CHALMERS

Development of tools and methods for scientometrics
The case of a bibliometric study on batteries and fuel cells

Master of Science Thesis

HICHAM SABIR

Department of Applied Physics
Chalmers University of Technology
Göteborg, Sweden, 2010

Development of tools and methods for scientometrics: the case of a bibliometric study on batteries and fuel cells
HICHAM SABIR

hicham.sabir@imelavi.fr

Department of Applied Physics
Chalmers University of Technology
SE-412 96 Göteborg
Sweden

Collaborator:

Dr. Eric Archambault
Science-Metrix Inc., 1335. Mont-Royal E. Montréal, Québec, H2J 1Y6, Canada
eric.archambault@science-metrix.com
www.science-metrix.com

Abstract

This report sums up our work on the development of bibliometric indicators. It focuses on citation analysis, collaboration systems and the interpretation of bibliometric point clouds. We particularly focused on the role of power-laws in the description of bibliometric systems, the introduction of the probabilistic collaborative index, fitting bibliometric datasets using spline interpolation and smoothing and looking at the result of moving average methods on point clouds.

These methods were applied to the fuel cell and battery science systems. This allowed us to draw a picture of the landscape around the different actors and technologies, which can play a key role in the future of renewable energies. We saw in bibliometrics a powerful tool to globally evaluate existing technologies, study interaction networks between the main research groups and evaluate the impact of scientific research outside the academic world. We were able to highlight the important place of Asia in the international collaboration system, the quasi-absence of European central hub and many features which could give a guideline to European policy makers in these two fields. We also used bibliometrics as a tool for innovation by looking at research trends on the different types of fuel cells.

À Madeleine, Tarcizio, Mina, Hmed …

ترســــيزيو و مـــدلين ،حمـــد ،لمينـــة

Table of Content

Acknowledgement

I would like to thank Dr. Éric Archambault for welcoming me at Science-Metrix in Montreal, a very dynamic and fast growing company, specialized in science and technology evaluation. I am also very grateful to David Campbell and Grégoire Côté for sharing their knowledge and experience on bibliometrics. These lines would be incomplete without a few words to Science-Metrix team which helped me have a wonderful time both at work and in Montreal.

I would also like to thank my examiner Pr. Aleksandar Matic from the department of applied physics of Chalmers University of Technology for his support and his precious insight into battery and fuel cell technologies.

Table of figures

Introduction

Scientometrics is at the intersection between mathematics, information science and sociology and provides decision-makers with a solid and approved theory, based on indicators and trend analysis. The most widely known application of scientometrics is the international ranking of universities. It is a field which is often unknown or neglected by physicists. However, researchers are more and more aware of the impact it had on the assessment of their research. In fact, scientometrics can be simply viewed as the formalization of methods and approaches which are part of scientist's daily activity.

Bibliometrics on the other hand is the discipline of scientometrics which quantizes the importance of emerging ideas, trends and fields in science. It uses tools such as citation analysis and content analysis which we combined with innovation-oriented concepts such as S-curves analysis for innovation.

The objective of this work is to use bibliometrics to draw a picture of the technological landscape around batteries and fuel cells. Since bibliometrics is often used for the publication of reports for decision makers, we decided to apply its methods to the particular case of battery and fuel cell technologies. This choice was motivated by the difficulty we face today when trying to picture the landscape around the different actors and technologies, which can play a key role in the future of renewable energies. We saw in bibliometrics a powerful tool to globally evaluate existing technologies, study interaction networks between the main research groups and evaluate the impact of scientific research outside the academic world.

In order to reach these objectives, this report will start with a presentation of the main tools and methods of bibliometrics. It will later lead us to consider more carefully certain phenomenon and develop new approaches to solve some of the resulting contradictions. After presenting the role of power laws in the bibliometric system, we will highlight some of the flaws of the indicators currently used. We will also investigate the particular case of the collaboration system to present a new probabilistic approach to evaluate the expected collaboration score of countries and institutions.

This report is constructed around two main parts: the first one deals with the development of bibliometric methods and will present our findings on matters such as scale free indicators and probabilistic models for the description of international collaboration. The second one will use these techniques to perform a detailed scientometric description of the landscape in battery and fuel cell research.

I. Bibliometric Methods

Scientometrics and bibliometrics are usually not very well known among researchers today. However, they represent a powerful tool to study the global behavior of research and scientific production. Scientometric tools can be used to measure and compare the scientific activities at various levels of aggregation including institutions, sectors, provinces and countries. They can also be used to measure research collaborations, to map scientific networks and to monitor the evolution of scientific fields (Science-Metrix Corp., 2010).

We will detail here the development of a few bibliometric indicators and methods which will be used later in the fuel-cell and batteries case study. The technical details described here led to three papers which have been submitted for publication[1].

1. *Power Laws in Bibliometrics*

a. Description

Power laws were identified to be relevant in the study of scientific publication systems in the seventies through the bibliographic studies of journals, articles and citations (Naranan, 1970),(Naranan, 1971). Although the implications of this finding were not immediately understood (Hubert, 1976), the model appeared to be efficient in developing the understanding of scientists productivity (Allison, 1980) and of the behavior of the science systems (Katz J. S., 1999). After the fractal behavior of systems following a power law had been described in finance (Mandelbrot, 1997), arguments in favor of a scale-independent description of bibliometrics were proposed (Katz J. S., 2000).

In our case, we aim at using this power law description of bibliometric systems to derive an index of performance. When considering the number of citations received by a country, not only do we want to determine the general behavior of the system by fitting the data with a power law, but we also want to quantize the performance of each country. The objective is to identify the countries or institutions who are receiving more citations than they should according to their size, revealing an unusually high impact.
Another example where the same index would be used is the evaluation of the collaborative performance of countries or institutions according to their size. Figure 1 shows on a log-log scale the number of collaboration versus the number of papers for

[1] On June 16, 2010 the submission process was not over and changes were still being made to the articles which prevents us from giving a full text or reference here.

the 100 biggest countries[2]. As can be seen in the graph, the exponent of the law relationship is 0.86 which means that the bigger a country is, the harder it is for it to have international collaborations. Moreover, the fit using a power law is rather satisfactory, with an R-squared coefficient of 0.97.

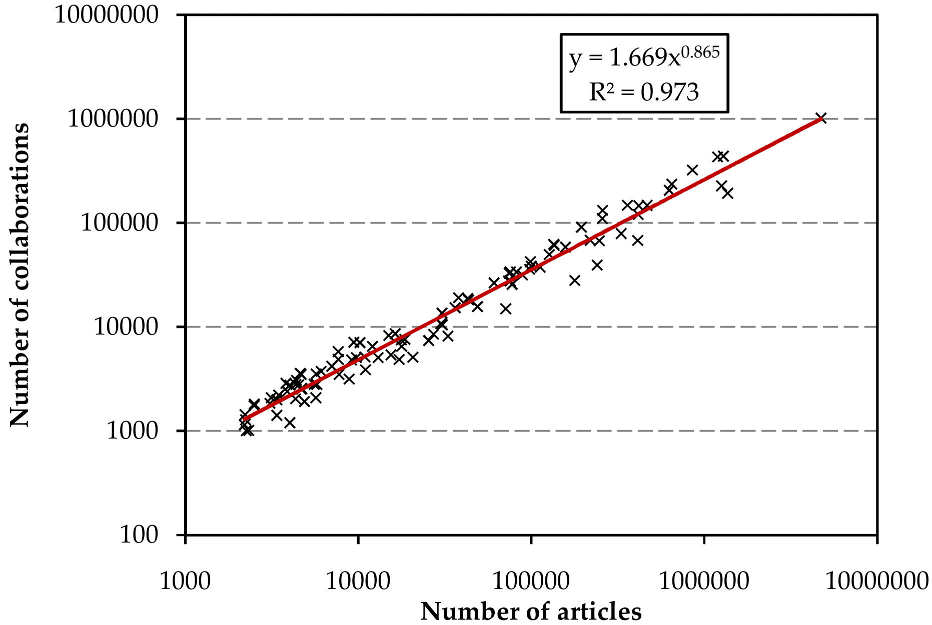


Figure 1: The international collaboration system

The simplest index which has been used will be called here the *Simple Ratio* (SR). With this index, if a country publishes 100 papers of which 60 are international collaborations, its simple ratio is SR=0.6. The countries can then be ranked according to their SR.
However, this approach presents a major flaw: it does not take account of the general behavior of the system. A new index was then developed, called here the *Scale Independent Ratio* (ScIR) based on the underlying power law behavior of the system.

As exposed earlier, the simple ratio indicator does not take account of the global scale of the system. The objective here will be to analytically describe where the bias lies, propose an alternative unbiased indicator and compare its behavior with the classic one.

Let's consider a statistical data set following a power law, i.e. the number of citations received by a country versus the number of publications of that country. The data set can be written $\{C_i\}$ where $\forall\ i \in [1;n]\ C_i = (x_i, y_i)$.

---

[2] By biggest countries we mean the countries with the largest number of peer-reviewed publications between 1996 and 2009.

Because the data set follows a power law, we can define c and p such as $\forall\, i \in [1;n]$ $\hat{y}_i = c\, x_i^p$ where $\hat{y}$ is the estimated value of y.

The first indicator using a simple ratio (SR) is defined by equation (1) and a second one, introduced previously is defined by equation (2):

$$\forall i \in [1;n]\; SR_i = \frac{y_i}{x_i} \tag{1}$$

$$\forall i \in [1;n]\; ScIR_i = \frac{y_i}{\hat{y}_i} \tag{2}$$

b. The problem of scale dependence

The purpose of the indicators introduced previously is to evaluate the performance of each element of the system. For instance, when considering the citations received by a sub-group (e.g. country or institution), a high value of the indicator will point out that the country's scientific production is of good quality.

This indicator is used to establish a ranking between the countries or institutions. We will therefore see what the simple ratio (SR) tells us in the case of two countries C_i and C_j with the same score:

C_i and C_j have the same score

$\Leftrightarrow SR_i = SR_j$

$\Leftrightarrow \frac{y_i}{x_i} = \frac{y_j}{x_j} \Leftrightarrow \frac{c\, x_i^p}{x_i} = \frac{c\, x_j^p}{x_j} \Leftrightarrow x_i^{p-1} = x_j^{p-1}$

$\Leftrightarrow x_i = x_j$ or $p = 1$

The first outcome $x_i = x_j$ is absurd since C_i and C_j are two different countries.

