initial set.* The content of each research area was analyzed pragmatically. A cleaning task was developed by selecting terms to eliminate part of the initial set of nanocellulose publication. This step provided a final set of nanocellulose publication clusters and enhanced the precision of research area assigned to nanocellulose;
4. *Final retrieval and selection of relevant nanocellulose research areas.* After cleaning the initial set of nanocellulose publication, the research areas (publication clusters) were retrieved again. Finally, as the number of topics retrieved was high, a selection that relies on the 80/20 rule was conducted reaching the final research areas associated with nanocellulose.

After performing the delineation, an independency test was conducted to verify how the steps from process affected the number of publication from the main authors in nanocellulose. We also had a look on the arrangement of the selected research topics on a map and discussed how relevant they are to support nanocellulose developments.



Source: authors.

Fig. 2 Iterative process of the overall procedure proposed.

Determine an initial set of publication on nanocellulose

A search expression was developed considering several terms and synonyms found in nanocellulose literature and recommended by experts (Klemm et al., 2011; Milanez et al., 2013; Siqueira et al., 2010; Siró & Plackett, 2010), as can be seen from Table 1. The search expression encompassed different words that refer to cellulose nanocrystals, cellulose nanofibrils, and bacterial cellulose as well as other generic forms, such as nanocellulose, cellulose nanoparticles, and cellulose nanofiller.

Besides the fact that there are two main types of nanocellulose (Klemm et al., 2011), the variety of search terms is partially consequence of research initiatives scattered across the twentieth century, especially in the 1980s, when researcher from the International Telephone and Telegraph Corporation patented a mechanical method to produce microfibrillated cellulose (Klemm et al., 2011; Tubark, Snyder & Sandberg, 1983). Bacterial cellulose was an object of research at least since its first documentation in a paper describing cellulose as result of bacterial fermentation (Brown, 1886). Among other initiatives, the international organization Technical Association of the Pulp and Paper Industry (TAPPI) performed, in 2011, a roadmap to standardize the nanocellulose terms to cellulose nanofibrils and cellulose nanocrystals (TAPPI, 2011).

There are some words on Table 1 using the prefix “micro”, which could indicate an idea of micrometric scale; however, according to some experts in nanocellulose (Isogai, 2013; Klemm et al., 2011; Siqueira, Bras, & Dufresne, 2010), these materials are cellulose in nanoscale due to the fact that there is at least one dimension that scale. In fact, the use of prefix “nano” as part of nanocellulose set of terms started only in 2001, concurrently with the unfolding of many nanotechnology research initiatives (Milanez et al., 2014).

The search was conducted in March 31th 2014 in the online Web of Science database (topic search) using the search expression from Table 1. Only articles that were contained in the CWTS Web of Science publication-level classification system criteria were used. The classification system takes into account only article, letter and review published from 2000 to 2013 and indexed in the Science Citation Index Expanded and the Social Science Citation Index. Moreover, to be part of one research area, a publication must be related, either directly or indirectly, to at least 49 other publications in terms of citation (Waltman & van Eck, 2012).

Table 1 Boolean search expression to retrieve the initial set of nanocellulose publications.

("bacterial cellulos*") OR ("cellulos* crystal*") OR ("cellulos* nanocrystal*") OR ("cellulos* whisker*") OR ("cellulos* microcrystal*") OR ("cellulos* nanowhisker*") OR ("nanocrystal* cellulos*") OR ("cellulos* nano-whisker*") OR ("cellulos* nano-crystal*") OR ("nano-crystal cellulos*") OR ("cellulos* micro-crystal*") OR ("cellulos* microfibril*") OR ("microfibril* cellulos*") OR ("cellulos* nanofibril*") OR ("nanofibril* cellulos*") OR ("micro-fibril* cellulos*") OR ("nano-fibril* cellulos*") OR ("cellulos* micro-fibril*") OR ("cellulos* nano-fibril*") OR ("cellulos* nanofiber*") OR ("nanocellulos*") OR ("cellulos* nanoparticle*") OR ("nano-cellulos*") OR ("nanoparticl* cellulos*") OR ("nanosiz* cellulos*") OR ("cellulos* nanofill*") OR ("nano-siz* cellulos*") OR ("cellulos* nano-fiber*") OR ("cellulos* nano-particle*") OR ("cellulos* nano-fill*") OR ("nano-particl* cellulos*")

Source: Developed considering nanocellulose-focused terms found in the literature (Klemm et al., 2011; Milanez, Amaral, Faria, & Gregolin, 2013; Milanez et al., 2014; Siqueira, Bras, & Dufresne, 2010; Siró & Plackett, 2010) and expert opinions.

Prior retrieval of nanocellulose research areas

Research areas that contained at least one publication from the nanocellulose set were retrieved from the CWTS Web of Science Publication-level database. In total, 533 research topics (clusters from the classification system) were found. These clusters showed large differences in terms of volume (number of publications included). The largest cluster contains 2,751 publications whereas the smallest one covers only 50 publications. Almost 80% of these clusters contained less than three publications from the initial set.

Interestingly, we found that two research areas (clusters) included 56.3% of the initial nanocellulose set of publications. Moreover, in these two clusters, more than 80% overlapped with the initial set. Their descriptive labels also pointed towards nanocellulose research. Therefore, they were considered as nuclei of research in nanocellulose. Other clusters in which the representation of the initial set was much lower, were considered peripheral research areas and their relevance to nanocellulose research was evaluated (see next section).

Analysis of retrieved research area and cleaning of the initial set

An analysis of the content of publications in the peripheral research areas was conducted. We wanted to check whether these articles focused on the nanomaterial as an object of research. If not, they were considered noise. Because an evaluation of all research area retrieved would be too labour intensive, we made a selection. The checking task was performed only on those clusters that matched one of the following criteria:

- Research topics that contained at least 20 publications from initial dataset;
- Research topics of which at least 5% overlapped (percentage proportion) with the initial set.

A total of 20 (peripheral) clusters were evaluated. The analysis regarded only articles from the initial dataset of nanocellulose publications assigned to these clusters. The task involved reading each title to decide whether the article was a study focused upon nanocellulose or not. When the title was not clear, the abstract was also consulted.

Once the checking process was completed, specific terms were identified and considered to clean the initial set of nanocellulose publications. Only research topics with high percentage of “noise publication” were used. The presence of “noise publications” is usual in bibliometric analysis because there is no exhaustive search, but the idea is to gain more precision during the cluster retrieving.

Noun-phrases were obtained with support of VOSviewer corpus map analysis applied to titles and abstracts from publications belonging to these clusters, which were processed separately. The cleaning-terms extraction took into account only high frequency terms from clusters contents. However it is important to highlight that not every highly frequent words found were used. For instance, “ethanol” (alone) is one of the most frequently word from cluster 16.3.2, but it was not incorporated on the set of cleaning terms from Table 2 due to the fact that nanocellulose is a sub-product of ethanol produced from cellulose enzymatic fermentation, also called, second generation ethanol (Kangas et al., 2014; Song, Winter, Bujanovic, & Amidon, 2014; Tsukamoto, Durán, & Tasic, 2013; J. Y. Zhu, Sabo, & Luo, 2011). Thus, excluding publications containing “ethanol” would have an undesired effect on the final set of nanocellulose publications. This not happened with the word “ethanol yield”, though. In order to avoid such undesired effects, we studied the effect of each potential cleaning-term on the nuclei clusters before selecting the final ones.

Table 2 presents the final terms used to clean the nanocellulose-focused publications retrieved using the search expression from Table 1. They were applied on the title, abstract, author’s keyword and keyword plus search field. The effect of this cleaning task on the nuclei clusters and the peripheral clusters we used will be discussed in the results.

Table 2. Boolean expression of terms used to clean the nanocellulose-focused publications.

"gene" OR "xyloglucan" OR "microtubule" OR "*cyto*" OR "kinesi" OR "tubulin" OR "*cell wall*" OR "spindle" OR "phragmoplast" OR "mitosis" OR "preprophase" OR "phenotype" OR "*plant growth*" OR "meiosi" OR "*lignin distribution*" OR "delignification" OR "hemicellulose" OR "saccharification" OR "ethanol yield" OR "lignocellulos*" OR "glucosidase" OR "xylanase"

Source: Authors.

Final retrieving and selection of relevant research areas

The final set of nanocellulose publication comprised 2,600 nanocellulose publications (named now as core-nanocellulose) and they were assigned to 428 research areas, which still would be a highly number of cluster to be evaluated. In fact, 81.0% of these clusters included only one or two publications from the core-nanocellulose publication, which raised questions about their actual relevance to the advances on nanocellulose studies: the

maximum proportions are less than 3.3% and, in the cases where the absolute numbers are high, the proportion is less than 0.2%. Therefore, a selecting step was introduced.

We introduce here the *Pareto Principle* (or 80/20 rule). This principle states that “roughly 80% of the effects come from 20% of the causes” (Juran & Godfrey, 1998) and is found in bibliometric and library studies (Gupta, 1989; Kao, 2009; Stephens, Hubbard, Pickett, & Kimball, 2013). We hypothesize that 80% of the core set will be assigned to 20% of the areas. To reach these relevant research areas, the steps below were carried out:

1. The research areas were listed in descending order of the total number of publications from the core-nanocellulose;
2. Research topics with one or two publications from the core-nanocellulose were excluded. This yields 86 research areas;
3. The first 17 clusters were selected corresponding to 20% from 86 clusters.

We do not claim that our selecting procedure is perfect, but a quick analysis of the chosen research topics showed themes currently found in nanocellulose literature.

Independency test

An independency test was conducted to evaluate the effectiveness of the procedure proposed. The test involved retrieving the number of publication from the top five authors before and after cleaning the initial set of nanocellulose publications, obtaining the percentage of decrease and the position on a ranking of authors. We performed the analysis to the top authors from each nuclei research area.

Nanocellulose selected research area map and corpus analysis of nuclei

A map containing the seventeen selected research topics was created with support of VOSviewer software (van Eck & Waltman, 2014). The map positions the research topics on the basis of their citation relations, i. e. the closer two topics, the more frequent the citation traffic between them. The node labels also match the main content of the clusters.

In order to understand the difference between the two nuclei research areas, a map of noun-phrases based on text corpus (titles and abstracts) was created. The map compares the topics that characterize each nuclei cluster and provide themes that they share. Again, the map was developed with support of VOSviewer software (van Eck & Waltman, 2011, 2014).

In the next section we discuss the effect of cleaning up the core set of publications by using ‘cleaning terms’, i.e., terms to increase the accuracy of our initial set. Moreover, we present a basic structure of the field on the basis of the delineation we developed and further analysis.