These lines show that indicators using a simple ratio are biased and are only valid for linear systems, i.e. p = 1.

Moreover, it is impossible to derive an average or expected behavior, noted $\overline{SR}$, in the general case:

For linear systems:

$$\forall i \in [1;n]\; SR_i = \frac{a\, x_i + \varepsilon_i}{x_i} = a + \frac{\varepsilon_i}{x_i} \;\; and \;\; \overline{SR} = a$$

where ε_i is the distance between the i-th element and the expected behavior. This distance can be considered as a random fluctuation around the expected value.

If the k-th element is medium, $\varepsilon_k = 0 \Rightarrow SR_k = a = \overline{SR}$ which satisfies the initial definition for the indicator.

For non linear systems (case of a power law):

$$\forall i \in [1;n]\; SR_i = \frac{c\, x_i^p + \varepsilon_i}{x_i} = c\, x_i^{p-1} + \frac{\varepsilon_i}{x_i} \;\; and \;\; \overline{SR} = \frac{c}{n} \sum_i x_i^{p-1}$$

If the k-th element is medium, $\varepsilon_k = 0 \Rightarrow SR_k = c\, x_i^{p-1} \neq \frac{c}{n} \sum_i x_i^{p-1}$, which does not satisfy the initial definition for an indicator.

c. A relevant scale dependent operator

The power of the new operator lies in the fact that it can handle both linear and non-linear behaviors.
As done previously, let's consider two countries Ci and Cj:

C_i and C_j have the same score

$\Leftrightarrow ScIR_i = ScIR_j$

$$\Leftrightarrow \frac{y_i}{c\, x_i^p} = \frac{y_j}{c\, x_j^p} \Leftrightarrow \frac{c\, x_i^p + \varepsilon_i}{c\, x_i^p} = \frac{c\, x_j^p + \varepsilon_j}{c\, x_j^p} \Leftrightarrow 1 + \frac{\varepsilon_i}{c\, x_i^p} = 1 + \frac{\varepsilon_j}{c\, x_j^p}$$

$$\Leftrightarrow \frac{\varepsilon_i}{c\, x_i^p} = \frac{\varepsilon_j}{c\, x_j^p}$$

$\Leftrightarrow$ they have the same deviation from the model.

Moreover, we can easily define the average or expected behavior: An item k is here said to have a medium behavior with respect to the new indicator if $ScIR_k = 1$.

For a system governed by a given relation f:

$$\forall i \in [1;n]\; ScIR_i = \frac{y_i}{\widehat{y_i}} = \frac{f(x_i) + \varepsilon_i}{f(x_i)} = 1 + \frac{\varepsilon_i}{f(x_i)}$$

If the k-th element is medium, $\varepsilon_k = 0 \Rightarrow ScIR_k = 1$, which satisfy the initial definition for this indicator.

d. Validation

The indicator we defined takes in account the underlying non-linearity of the statistical data set. However, it is coherent with the old one since there is a simple relation between both operators:

$$\forall i \in [1;n]\ \frac{ScIR_i}{SR_i} = \frac{y_i}{\hat{y}_i}\frac{x_i}{y_i} = \frac{x_i}{c\,x_i^p}$$

Therefore,

$$\forall i \in [1;n]\ \ ScIR_i = \frac{1}{c}x_i^{1-p}SR_i \tag{3}$$

When the non linearity of the system is weak ($p\sim1$), this relation becomes linear

$$ScIR = \frac{1}{c}SR \tag{4}$$

This work allowed us to propose a mathematical description of a scale independent index which gives a coherent quantization of countries and institutions' performances. Table 6 page 78 sums up the $ScIR$ values obtained on the world collaboration system of Figure 1.

2. ***Probabilistic Collaborative Index***

Collaboration is often a central feature in bibliometric analysis. Several indexes have been developed in order to quantify and evaluate the collaborative effort of researches, institutions and countries (Narvaez-Berthelemot, 1990), (Miquel & Okubo, 1994). A fairly intuitive approach consists in saying that the more productive a country is, the harder it is for him to have international collaborations.

Collaboration is here defined by the association of two or more researchers for the publication of a work. From a strictly probabilistic point of view, a scientist who belongs to a country producing 80% of the world's scientific production has only 1/5 chances to collaborate with a foreign scientist if picked randomly. In the mean time, a scientist who belongs to a country producing 1% of the world scientific production has 9/10 chances to collaborate with a foreign colleague on his next article.

This approach was used to develop indicators based on the comparison between the real score of countries and that expected from a random behavior (Zitt, Bassecoulard, & Okubo, 2000). This so-called probabilistic affinity index (PAI) is derived, for two given countries, as the ratio between the observed number of bilateral collaborations and the expected number of collaborations given a random distribution of each country's international collaborations.

Another method was derived from the existing probabilistic methods and used Montecarlo simulations instead of probability calculations in order to determine the expected number of collaboration (Yamashita & Okubo, 2006). This so-called probabilistic partnership index (PPI) was initially used to investigate inter-sectoral cooperation between France and Japan.

When evaluating the scientific potential of a country, its performance regarding international collaboration is a key issue. However, as explained in the last section, collaboration on the world scale can be described using power laws. It appears that, when the PAI and PPI were developed, no observation was made regarding the linearity of the data distribution. Moreover, the PAI is not suitable to quantize the global collaborative score of a country, since it takes it as input. In order to calculate the expected values between two countries, the PAI re-distributes the total number of collaborations of that country randomly among all the others (Zitt, Bassecoulard, & Okubo, 2000).

To obtain such information, we developed a new index called the Probabilistic Collaborative Index (PCI). The concept behind the PCI is quite similar to that of the PAI except that we take the total amount of collaborations in the world and re-distribute it among the different countries, according to their size. For a given country, it is defined as the ratio between the observed number of collaborations and the expected one.

a. Theory

For one subgroup A the number of international bilateral collaborations which can be constructed from all the articles it appears in can be calculated according to eq. (5):

$$C_A = \sum_{i\ articles} \left[\binom{n_i}{2} - \binom{n_i - na_i}{2} - \binom{na_i}{2}\right] = \sum_i [(n_i - na_i)\, na_i] \tag{5}$$

Where n_i is the number of collaborators in the article i and na_i is the number of collaborators from country A in the same article i. $\binom{na_i}{2}$ is then the number of bilateral national collaborations between elements form A in the article i.

Note that this demonstration will group publications by countries but A could also represent an institution, a research group, a field, etc.

Equation (5) will help us compute the observed collaborative value more easily, only knowing the share of country A in the total contributions to the papers. To illustrate this point, let's consider a paper with 3 contributions from country A, 2 from country X and 4 from country Y. The number of international collaborations for A in this paper is 3x(9-3)=18. C_A is then the sum over all the articles.

In order to calculate the PCI for this country, we need to evaluate the expected value of collaborations ($\hat{C}_A$) using the probabilistic approach detailed earlier. The probability to obtain international collaborations (P_{int}) is given by equation (6). They are obtained by dividing the maximum number of national couples that a country can have, with the total over all the countries.

$$P_{int}(A) = \frac{\binom{N}{2} - \binom{N-N_A}{2} - \binom{N_A}{2}}{\sum_X \binom{N}{2} - \binom{N-N_X}{2} - \binom{N_X}{2}} = \frac{N_A(N - N_A)}{\sum_X N_X(N - N_X)} \tag{6}$$

N_A is the number of contributions of the country A and N is the total number of contributions: $N_A = \sum_i na_i$ and $N = \sum_X N_X$.
To translate this probability into expected collaboration values, we multiply it by the sum of all international collaborations observed. This constant gives the scaling of the system.

For a given country, the Probabilistic Collaborative Index (PCI) is then given by the ratio between the observed and the expected number of collaborations, as shown in eq. (7).

$$PCI(A) = \frac{C_A}{\hat{C}_A} = C_A \left[P_{int}(A) \sum_X C_X \right]^{-1} \quad (7)$$

b. Validation

In order to prove the validity of our calculations, we used *Visual Basic for Applications* to build a simulation of a probabilistic model based on 10,000 articles and 5 countries (A, B, C, D and E) with different contribution sizes. The initial values of the problem are detailed in Table 1. Each contribution was randomly assigned to a specific paper.

After all the contributions were distributed among the 10,000 papers, we used the method above to describe it. First, the number of international collaborations was calculated for each country using eq. (5). Then, the expected value was evaluated using the probability given by eq. (6) and finally, the PCI was calculated using eq (7). The results of this simulations are summed up in Figure 2a and Figure 2b.

Country	Contributions	Appears in	
A	5000	3991	papers
B	4000	3285	Papers
C	3000	2546	papers
D	2000	1816	papers
E	250	246	papers

Table 1: Probabilistic model on 10,000 articles – Initial conditions

As can be seen in Figure 2a, the calculated expected value fits properly the observed data. As a result, the PCI for each country oscillates very little around 1. Since the expected value corresponds to the outcome of the experience for an infinitely large number of papers, the higher this number is, the smaller the oscillations.

On the other hand, the arguments in favor of a fit using a quadratic curve will be detailed in section I-c.