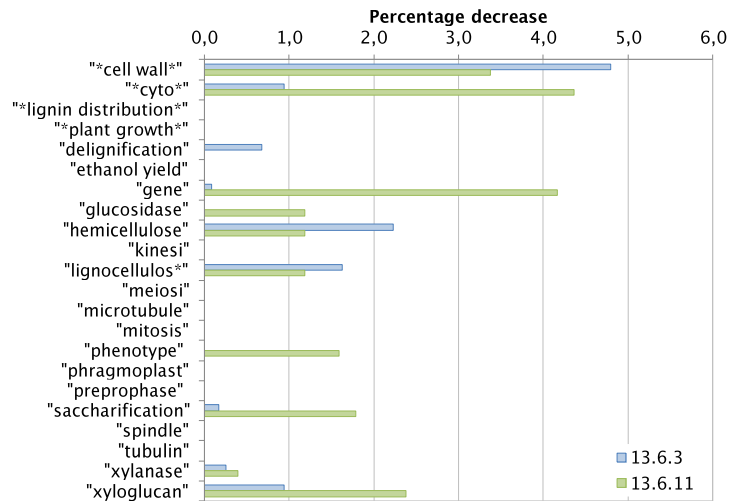
Results and discussion

Effect of cleaning the initial set of nanocellulose publications

Half of the 22 terms we used to clean the nanocellulose search strategy did not affect the coverage of core-nanocellulose publications in the nuclei research areas, as depicted in Figure 3. To the other half, none term could reduce the coverage in more than 5%. The terms that influenced research area 13.6.4 the most were “*cell wall*” and “hemicelluloses” while “*cyto*”, “gene” and “*cell wall*” were the ones that decreased the most core-nanocellulose coverage in cluster 13.6.11.

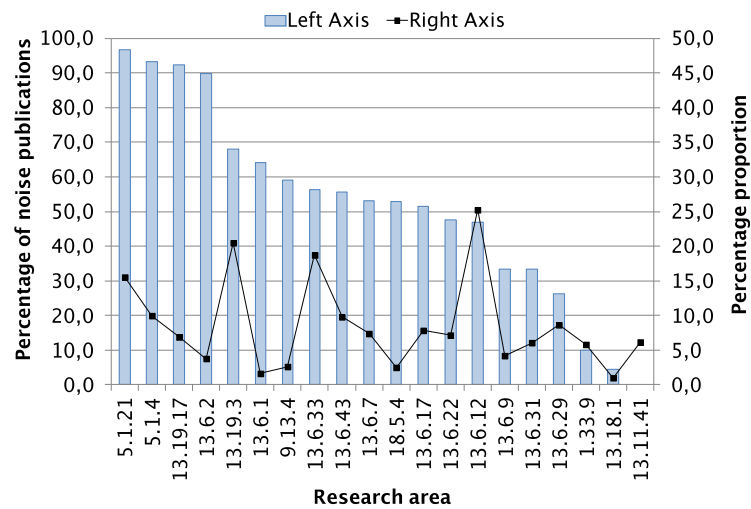
Overall, research topic 13.6.11 had its core-nanocellulose publication reduced in 17.5% while the decrease to cluster 13.6.3 was 10.2%. Nonetheless, both clusters still concentrated publication from the core-nanocellulose after the cleaning tasks (the proportion was 74.0% to research area 13.6.3 and 72.1% to 13.6.11). Therefore, they still had the status of nuclei research areas.

To the 20 peripheral research topics whose nanocellulose set of publication were evaluated, no direct correlation was observed between the proportional relevance of each clusters and the percentage of noise, according to Figure 4. Four research topics had a high percentage (>70%) of ‘noisy’ publications mainly focusing on biological issues of plants, ethanol production, and enzymes aspects, not having the nanomaterial as a final object of research. Since these four were used to select the cleaning terms, the cleaning affected them highly. Two of them were even eliminated (5.1.21 and 13.19.17). Furthermore, other peripheral clusters had their nanocellulose publication coverage diminished, as shown on Figure 5.



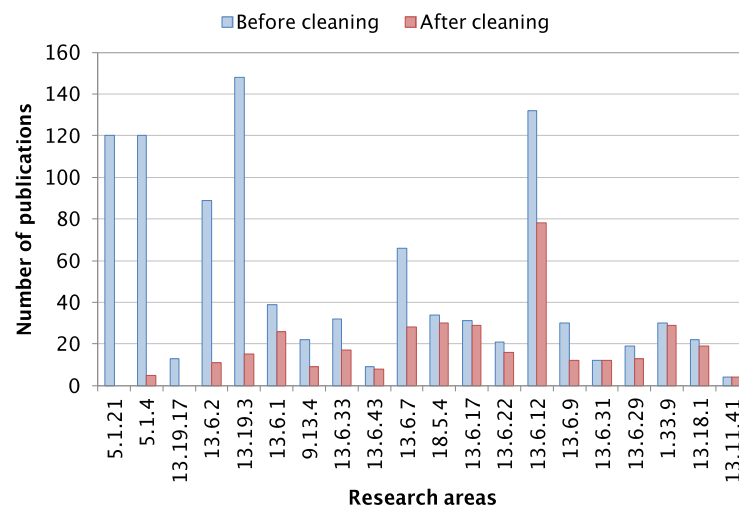
Source: CWTS Web of Science Publication-level database.

Fig. 3 Effect of cleaning terms on the number of publication from nuclei research areas.



Source: CWTS Web of Science Publication-level database.

Fig. 4 Percentage of noise of core-nanocellulose publications and proportion between core-nanocellulose publications and total number of publications over research area.



Source: CWTS Web of Science Publication-level database.

Fig. 5 Effect of cleaning terms on the number of publication from selected peripheral research areas.

Figure 6 provides an overview of the number of nanocellulose-focused publication and number of clusters (research areas) in four moments of the delineation procedure proposed. Considering the only first and the last moment, the number of nanocellulose-focused publication diminished 52.6% while the number of clusters decreased 96.8%. Nanocellulose-focused publications were more affected between moments 1 and 2, when the value reduced 30.1%. In contrast, the number of clusters diminished 79.9% and 80.2% respectively among the others moments.