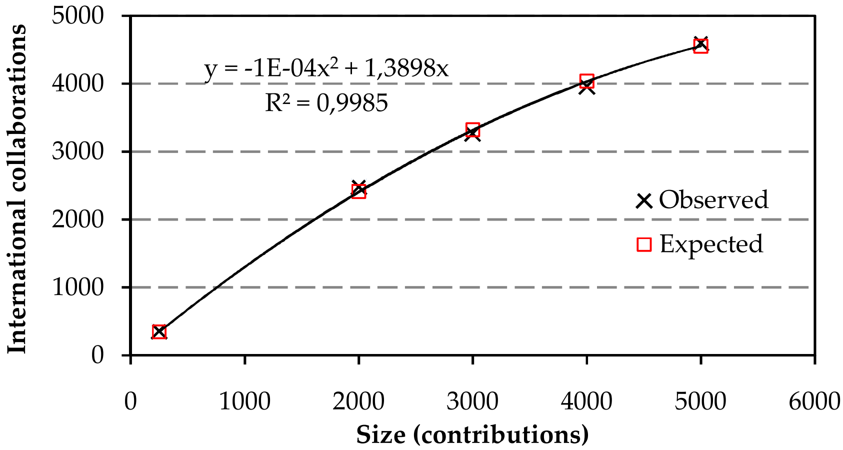


Figure 2a: Probabilistic model on 10,000 articles - Results

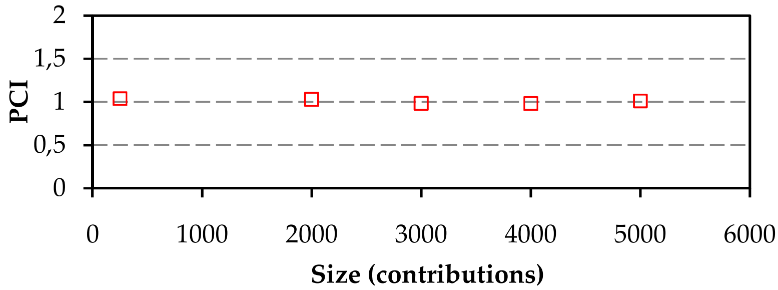


Figure 2b: Probabilistic model on 10,000 articles - PCI

The PCI will be used to evaluate the collaborative efforts of the different subgroups (here countries). PCI values smaller than 1 indicate that the researches in these countries do not cooperate as much on the international level as they should or could.

The next step in the validation of this model is to apply it to real-world problems. In order to do so, we took the international collaboration system already presented in Figure 1, page 14. When fitting the observed values with the predictions of the probabilistic model, both G and χ^2 statistical tests are more favorable than when fitting with a power law on the log-transformed data (Table 2).

Model	G	χ^2
Probabilistic	**31664**	**31878**
Power law	**1388838**	**41340**

Table 2: Performance of the probabilistic model, applied to the world system

In order to demonstrate the validity of this approach, we selected all the articles referenced in Scopus about Superconductors. The expected values were calculated using the probabilistic method and the observed ones were counted using an SQL algorithm. Both values are represented in Figure 3a.

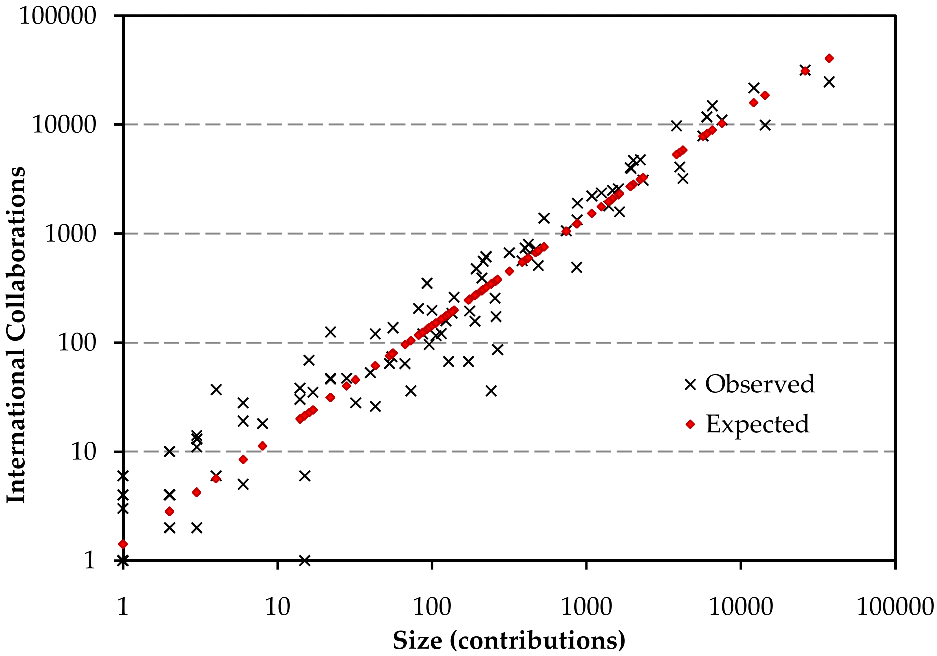


Figure 3a: Probabilistic description - Superconductors

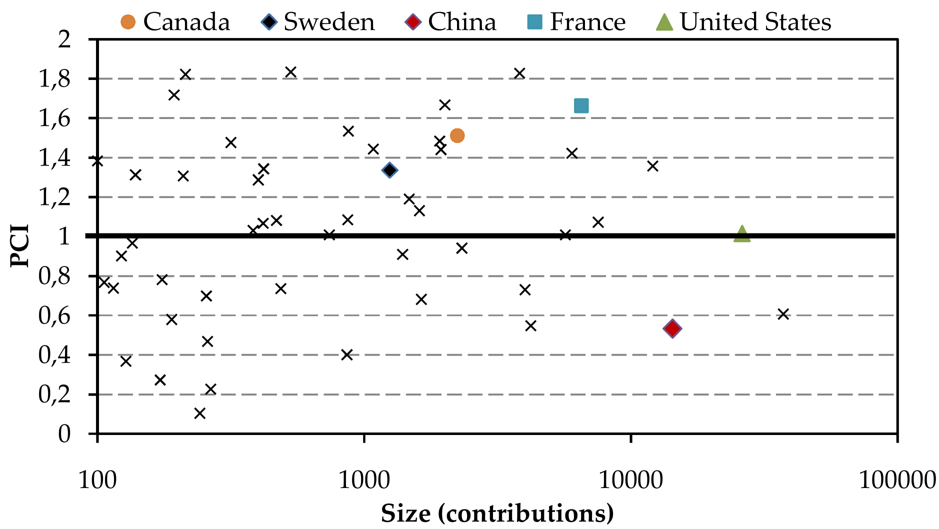


Figure 3b: PCI - Superconductors

c. Findings

It was shown in section I-1 that several bibliometric phenomena, such as international collaboration, could be described using power laws. These curves describe a very specific behavior, often associated with fractals, carrying specific structures and scale properties (Schroeder, 1991). It is rather unexpected that such a structured system could be governed by random processes.

In order to understand the structure of the system we built here and its prediction capabilities, let us introduce the following notations:

$$S_1 = \sum_X C_X$$

$$S_2 = \sum_X \binom{N}{2} - \binom{N-N_X}{2} - \binom{N_X}{2}$$

With these notations, the expected value of international collaboration for a country A ($\hat{C}_A$) can be expressed using equation (6). The resulting expression shows that our system is governed by second order polynomials (eq. (8)). This new essential feature has important consequences on the predictions our model can make.

$$\hat{C}_A = \frac{S_1}{S_2} N\, N_A - \frac{S_1}{S_2} N_A^2 \qquad (8)$$

Since S_1, S_2 and N are all positive constants, we expect $\hat{C}_A$ to decrease when N_A gets very large. To our knowledge, no observation of this feature has been reported in the literature.
Contrary to a power law, the representation of a second order polynomial on a log-log plot is not a straight line; it is curved at high values. Figure 4 illustrates this difference and shows why the confusion between both models can be hard to detect. In most classic cases, the system doesn't reach the region of the curvature. In these cases we seem to be fitting a second order polynomial with a power law.

By seeking for this specific prediction of our model, we look for the invalidation of either our probabilistic parabolic model or the power law model.

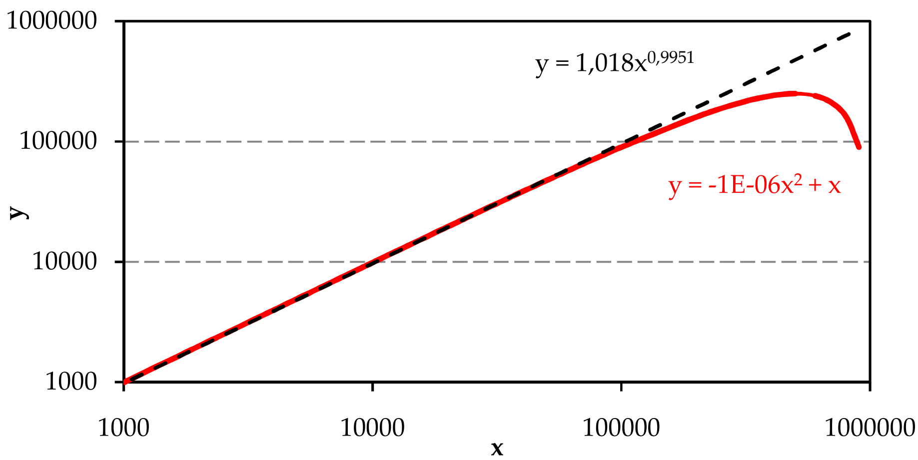


Figure 4: Divergence of power and quadratic curves

Such a situation is encountered when a certain country if far more productive on a specific topic than the others. In this configuration, we enter the zone where power laws and second order polynomials diverge significantly.

We found several of these unbalanced systems. We chose here to present a study of the literature on Lake Michigan in the United States. As expected, the majority of contributions (92%) belong to the United States (Table 3). As can be seen in Table 4, our prediction using equation (8) gives a much better fit than the power-laws do[3]. Moreover, the deviation between our predicted values and the optimal quadratic fit given in Figure 5 is very small. We calculated the coefficients of the second order polynomial given by equation (8) and found a 3.49% error from the optimal fit given on Figure 5. This highlights two major conclusions of this work: first, with the same degree of freedom (i.e. two coefficients to be adjusted), the fit with a second order polynomial gives a better result than the fit with a power law (Table 4). Then, this quadratic fit is very well described by our model and the resulting equation (8).

To us, power laws only describe the first portion of collaboration systems. This description using second order polynomials built on a probabilistic model not only coincide with the power law description on the first section but gives a more complete description of the system.

[3] Fitting with a power law can be done either by fitting the raw data directly or by fitting a line after a log-transformation. These methods give different results and discussions are still going on to determine the proper method which should be used.