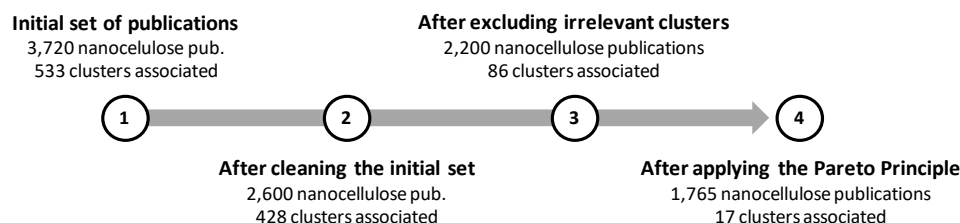


Fig. 6 Overview of the number of nanocellulose-focused publication and number of clusters (research areas) in four moments of the delineation procedure proposed.

Effect of cleaning procedure on top authors (independency test)

A second test verified the effect of the cleaning process on the coverage of key-authors (top 5) from each nuclei research area (13.6.3 and 13.6.11) in nanocellulose, as can be seen from Tables 3 and 4. Moreover, it is important to highlight that the top authors from cluster 13.6.3 are not the same top authors from the cluster 13.6.11 due to the fact that they simply did not overlap.

Table 3 Effect of cleaning step on the top authors from nucleus cluster 13.6.3.

Author	Number of Publications		Decrease (%)	Ranking	
	Before*	After*		Before*	After*
13.6.3 A	79	74	-6,33	1	1
13.6.3 B	47	40	-14,9	2	2
13.6.3 C	43	35	-18,6	3	5
13.6.3 D	42	37	-11,9	4	4
13.6.3 E	39	39	0	5	3

* Before and after the cleaning step.

Source: CWTS Web of Science Publication-level database.

To the ranking of authors from cluster 13.6.3, the number of publications decreases more significantly to authors B, C and D and a change on the ranking position occurred between author and author E. The cleaning step did not affect the publications from author E. In the case of cluster 13.6.11, the ranking of authors stay constant after the cleaning step despite the decrease in the number of publications from the authors (except author B), notably for authors A and C. Besides the effects verified, it is clear that the cleaning step did not affect the overall ranking of authors in both nuclei research areas. This fact supports the effectiveness of the delineation procedure used to retrieved nanocellulose research areas.

Table 4 Effect of cleaning step on the top authors from nucleus cluster 13.6.11.

Author	Number of Publications		Decrease (%)	Ranking	
	Before*	After*		Before*	After*
13.6.11 A	31	23	-25,8	1	1
13.6.11 B	22	22	0	2	2
13.6.11 C	21	17	-19,0	3	3
13.6.11 D	17	16	-5,88	4	4
13.6.11 E	17	16	-5,88	5	5

* Before and after the cleaning step.

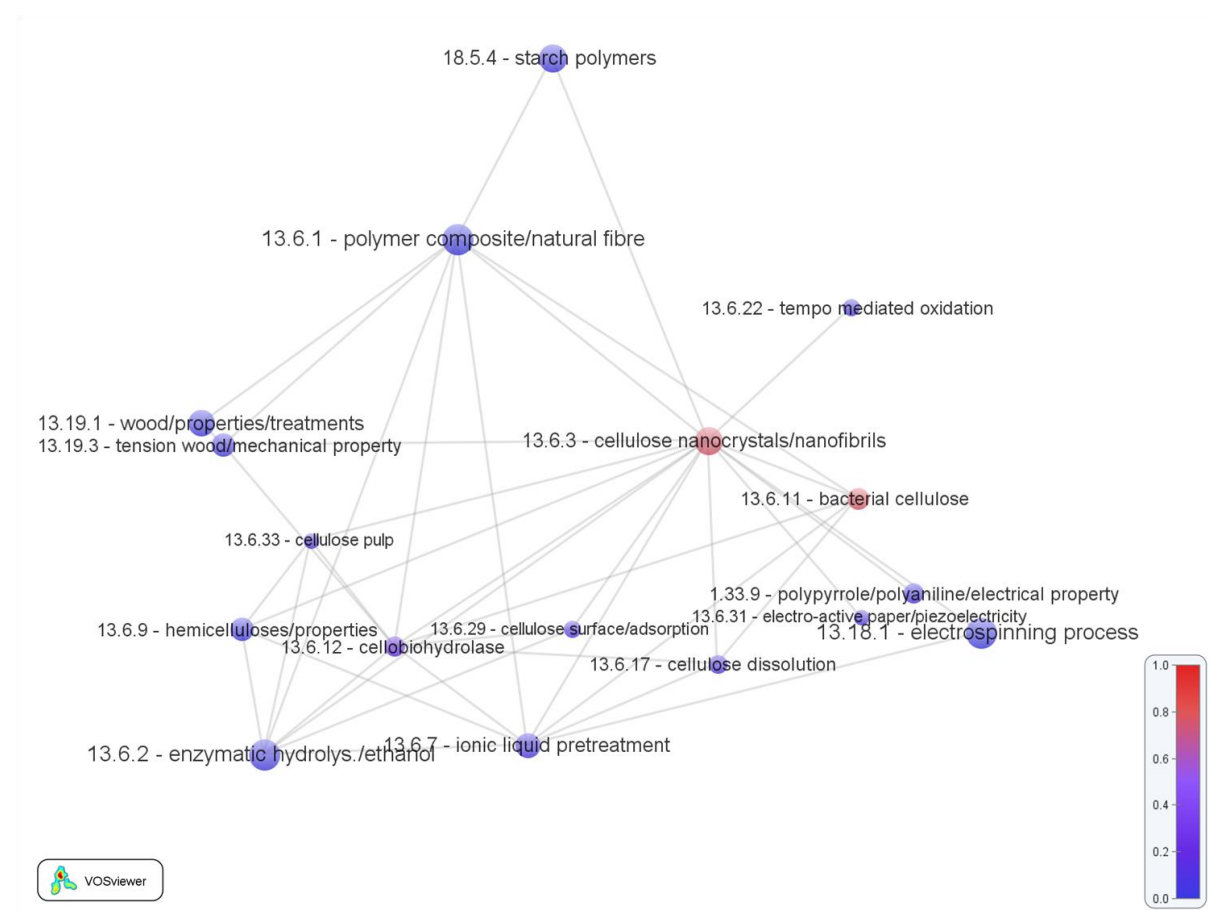
Source: CWTS Web of Science Publication-level database.

Map of the nanocellulose research topics

The delineating approach was able to retrieve two nuclei research areas, one associated with cellulose nanocrystals and nanofibrils and other to bacterial cellulose. The peripheral research topics regards biodegradable polysaccharides (starch polymers), hemicelluloses, polymer composites based on natural fibres, intrinsically conducting polymers, and cellulose electro-active paper. Other peripheral research areas included enzymatic hydrolyses and ethanol production, cellobiohydrolyse, cellulose pulp and cellulose dissolution, cellulose surface and ionic liquid pre-treatment. Wood, tension wood, electrospinning process and tempo mediated oxidation, which is an treatment that uses the chemical compound (2,2,6,6-Tetramethylpiperidin-1-yl)oxy (TEMPO), were also part of the final selection. These themes appears frequently in nanocellulose-focused studies (Azizi Samir, Alloin, & Dufresne, 2005; Charreau, Foresti, & Vazquez, 2013; Chirayil, Mathew, & Thomas, 2014; Dai et al., 2014; Domingues, Gomes, & Reis, 2014; Durán, Lemes, & Seabra, 2012; Eichhorn et al., 2010; Isogai, 2013; Klemm et al., 2011; Moon et al., 2011; Orts et al., 2005; Pääkkö et al., 2007; Siqueira et al., 2010; Siró & Plackett, 2010)

Figure 7 presents a map with these research topics (nodes). In this map all clusters (topics) are placed using the citation traffic among them. Clusters with dense citation traffic between them are in each other's vicinity while those with little traffic are remote from each other. The grey scale indicates the proportion of core publications being covered.

The nuclei research areas are darker and positioned slightly to the right centre of the map. Research area 13.6.3 (cellulose nanocrystals/microfibrillated cellulose) has citation connections to all clusters. On the other hand, research topic 13.6.11 (bacterial cellulose) is connected only with four other clusters, which might indicate its lower relevance than the other nucleus research area. At the top of the map, two research areas are located corresponding to starch polymers and polymer composites based on natural fibres. These research topics regard the development of sustainable materials (Durán et al., 2012; Moon et al., 2011; Siqueira et al., 2010; Isogai, 2013).



Source: CWTS Web of Science Publication-level database.

Fig. 7 Selected research areas according to the procedure proposed.

On the left part of the map research areas are found concerning wood and its mechanical properties. It is important to highlight that wood is a source of nanocellulose, especially because of interest from pulp industry (Moon et al., 2011; Klemm et al., 2011; Siqueira et al., 2010; Isogai, 2013; Milanez et al., 2014). These clusters

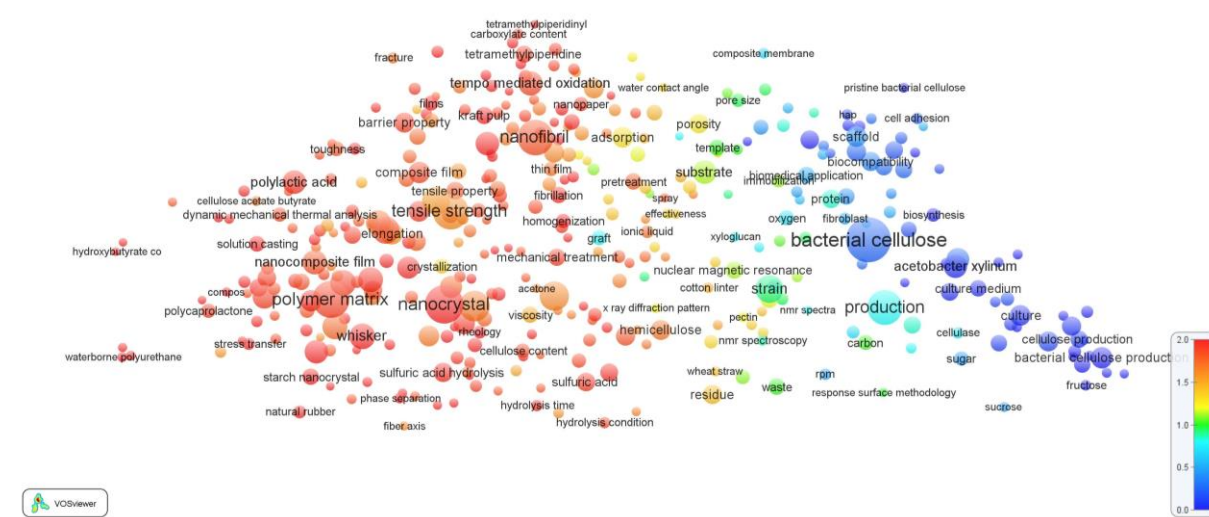
are near the cellulose pulp research area and the knowledge developed by the pulp industry in treatments of wood. The effect on the mechanical property may be useful for nanocellulose developments.

The research area of enzymatic hydrolysis (bottom left of the map) is close to the research topic cellobiohydrolase, i.e., enzymes that perform the process of hydrolyse, and ionic liquid pre-treatment, which also relies on enzymatic approaches. Indeed, one of them was considered as highly noisy (13.6.2), but we should take into account that nanocellulose obtainment has been also studied as a secondary product of bio-ethanol production (Beecher, 2007; Zhu, Sabo, & Luo, 2011). Enzymatic pre-treatment has been researched to improve nanocellulose defibrillation (Pääkkö et al. 2007; Moon et al., 2011; Klemm et al., 2011; Siqueira et al., 2010; Isogai, 2013). Moreover, hemicelluloses research area is also near these “pre-treatment-related” clusters as it is a polymer usually removed from plant cell walls when nanocellulose is produced (Moon et al., 2011; Klemm et al., 2011; Siqueira et al., 2010).