Country	Number of contributions	Share of total
U.S.A	**3235**	**92.0%**
Canada	**161**	**4.6%**
France	**11**	**0.3%**
United Kingdom	**11**	**0.3%**
Germany	**10**	**0.3%**
Japan	**10**	**0.3%**
Rest of the world	**79**	**2.2%**

Table 3: Scientific production by country on Lake Michigan

Model	R^2	G	χ^2
Probabilistic	**0.993**	**196**	**65 200**
Optimal quadratic fit	0.995	258	66 941
Power law fit	0.887	100	5 229 384
Power-log fit	0.697	242	116 630 894

Table 4: Comparison of the different models on the Lake Michigan case

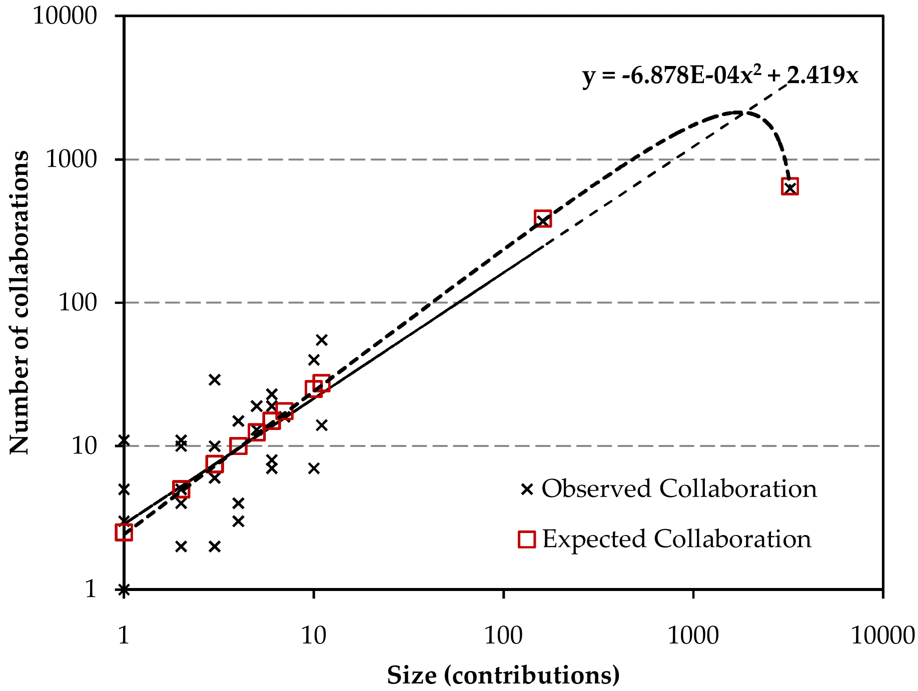


Figure 5: Results on the Lake Michigan system

d. Limitations

A second objective was to try to apply the same approach to predict the national collaboration effort, thereby having a complete description of countries' collaboration.

As can be seen in Figure 6, the distribution of national contributions with respect to the size of the countries seems to follow a similar power law. However, after trying to adapt our calculation to describe national collaboration, we couldn't fit the data with any of the available derivation of our probabilistic model. Moreover, if it was possible to describe the national collaboration system with a power law or a quadratic law, we would have a full description of the overall system. There would therefore be a natural relationship between the number of national collaborations C_{nat} and the total number of collaborations (C_{tot}):

$$f(N_A) = \frac{C_{nat}(N_A)}{C_{tot}(N_A)} = \frac{C_{nat}(N_A)}{C_{inter}(N_A) + C_{nat}(N_A)}$$

where N_A is the total number of contributions for country A and C_{inter} the number of international collaborations calculated with the probabilistic model. f would then be a function of N_A defined as a ratio of two power-laws or second order polynomials.

However, it was found that no general trend could be identified for the description of this function f. Indeed, as can be seen in Figure 7, the repartition of national collaboration among the international one seems totally random. We therefore believe no general description can be given of this particular system.

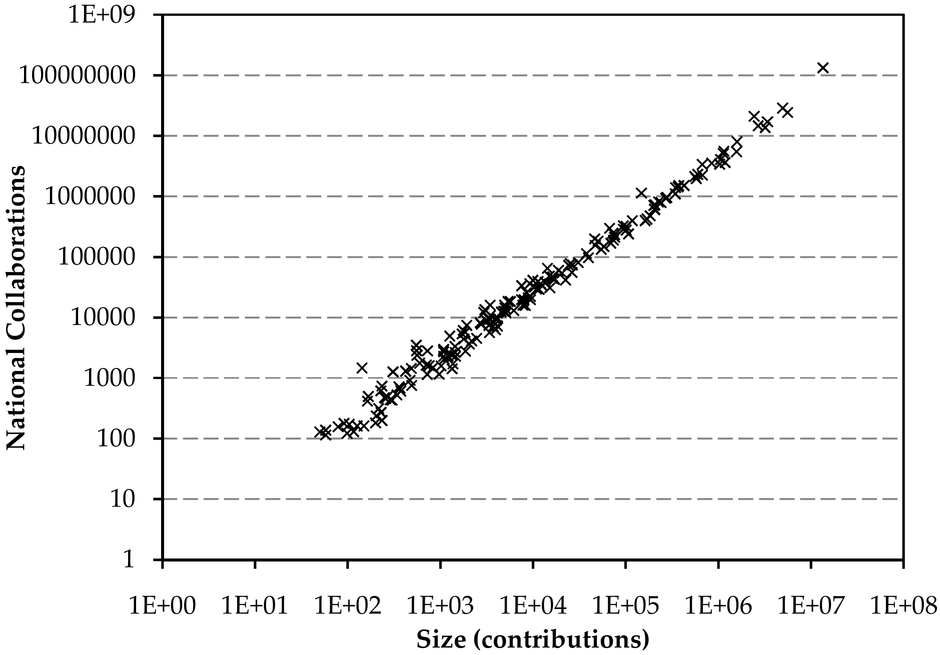


Figure 6: The national collaboration system

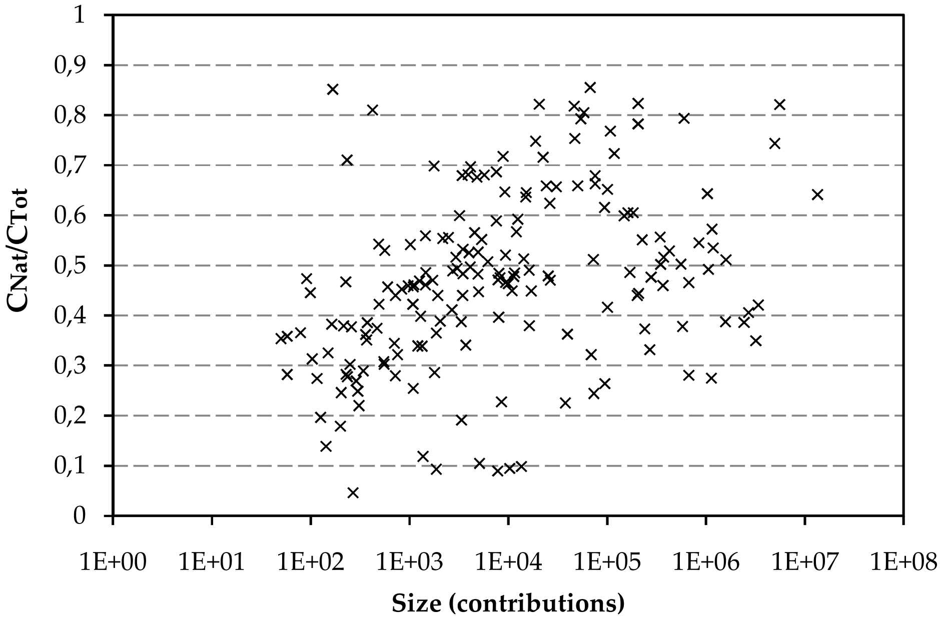


Figure 7: Share of national collaborations

3. *Spline Interpolation*

a. Interpolation

In numerical analysis, interpolation consists in constructing new data points from a given discrete set of known data points. Two main approaches can be considered when trying to interpolate a set of data points. The first one consists in fitting a known function to the data by determining its parameters. This requires having a model or a theory describing the behavior of the observed phenomenon.
The second consists in building a function which would fit all the data points. The value of that function between the data point is an interpolation.

In our case, since we are looking for a method which would apply to a variety of observations, we considered the second option. However, there again, several techniques exist to build the interpolated curve. The efficiency of each of these methods depends on the global shape of the data: Lagrange polynomials are used for sets of data in the Lagrange form, linear interpolation for dense data sets, Whittaker–Shannon interpolation for continuous-time band-limited signals, nearest-neighbor interpolation for slowly varying phenomena.

We decided to implement another method, called spline-interpolation, for three main reasons:

- The method can deal with non-equally spaced data sets.
- Its implementation can be easily programmed.
- It avoids the problem of Runge's phenomenon[4].

By definition, a spline is a piecewise-defined function using polynomials. It was shown that without a relevant damping, quadratic spline interpolations were susceptible to severe oscillation effects in the case of quickly varying data. Cubic polynomials are a good compromise between stability, differentiability, damping of oscillations and computational cost (Ueberhuber, 1995).

This work on the application of the interpolation theory using cubic splines is largely inspired by the work published in the book of Green & Silverman, *Nonparametric Regression and Generalized Linear Models*, published in 1994.

[4] When interpolating a data set with high degree polynomials for better precision, Runge's phenomenon is a problem of oscillation that occurs at the edges of the data set.

- Implementation:

We have used here a quadratic spline interpolation. Given n + 1 data points (x_0, y_0) to (x_n, y_n) we defined n spline functions S_n by the following conditions:

- Interpolation:

$$S_i(x_i) = y_i \text{ for } 1 \leq i \leq n-1 \text{ and } S_{n-1}(x_n) = y_n \tag{9}$$

- Continuity:

$$\text{for } 1 \leq i \leq n-1 : \begin{cases} S_{i-1}(x_i) = S_i(x_i) \\ S'_{i-1}(x_i) = S'_i(x_i) \\ S''_{i-1}(x_i) = S''_i(x_i) \end{cases} \tag{10}$$

Each S_i is a cubic polynomial defined as $S_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i$. There are therefore 4n unknowns to be calculated using 4n-2 equations. As described in the literature, we introduced two additional conditions and defined **natural cubic splines** (Green & Silverman, 1994):

$$S''(x_0) = S''(x_n) = 0 \tag{11}$$

- Computation

As explained before, we need to solve 4n equations to determine all the constants defining our piecewise-defined function. Before starting the resolution, we used the following notations:

Since each S_i is a cubic polynomial, we can write its second derivative as $S''_i(x) = z_i + m_i(x - x_i)$ where $m_i = \frac{z_{i+1} - z_i}{h_i}$ is the slope of $S''_i(x)$.
By integrating this expression twice we get:

$$S_i(x) = \frac{z_{i+1}}{6\,h_i}(x - x_i)^3 + \frac{z_i}{6\,h_i}(x_{i+1} - x)^3 + C_i(x - x_i) + D_i(x_{i+1} - x) \tag{12}$$

Where C_i and D_i are integration constants which are calculated from the conditions $S_i(x_i) = y_i$ and $S_i(x_{i+1}) = y_{i+1}$.