At the bottom right of the map, electrospinning process, conductive polymers, and electro-active cellulose-based papers are positioned together. There is, however, no relevant citation connection between them. Electrospinning is a technique used to produce micro- and nano-sized polymer-based fibres, and nanocellulose has been studied to improve the mechanical property of the final fibre (Dai et al., 2014). Nanocellulose electrical and magnetic properties have also been explored to be used with conductive polymers and optically transparent electro-active papers (Moon et al., 2011; Klemm et al., 2011). The other three research areas (cellulose dissolution, cellulose surface, and tempo mediated oxidation) are the smallest ones and their development must be monitored as they are associated to some issues concerning nanocellulose. For instance, tempo mediated oxidation is a current technique to perform pre-treatment of nanocellulose while there is high interests on cellulose surface research to develop new surface modifier to prevent aggregation from nanoscale cellulose particles (Klemm et al., 2011; Isogai, 2013).

Main topics shared by nuclei research areas

An analysis from the topics shared by the nuclei areas was performed with support of VOSviewer. These topics were extracted using NLP applied on titles and abstracts and the resulting noun-phrases are presented on Figure 8. The more the noun-phrase is located in the left part of the map the more it appeared on titles and abstracts from papers addressed to cluster 13.6.3 (red). Otherwise, the noun-phrase more to the right belongs to cluster 13.6.11 (blue). Noun-phrases between these areas represent topics that occurred constantly in both clusters.



Source: CWTS Web of Science Publication-level database.

Fig. 8 Noun-phrases from the nuclei research areas 13.6.3 (left side) and 13.6.11 (right side).

Noun-phrases from nuclei cluster 13.6.3 (cellulose nanocrystals/microfibrillated cellulose) refers to nanocelluloses produced by top-down approaches, its variables of processes, and its uses mainly in nanocomposites. On the other hand, nuclei cluster 13.6.11 (bacterial cellulose) contain noun-phrases related to nanocellulose produced using bottom-up approaches, such as biosyntheses, and its aspects. It was also verified that the subjects shared by these nuclei clusters refers mainly to properties and morphology of nanocellulose (for example, electrical conductivity, high tensile strength, high mechanical strength, cytotoxicity, crystallinity index, crystallite size, pore size, etc.); and characterization techniques (for instance, x-ray diffraction, nuclear magnetic resonance spectroscopy, scanning electron microscope image, Fourier transformation infrared spectroscopy).

Conclusion

The proposed delineation procedure enabled us to retrieve relevant publications from research areas involving nanocellulose. Seventeen research topics were identified, mapped and associated with current research challenges on nanocellulose. Two of them were highlighted as nuclei since they contain most part of the initial set of publications. The effect of the cleaning step on nuclei and peripheral clusters provided valuable feedback and demonstrated its importance to establishing relevant clusters afterwards. The independency test showed that the cleaning procedure did affect the number of publication from most of the top author. However, the author's ranking did not change significantly. In future research, we intend to develop new research aiming at excluding the manual checking on the peripheral clusters and include a procedure more automated.

Delineating scientific fields is a complex task as boundaries are not frequently well established since scientific studies have become more complex and interdisciplinary. More and more exchange of knowledge between scientists from different disciplines is involved. Our approach retrieves and delineates the real nuclei and the peripheral research areas concerning nanocellulose studies. This clear separation provides suggestions for further research, putting the nuclei research in context. We already check in this paper the share of topics between the nuclei research areas, but we also intend to understand the knowledge flow from peripheral research topics to the nuclei areas. Another idea involves the mapping of how the peripheral areas provide the necessary knowledge to face nanocellulose current challenges, and how countries and scientific institutions are contributing to this evolution. The investigation will evaluate the added value of peripheral clusters to the nuclei and what type of analyses they are necessary and in which cases they can be ignored.

Acknowledgments

The authors are grateful to the São Paulo Research Foundation (process number 2012/16573-7) and comments from researchers of CWTS, NIT/Materiais, and ISSI Conference. We are also thankful to the Graduate Program in Materials Science and Engineering at the Federal University of São Carlos for supporting this work. We also acknowledge the nanocellulose researchers for their valuable opinion on the research areas and on the map of topics.

References

- Azizi Samir, M. A. S., Alloin, F., & Dufresne, A. (2005). Review of recent research into cellulosic whiskers, their properties and their application in nanocomposite field. *Biomacromolecules*, 6(2), 612–26. doi:10.1021/bm0493685
- Beecher, J. (2007). Wood , trees and nanotechnology. *Nature Nanotechnology*, 2(August), 466–467.
- Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, 64(3), 351–374.
- Boyack, K. W., Newman, D., Duhon, R. J., Klavans, R., Patek, M., Biberstine, J. R., ... Börner, K. (2011). Clustering more than two million biomedical publications: comparing the accuracies of nine text-based similarity approaches. *PloS One*, 6(3), e18029. doi:10.1371/journal.pone.0018029
- Brown, A. J. (1886). On an acetic ferment which forms cellulose. *Journal of the Chemical Society, Transactions*, 49, 432–439. doi:10.1039/ct8864900432
- Charreau, H., Foresti, M. L., & Vazquez, A. (2013). Nanocellulose patents trends: a comprehensive review on patents on cellulose nanocrystals, microfibrillated and bacterial cellulose. *Recent Patents on Nanotechnology*, 7(1), 56–80. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/22747719>
- Chirayil, C. J., Mathew, L., & Thomas, S. (2014). Review of recent research in nanocellulose preparation from different lignocellulosic fibers. *Review of Advanced Materials Science*, 37, 20–28.
- Dai, L., Long, Z., Ren, X., Deng, H., He, H., & Liu, W. (2014). Electrospun polyvinyl alcohol/waterborne polyurethane composite nanofibers involving cellulose nanofibers. *Journal of Applied Polymer*, 41051, 1–6. doi:10.1002/app.41051
- Domingues, R. M. A., Gomes, M. E., & Reis, R. L. (2014). The potential of cellulose nanocrystals in tissue engineering strategies. *Biomacromolecules*, 15, 2327–2346.
- Dufresne, A. (2013). Nanocellulose: A new ageless bionanomaterial. *Materials Today*, 16(6), 220–227.