The last step is to calculate the Z_i using the continuity condition $S''_{i-1}(x_i) = S''_i(x_i)$.

This gives us the following expression:

$$h_{i-1}\, z_{i-1} + 2\,(h_{i-1} + h_i)\, z_i + h_i\, z_{i+1} = 6\left(\frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}}\right) \tag{13}$$

In order to determine the spline's coefficients and to be able to plot it, eq. (13) has to be solved. This can be achieved by transforming the expression into a matrix equation. Eq. (14) can be therefore used instead of the set of equations defined by the expression (13) to efficiently calculate the vector containing all the z_i coefficients.

$$\begin{bmatrix} 2(h_0+h_1) & h_1 & 0 & 0 & \cdots & 0 \\ h_1 & 2(h_1+h_2) & h_2 & 0 & \cdots & 0 \\ 0 & h_2 & 2(h_2+h_3) & h_3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & h_{n-3} & 2(h_{n-3}+h_{n-2}) & h_{n-2} \\ 0 & 0 & 0 & 0 & h_{n-2} & 2(h_{n-2}+h_{n-1}) \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{n-2} \\ z_{n-1} \end{bmatrix} = \begin{bmatrix} 6(b_1-b_0) \\ 6(b_2-b_1) \\ 6(b_3-b_2) \\ \vdots \\ 6(b_{n-2}-b_{n-3}) \\ 6(b_{n-1}-b_{n-2}) \end{bmatrix} \quad (14)$$

Let us write eq. (14) $M \times Z = B$.
By noticing that the matrix M is strictly diagonally dominant by column, we can use the Levy–Desplanques theorem to say that it is non-singular (Taussky, 1949). Therefore, we know that M^{-1} exists and the coefficients can be calculated using $Z = M^{-1} \times B$.
Moreover, the fact that M is tridiagonal and similar to a Hermitian matrix reduces the computational cost of the inversion.

b. Smoothing

When considering a dense set of data, it might not be optimal to fit perfectly the data. Indeed, error fluctuations in the data may have an important impact on the general shape of our fitting curve. However, since we are looking at continuous, slowly varying phenomena, these fluctuations can be damped using a linear smoothing. This method will associate each data point with its damped image. The damping is evaluated by comparing the position of each data point with that of its neighbors. We introduced a smoothing parameter which quantifies the correlation between the data points.

In this work we used a smoothing parameter of 0.5, which gives the best fit for the data sets we had. Note that even if doesn't appear in this work, an analytical approach exists to determine the optimal smoothing parameter for a dataset.

- <u>Implementation:</u>

Let $Y = (y_1, \dots, y_n)$ be the initial data set and $g = (g_1, \dots, g_n)$ the output (smoothed) data set. The key idea behind smoothing a curve is to find a compromise between two key parameters in curve fitting:

- The sum of squares which quantize the distance between the data points and the fitting curve
- The roughness of the function

The roughness is evaluated using an intuitive way: when a flexible piece of structure (i.e. a spline) is bent to the shape of a known curve, then the strain energy is proportional to the integral of the second derivative's square of the function. This approach has become a standard way of evaluating the roughness since Reinsch's work on splines in 1967 (Eilers & Marx, 1996) (Reinsch, 1967). The roughness r is therefore evaluated using equation (15) below.

$$r = \int_a^b \left(g''(t)\right)^2 dt \tag{15}$$

We are therefore able to introduce the penalized sum of squares Pss given in eq (16) where Y is the vector defined by $Y = (y_1, \dots, y_n)$, g is any twice differential function on [a,b] and α is the smoothing parameter. Note that $\{g_i\}$ is defined by $\forall i \in [1;n]\ g_i = g(t_i)$

$$Pss(g) = \sum_{i=1}^{n} \{Y_i - g(x)\}^2 + \alpha \int_a^b \left(g''(t)\right)^2 dt \tag{16}$$

Let γ be the vector defined by $\gamma = (\gamma_2, \dots, \gamma_{n-1})$ where $\forall i \in [2;n-1]\ \gamma_i = g''(x_i)$. Knowing both vectors Y and γ gives a full description of the curve of g.

$$Q^t\, g = R\, \gamma \tag{17}$$

Where:

$$Q = \begin{pmatrix} h_1^{-1} & 0 & 0 & \cdots & 0 \\ -h_1^{-1} - h_2^{-1} & h_2^{-1} & 0 & \cdots & 0 \\ h_2^{-1} & h_2^{-1} - h_3^{-1} & h_3^{-1} & \cdots & 0 \\ 0 & h_3^{-1} & h_3^{-1} - h_4^{-1} & \cdots & 0 \\ 0 & 0 & h_4^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & h_{n-1}^{-1} \end{pmatrix}_{n\times(n-2)} \tag{18}$$

$$R = \begin{pmatrix} \frac{1}{3}(h_1 + h_2) & \frac{1}{6}h_2 & \cdots & 0 \\ \frac{1}{6}h_2 & \frac{1}{3}(h_2 + h_3) & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \frac{1}{3}(h_{n-2} + h_{n-1}) \end{pmatrix}_{(n-2)\times(n-2)} \tag{19}$$

where $\forall i \in [1;n]\ h_i = x_{i+1} - x_i$

The roughness can then be written $r = \int_a^b \left(g''(t)\right)^2 dt = \gamma^t R\, \gamma = g^t\, K\, g$ where

$$K = Q\, R^{-1} Q^t \tag{20}$$

Since we aim at minimizing the penalized sum of squares, let's introduce the penalized least square estimator:

$$\hat{g} = \arg\min Pss(g) \tag{21}$$

To evaluate $\hat{g}$ we develop the expression of $Pss(g)$ into:

$$\begin{aligned} Pss(g) &= \sum_{i=1}^{n} \{Y_i - g_i\}^2 + \alpha\, g^t K\, g \\ &= (Y-g)^t\,(Y-g) + \alpha\, g^t K\, g \\ &= g^t(I + \alpha K\,)g - 2\,Y^t g + Y^t Y \end{aligned}$$

The penalized sum of squared reaches a minimum when $g = (I + \alpha K)^{-1}\,Y$; therefore

$$\hat{g} = (I + \alpha K)^{-1}\,Y \tag{22}$$

We finally calculate the output data set by the following linear transformation:

$$g = (I + \alpha Q R^{-1} Q^t)^{-1}\,Y \tag{23}$$

c. Results

In order to illustrate the results obtained with the described method, we created an arbitrary dataset, visible on the two figures below.
The red line in Figure 8a represents the spline created with the method described earlier, where each section between two data points is a third-degree polynomial. We can see that the result fulfills the initial criteria: the spline passes through all the points without any discontinuity or abrupt change in derivative. This is a validation of the algorithm used to solve equation (13).

The implementation of the smoothing algorithm is visible on Figure 8b, where a smoothing parameter $\alpha = 2$ was used. Here again, the initial dataset is represented by black crosses. However, the spline here was built on the smoothed dataset calculated with equation (23) and represented by red squares.
As expected, the smoothing gives the curve a more regular shape. It suppresses small fluctuation to focus on the global trend. This feature will be crucial in the final application of this tool, which will be developed in chapter II.

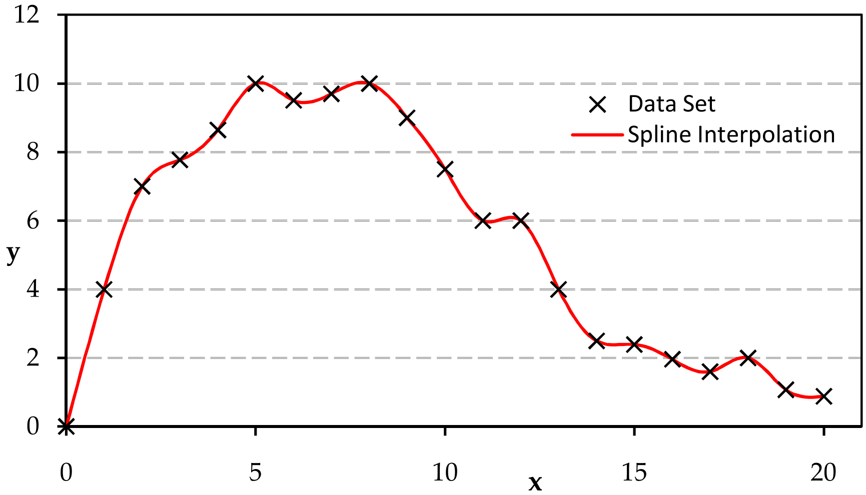


Figure 8a: Interpolation using a cubic spline

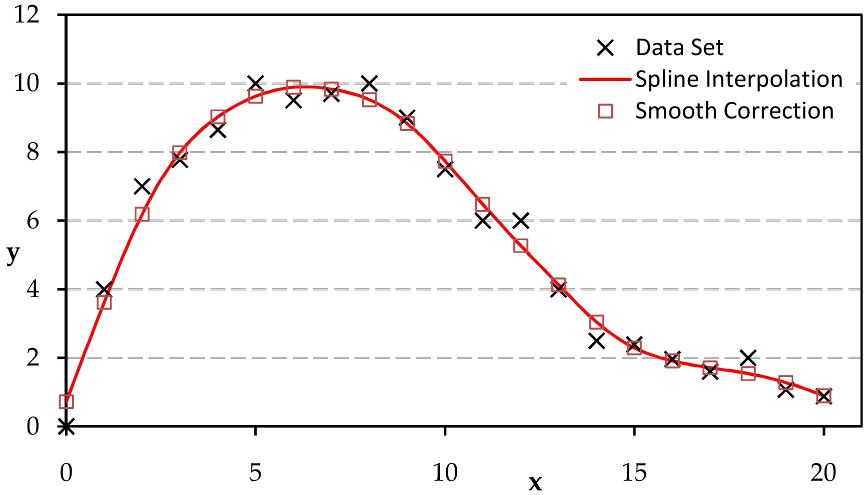


Figure 8b: Smoothing of the cubic spline, $\boldsymbol{\alpha} = \mathbf{2}$

4. *Moving averages*

When complex manipulations are performed on the raw data, scientists often face a problem when it comes to the fitting of the resulting point cloud. This paragraph contains a description of the method we used to fit and analyze the resulting set.
Since power laws are an important theoretical feature in the systems we study, we expect our data to follow a similar trend. If it doesn't, we hope to be able to identify the underlying trend and find a suitable model to describe it.

a. Proof of concept

The algorithm we developed is based on the so-called Mayer method, also known as a moving average. Usually, this technique is used to smooth out short term fluctuations. It is here used to give a single y-value to an x-range. This suppresses the artificial weighting coming from the abundance of information for small values of x. A typical example of such a data set is provided in Figure 9. It was constructed by generating random integers Y_{th} according to the criteria $0.64\sqrt{X_{th} - X_l} < Y_{th} < 0.35\sqrt{X_{th}} + Y_l$ where X_l and Y_l were defined arbitrarily to resemble de type of data we usually work with. These two upper and lower bounds were calculated in order for the underlying model to be described by the equation $Y_{th} = 0.5\sqrt{X_{th}}$.

As can be seen in the figure, X_l and Y_l were respectively chosen to be 400 and 5. The thick dark line represents the trend line obtained when fitting the initial data set and the red line represent the fit on the Mayer transformed data, using a least square fit to a power law.

It is clear from Figure 9 that using a moving average gives a much better fit. It was calculated by using arbitrary intervals of 10 X units. For example, we calculate the average of all the points which abscise is bigger than 50 and smaller than 60 and associate it with an abscise of 54.5.

The validity of the arbitrary interval (here 10) can be discussed. However, this example being only a proof of concept for the method, this will be done in the next section, when the moving average will be calculated on a real set of data.

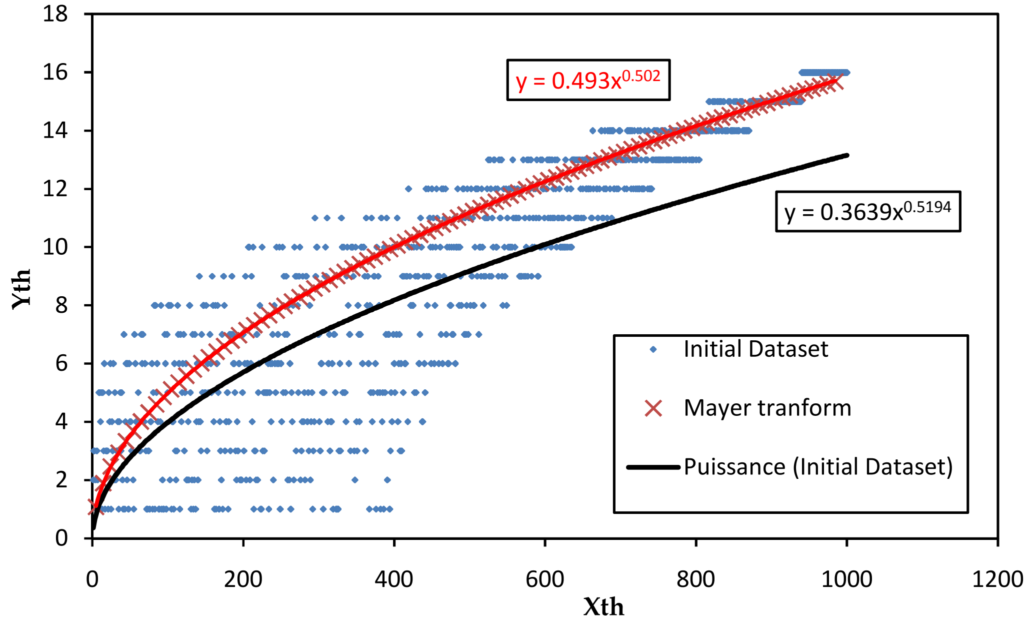


Figure 9: Fitting a point cloud using moving averages

b. Real-world example

Once the details of this method were understood, we were able to apply it to real world problems. A current bibliometric study on the design of a new performance indicator gave as a result the point cloud presented in Figure 10a. Since it is hardly impossible to extract any trend or information from the figure, a transformation of the dataset will be necessary.

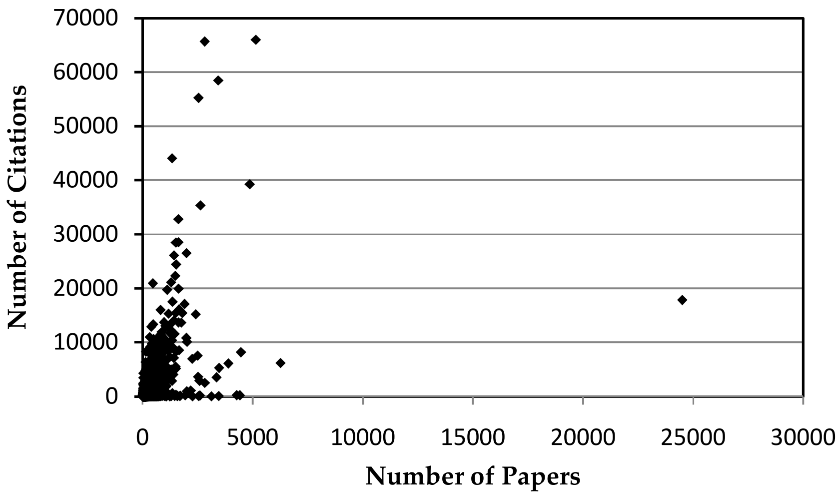


Figure 10a: Citations Vs Size – Linear scale

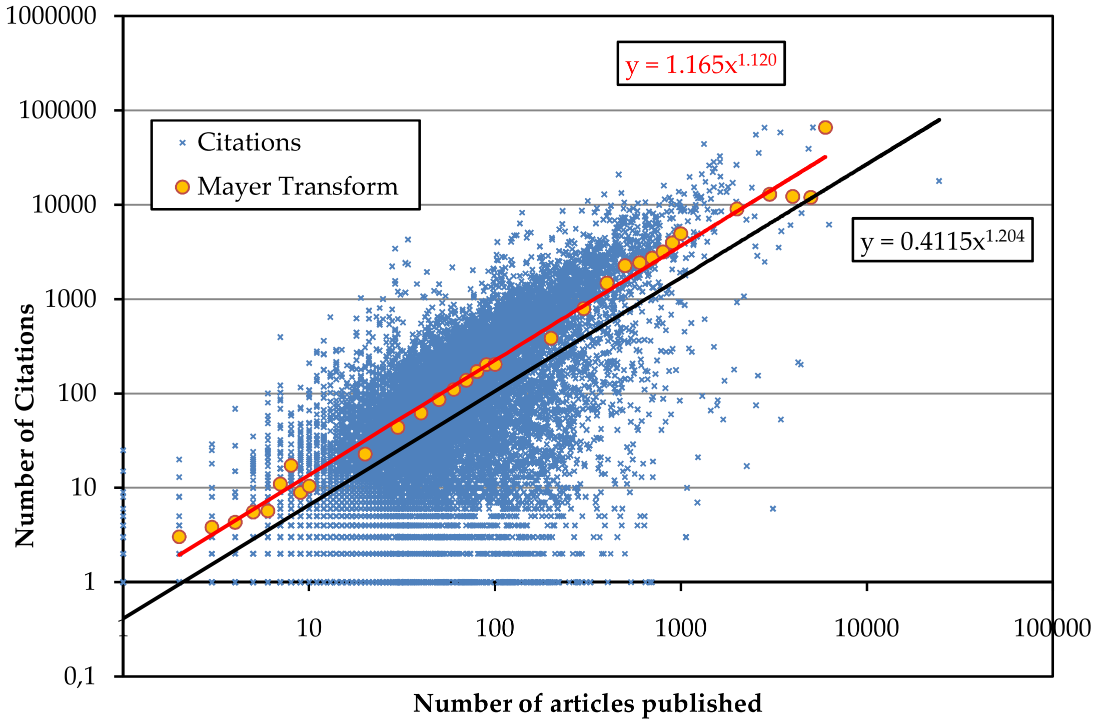


Figure 10b: Citation Vs Size – Logarithmic scale

To start with, since citations were shown to experience a Matthew effect (Katz J. S., 2000), we expect the best fit to be obtained with a power law. As can be seen in Figure 10b, using this conjecture to transform both axes into a logarithmic scale gives a much clearer point cloud. Even if the data look messier than the artificial cloud built in the last section, the same method will allow us to determine the underlying trend in the system.

The first issue to be solved when using Mayer's method is to determine the width of the range over which the average will be calculated. As opposed to what was done in the previous theoretical example, we are looking for a method which could be applied to any real world situation. The main disadvantage of the "fix range" approach is that its efficiency highly depends on the dataset. For example, using multiples of 10 as was done in Figure 9 would be irrelevant if the x-values range from 0.001 to 10. Also, if the x-values are defined over a long range, as in Figure 10a, it is irrelevant to calculate the average on a 10-articles window since it might not contain any data.

Because we want to average the studied behavior without using an a priori information on the scaling of the system, we decided to calculate Mayer's barycenters based on the logarithmic scale used. For instance, the width of the window would be 0.1 on the values ranging from 0.1 to 0.9, 1 on the values ranging from 1 to 9, 100 on the values ranging from 100 to 900, etc.

The result of Mayer's transformation on a log-log scale using this type of window is represented on Figure 10b. The red line corresponds to the fitting of this new dataset with a power law using a least square fit, while the black line corresponds to a fit on the initial dataset. The fit using the moving average is more satisfactory since the red line passes closer to the corner of the cone, which is where the data is the most accurate.

We do not aim here at presenting the details of the work on journal citations, which will be the main subject of an article soon to be published. This example aimed at exposing a method we developed in order to improve the treatment of dense statistical data sets in bibliometrics, based on Mayer's transformation.