- Durán, N., Lemes, A. P., & Seabra, A. B. (2012). Review of cellulose nanocrystals patents: preparation, composites and general applications. *Recent Patents on Nanotechnology*, 6(1), 16–28. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/21875405>
- Eichhorn, S. J., Dufresne, A., Aranguren, M., Marcovich, N. E., Capadona, J. R., Rowan, S. J., ... Peijs, T. (2010). Review: current international research into cellulose nanofibres and nanocomposites. *Journal of Materials Science*, 45(1), 1–33. doi:10.1007/s10853-009-3874-0
- Gardner, D. J., Opo, Oporto, G. S., Mills, R., & Samir, M. A. S. A. (2008). Adhesion and Surface Issues in Cellulose and Nanocellulose. *Journal of Adhesion Science and Technology*, 22(5-6), 545–567. doi:10.1163/156856108X295509
- Glanzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, 56(3), 357–367.
- Gupta, D. . (1989). Scientometric study of biochemical literature of Nigeria, 1970-1984: application of Lotka's Law and the 80/20-rule. *Scientometrics*, 15(3-4), 171–179.
- Hullmann, A., & Meyer, M. (2003). Publications and patents in nanotechnology An overview of previous studies and the state of the art. *Scientometrics*, 58(3), 507–527.
- Igami, M. (2008). Exploration of the evolution of nanotechnology via mapping of patent applications. *Scientometrics*, 77(2), 289–308. doi:10.1007/s11192-007-1973-8
- Isogai, A. (2013). Wood nanocelluloses: fundamentals and applications as new bio-based nanomaterials. *Journal of Wood Science*, 59(6), 449–459. doi:10.1007/s10086-013-1365-z
- Juran, J. M., & Godfrey, A. B. (Eds.). (1998). *Juran's quality handbook* (5th ed., p. 1730). New York: McGraw-Hill.
- Kangas, H., Tamminen, T., Liitia, T., Hakala, T. K., Vorweg, W., & Poppius-Levlin, K. (2014). Lignofibre (lgf) process – a flexible biorefinery for lignocellulosics. *Cellulose Chemistry and Technology*, 48(9-10), 765–771.
- Kao, C. (2009). The authorship and internationality of Industrial Engineering journals. *Scientometrics*, 81(1), 123–136. doi:10.1007/s11192-009-2093-4
- Klemm, D., Kramer, F., Moritz, S., Lindström, T., Ankerfors, M., Gray, D., & Dorris, A. (2011). Nanocelluloses: a new family of nature-based materials. *Angewandte Chemie (International Ed. in English)*, 50(24), 5438–66. doi:10.1002/anie.201001273
- Kostoff, R. N., Koytcheff, R. G., & Lau, C. G. Y. (2009). Seminal Nanotechnology Literature: A Review. *Journal of Nanoscience and Nanotechnology*, 9(11), 6239–6270. doi:10.1166/jnn.2009.1465
- Leydesdorff, L., Carley, S., & Rafols, I. (2013). Global maps of science based on the new Web-of-Science categories. *Scientometrics*, 94(2), 589–593. doi:10.1007/s11192-012-0784-8
- Leydesdorff, L., de Moya-Anegón, F., & Guerrero-Bote, V. P. (2014). Journal maps, interactive overlays, and the measurement of interdisciplinarity on the basis of Scopus data (1996–2012). *Journal of the Association for Information Science and Technology*, n/a–n/a. doi:10.1002/asi.23243
- Mariano, M., Kissi, N. El, & Dufresne, A. (2014). Cellulose nanocrystals and related nanocomposites: review of some properties and challenges. *Journal of Polymer Science*, 52, 791–806. doi:10.1002/polb.23490
- Milanez, D. H., Amaral, R. M. Do, Faria, L. I. L. De, & Gregolin, J. A. R. (2013). Assessing nanocellulose developments using science and technology indicators. *Materials Research*, 16(3), 635–641. doi:10.1590/S1516-14392013005000033
- Milanez, D. H., Conserva, A. C. A., Amaral, R. M., Faria, L. I. L., Gregolin, J. A. R., Carlos, A., & Conserva, A. C. A. (2014). Análise de bases de dados e termos de busca para estudos bibliométricos e monitoramento científico em nanocelulose. *Em Questão*, 20(3).
- Milanez, D. H., Faria, L. I. L., Amaral, R. M., Leiva, D. R., & Gregolin, J. A. R. (2014). Patents in nanotechnology: an analysis using macro-indicators and forecasting curves. *Scientometrics*. doi:10.1007/s11192-014-1244-4
- Moed, H. F., Glänzel, W., & Schmoch, U. (2004). *Handbook of quantitative science and technology research: the use of publication and patent statistics in studies of S&T systems*. (H. F. Moed, W. Glänzel, & U. Schmoch, Eds.) (Kluwer Aca., p. 785). New York: Kluwer Academic Publishers.
- Mogoutov, A., & Kahane, B. (2007). Data search strategy for science and technology emergence: A scalable and evolutionary query for nanotechnology tracking. *Research Policy*, 36(6), 893–903. doi:10.1016/j.respol.2007.02.005
- Moon, R. J., Martini, A., Nairn, J., Simonsen, J., & Youngblood, J. (2011). Cellulose nanomaterials review: structure, properties and nanocomposites. *Chemical Society Reviews*, 40(7), 3941–94. doi:10.1039/c0cs00108b
- Neuhaus, C., & Daniel, H. D. (2009). A new reference standard for citation analysis in chemistry and related fields based on the sections of Chemical Abstracts. *Scientometrics*, 78(2), 219–229. doi:10.1007/s11192-007-2007-2