II. Application to batteries and fuel cells

1. *Institutional Collaboration*

The objective in this section will be to assess who the main actors of the battery and fuel cell fields are in term of number of publications and collaborative efforts. Measuring the collaboration between the main institutions will allow us to determine the pivot actors which are the link between several research groups and identify the key research hubs. We will also focus on the relative position of European, North American and Asian institutions.
The main tools which will be used here are related to network analysis. We constructed several networks using Ucinet 6 (Borgatti, Everett, & Freeman, 2002) and the data extracted from Scopus: for each article in the Fuel-cell or the Battery envelope[5], we selected the institutions which the authors were affiliated to. From this information we built a so-called "collaboration matrix" where the number of collaboration is given for each couple of institutions.

The next sections will describe both quantitatively and qualitatively the global picture in fuel cell and battery related technologies and highlight situations where organizations could gain from a deeper collaboration with neighboring institutions.

a. Collaboration efforts

In order to describe the global collaboration system in the fuel-cell and battery environment we will start by implementing the probabilistic method detailed in section I-2. The description of the system we obtain with this approach is visualized in Figure 13 and Figure 14 where the red line represents the expected collaborative value. In this particular case, the expected collaboration value calculated from equation (8) is given by eq. (24) for fuel cells and eq. (25) for batteries, where N_A is the number of contributions from country A.

$$\hat{C}_A = 1.790\ N_A - 5.506\ 10^{-5} N_A^2 \tag{24}$$

$$\hat{C}_A = 1.721\ N_A - 1.409\ 10^{-4} N_A^2 \tag{25}$$

Table 7 and Table 8 give an overview of some values of the Probabilistic Collaborative Index (PCI) of both sectors which was calculated using the two equations above.

[5] The envelops were built by selecting the peer-reviewed articles published between 1996 and 2009 in English and related to research, technical or production details regarding batteries or fuel cells. Please refer to Annex 1 and 2 for the SQL code and list of key-words.

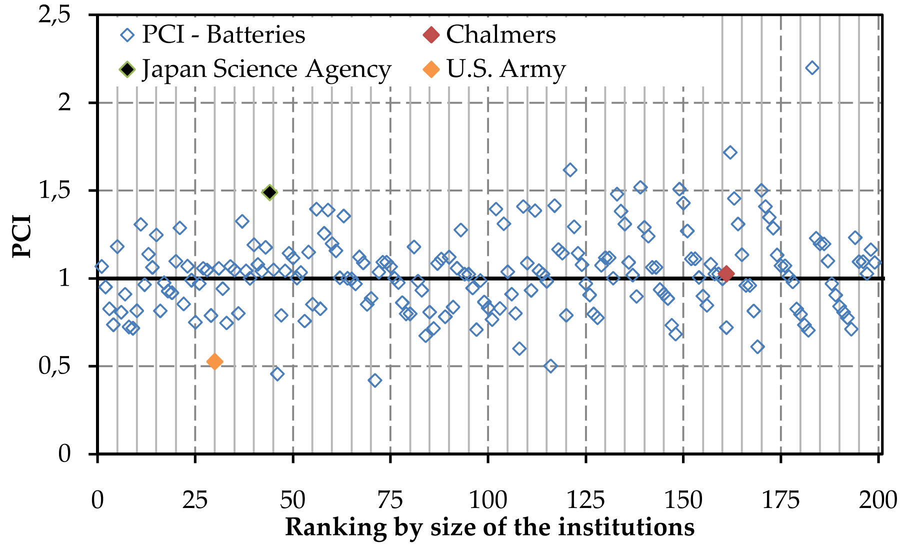


Figure 11: PCI values in the battery system

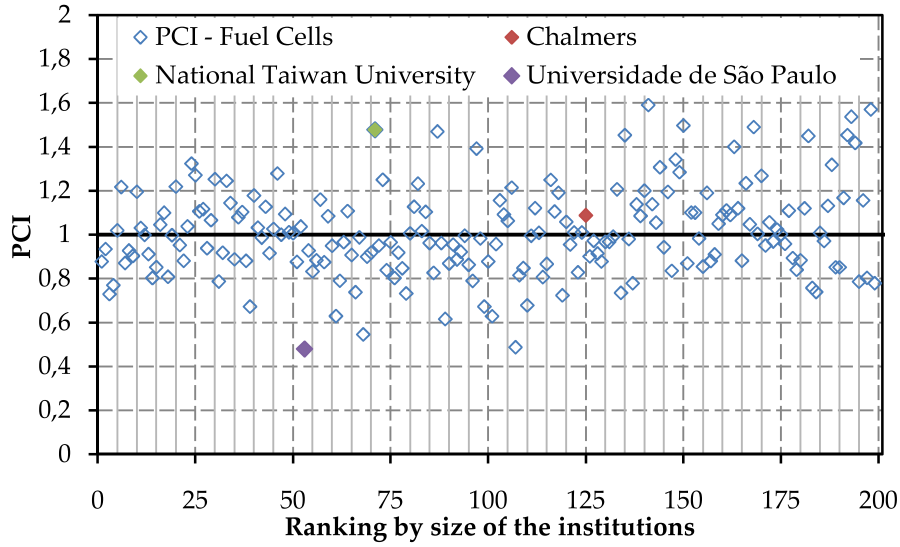


Figure 12: PCI values in the fuel cell system

We can see from Figure 11 and Figure 12 that some major producers of scientific papers have a very low PCI index. This is often explained either by the nature of the institution or by its geographic situation. In the first figure, it is actually no surprise that the U.S. Army, one of the biggest scientific institutions in the world, tends to be quite closed to international collaboration when it comes to research on batteries. On the other hand University of São Paulo (USP) has a very low collaboration index in the fuel cell system. Considering the prestige and role of the institution, this could be explained by its important size when compared to other Brazilian institutions. USP might play the role of a national hub, therefore being associated to a very high number of national projects.

On the other hand, the good score of Japanese science agency and National Taiwan University - two huge contributors – highlights their key position as regional hubs. A further study of this hypothesis will be done in section b.

A few words could also be said on the good position of Chalmers University of Technology which manages to be among the biggest contributors to both domains despite of its relatively small physical size. It is also an actor of the European collaboration on fuel cells and batteries, with a PCI slightly over 1 in both domains.

It is interesting to note that the three main institutions of the fuel cell system (Figure 13) belong to three different geographical groups. This could mean that the United State Department of Energy, the Chinese Academy of Science and Helmholtz-Gemeinschaft are three geographical hubs. On the other hand, no European institution is in the top 3 of the battery system, meaning that no European hub can be easily identified.
These hypotheses will be verified by looking more into detail at the collaboration networks presented in the next section.