- Okubo, Y. (1997). *Bibliometric Indicators and Analysis of Research Systems: Methods and Examples* (No. 01). doi:10.1787/208277770603
- Orts, W. J., Shey, J., Imam, S. H., Glenn, G. M., Guttman, M. E., & Revol, J.-F. (2005). Application of Cellulose Microfibrils in Polymer Nanocomposites. *Journal of Polymers and the Environment*, 13(4), 301–306. doi:10.1007/s10924-005-5514-3
- Pääkkö, M. et al. (2007). Enzymatic hydrolysis combined with mechanical shearing and high-pressure homogenization for nanoscale cellulose fibrils and strong gels. *Biomacromolecules*, 8(6), 1934–41. doi:10.1021/bm061215p
- Raan, A. F. J. Van. (2014). Advances in bibliometric analysis : research performance assessment and science mapping. In W. Blockmans, L. Engwall, & D. Weaire (Eds.), *Bibliometrics: Use and abuse in the review of research performance* (Vol. c, pp. 17–28). Portland.
- Reuters, T. (2014). Web of Science. Retrieved March 03, 2014, from http://apps.webofknowledge.com/WOS_GeneralSearch_input.do?product=WOS&SID=4CGNLFV3uyF6vP2Mvti&search_mode=GeneralSearch
- Ruiz-Castillo, J., & Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of Science. *Journal of Informetrics*, 9(1), 1–39. doi:10.1016/j.joi.2014.11.010
- Siqueira, G., Bras, J., & Dufresne, A. (2010). Cellulosic Bionanocomposites: A Review of Preparation, Properties and Applications. *Polymers*, 2(4), 728–765. doi:10.3390/polym2040728
- Siró, I., & Plackett, D. (2010). Microfibrillated cellulose and new nanocomposite materials: a review. *Cellulose*, 17(3), 459–494. doi:10.1007/s10570-010-9405-y
- Song, Q., Winter, W. T., Bujanovic, B. M., & Amidon, T. E. (2014). Nanofibrillated cellulose (NFC): A high-value co-product that improves the economics of cellulosic ethanol production. *Energies*, 7(2), 607–618. doi:10.3390/en7020607
- Stephens, J., Hubbard, D. E., Pickett, C., & Kimball, R. (2013). Citation Behavior of Aerospace Engineering Faculty. *The Journal of Academic Librarianship*, 39(6), 451–457. doi:10.1016/j.acalib.2013.09.007
- TAPPI. (2011). *Roadmap for the development of international standards for nanocellulose* (p. 36).
- Tsukamoto, J., Durán, N., & Tasic, L. (2013). Nanocellulose and bioethanol production from orange waste using isolated microorganisms. *Journal of the Brazilian Chemical Society*, 24(9), 1537–1543. doi:10.5935/0103-5053.20130195
- Tubark, A. F., Snyder, F. W. & Sandberg, K. R. Microfibrillated cellulose, a new cellulose product: properties, uses, and commercial potential. *Journal of Applied Polymer Science: Applied Polymer Symposium*, v. 30, p. 815-827, jan. 1983.
- Van Eck, N. J., & Waltman, L. (2011). Text mining and visualization using VOSviewer. *ISSI Newsletter*, 7(3), 50–54.
- Van Eck, N. J., & Waltman, L. (2015). VOSviewer. Retrieved July 10, 2015, from: <http://www.vosviewer.com/>
- Van Raan, A. F. J. (2004). Science measuring. In *Handbook of quantitative science and technology research: the use of publication and patent statistics in studies of S&T systems* (pp. 19–50). New York: Kluwer Academic Publishers.
- Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, 63(12), 2378–2392. doi:10.1002/asi.22748
- Waltman, L., van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629–635. doi:10.1016/j.joi.2010.07.002
- Wang, L., Notten, A., & Surpatean, A. (2012). Interdisciplinarity of nano research fields: a keyword mining approach. *Scientometrics*, 94(3), 877–892. doi:10.1007/s11192-012-0856-9
- Zhu, J. Y., Sabo, R., & Luo, X. (2011). Integrated production of nano-fibrillated cellulose and cellulosic biofuel (ethanol) by enzymatic fractionation of wood fibers. *Green Chemistry*, 13(5), 1339. doi:10.1039/c1gc15103g
- Zhu, J. Y., Sabo, R., & Luo, X. (2011). Integrated production of nano-fibrillated cellulose and cellulosic biofuel (ethanol) by enzymatic fractionation of wood fibers. *Green Chemistry*, 13(5), 1339. doi:10.1039/c1gc15103g
- Zhu, Y., & Yan, E. (2015). Dynamic subfield analysis of disciplines: an examination of the trading impact and knowledge diffusion patterns of computer science. *Scientometrics*, 335–359. doi:10.1007/s11192-015-1594-6
- Zitt, M. (2015). Meso-level retrieval: IR-bibliometrics interplay and hybrid citation-words methods in scientific fields delineation. *Scientometrics*, 102(3), 2223–2245. doi:10.1007/s11192-014-1482-5