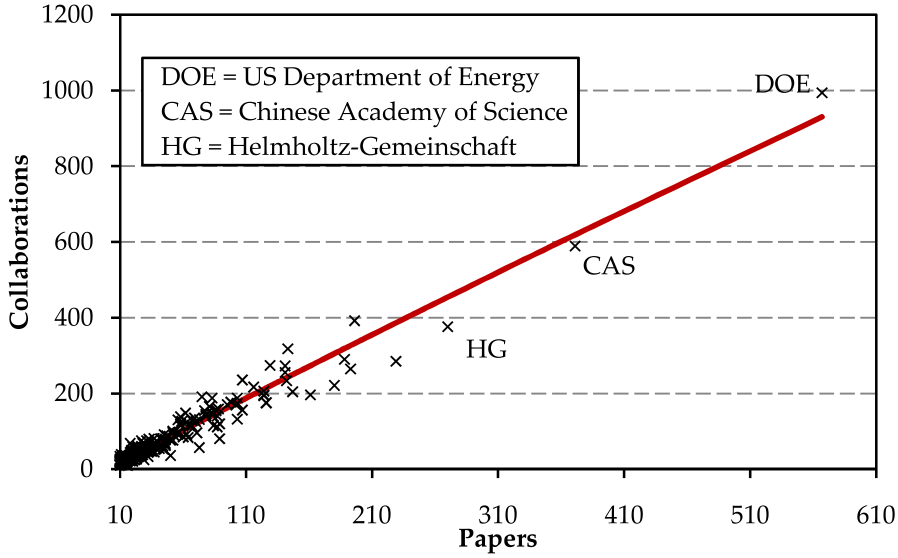


Figure 13: Institutional collaborations – Fuel Cells

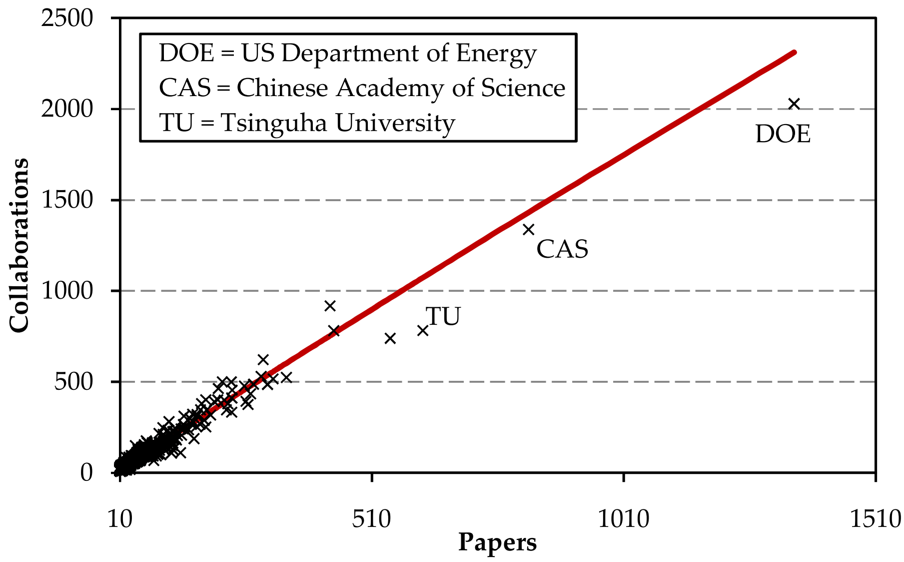


Figure 14: Institutional collaborations – Batteries

b. Collaboration networks

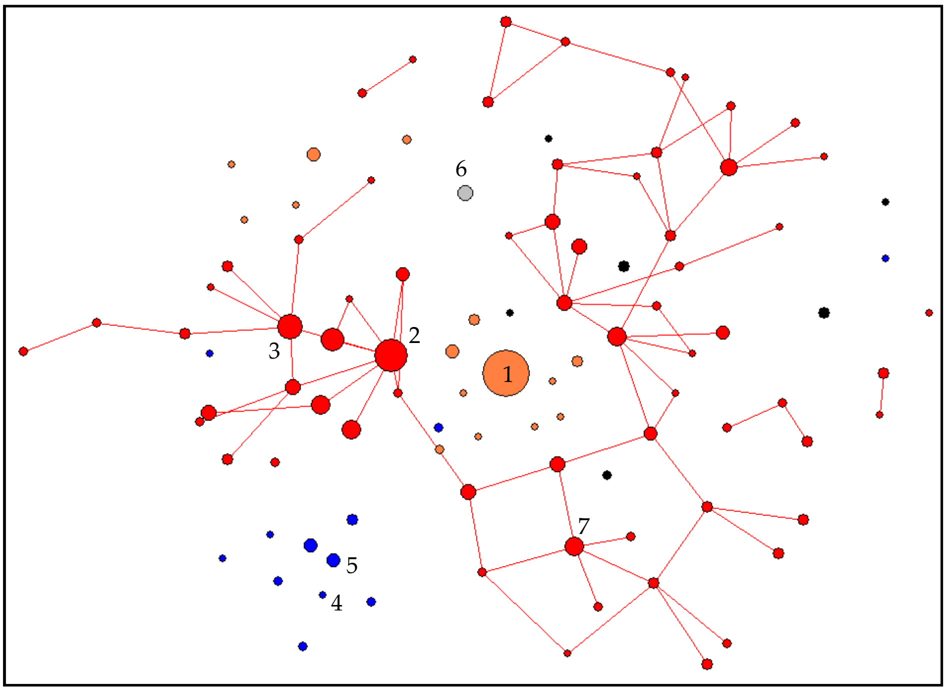
In order to describe the collaboration networks related to both systems we represented all institutions with more than 20 publications and the links where more than 5 bilateral collaborations appeared. Since it is hard to draw any conclusions from the total international pictures (Network 4 and Network 8a), we also produced networks where only internal cooperation was visible.

The networks are built using Ucinet 6 (Borgatti, Everett, & Freeman, 2002) and NetDraw by minimizing the distance between collaborators. A threshold of 5 bilateral collaborations was set to represent the links. Under that limit, the institutions were not considered as collaborators. The size of the circle depends on the number of publications associated with each institution. The position of an institution on the network is calculated to minimize the distance with its collaborators. When several solutions are possible, the institution is placed such as the distance with institutions from the same continent is minimal.

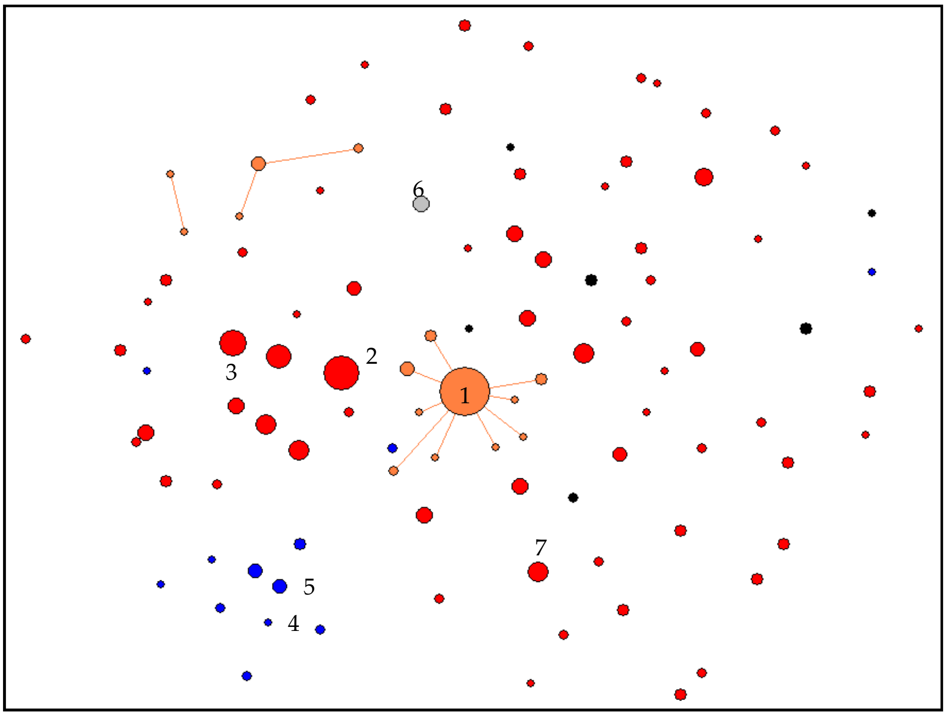
When considering the battery collaboration network, we can see that its entire structure is built with Asian bilateral collaborations. Network 1 shows how Asian institutions and publications represent the major part of the world scientific production on batteries. North America, and more particularly the U.S.A, maintains its key central position thanks to a very strong hub, namely the Department of Energy (DOE), which is a national (Network 2) and international (Network 4) key actor. The Chinese academy of Science plays a central role in the left part of the network (mainly Chinese institutions) but Asia gets its strength from a very dense collaboration network with actors of various sizes. University of Wollongong (UW), Australia, can be easily included in the Asian network regarding its research on batteries (Network 4). This would give the network a ring structure around the American one.

On the other hand, the European network, visible on Network 3, is very marginal. Not only are European publications isolated, but they are also less abundant: no European hub exists on batteries and several small institutions seem to be doing research without building a real European network.
The best strategy from a networking point of view for European groups would be to seek a deeper collaboration with DOE in order to gradually enter the Asia network. The creation of a European hub is a long-term option which would need an unlikely political consensus.

By giving research orientations and funding, institutional hubs allowed a national and regional growth of the field. The recent decision of DOE to allow $2.4 billion on U.S. battery research (Department of Energy, 2009) is precisely the kind of strategy Europe would need.



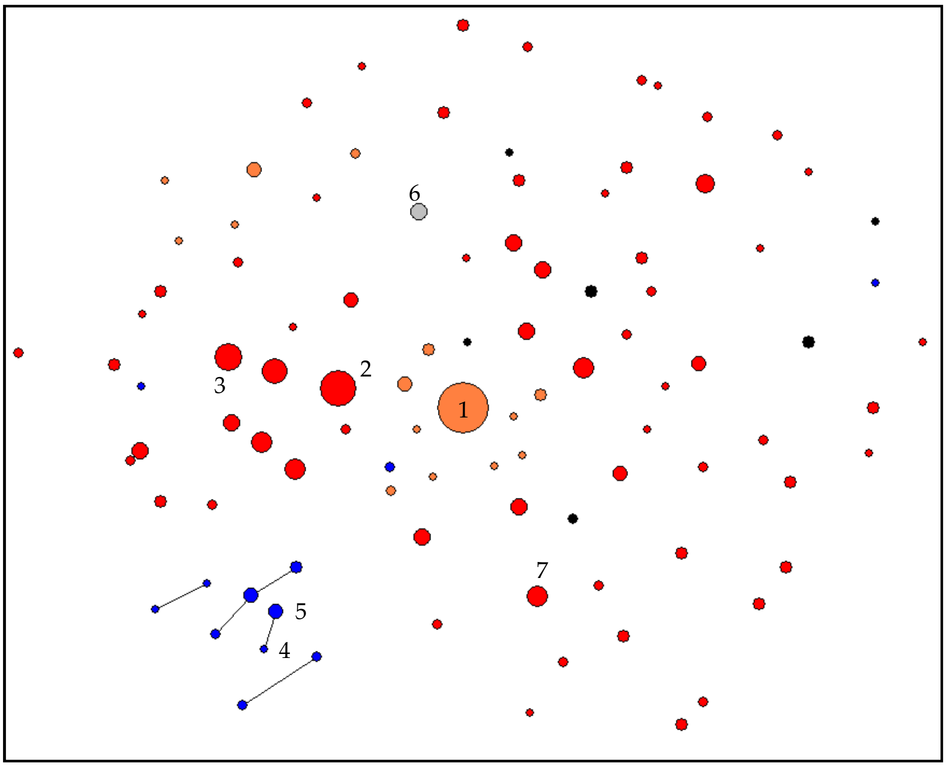
Network 1: Asia in the international battery network



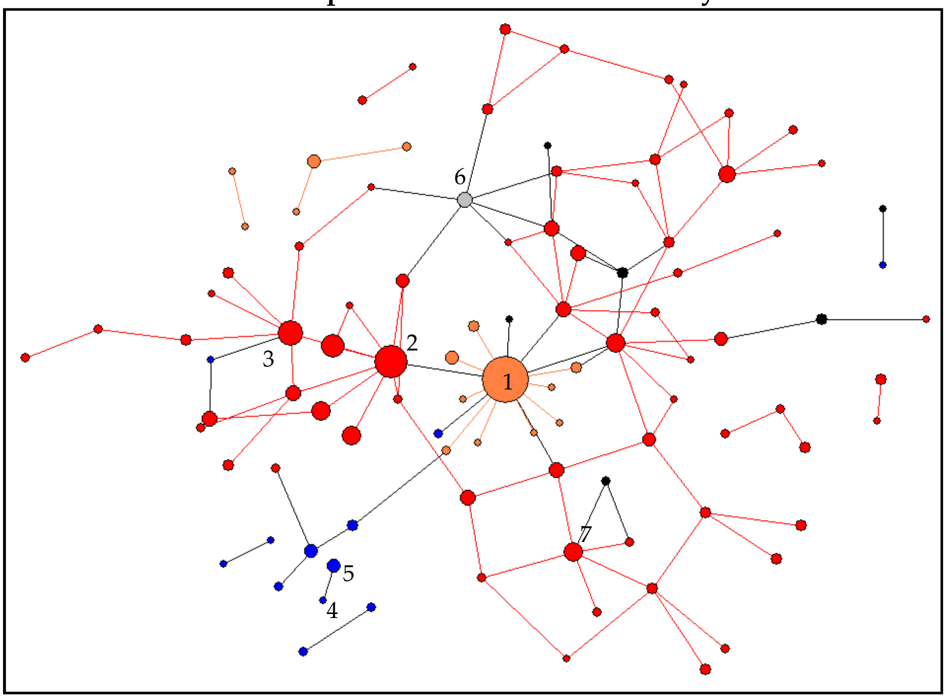
Network 2: North America in the international battery network

Asia North America Europe Private Company

1 : DOE - 2 : CAS - 3 : Tsinghua University - 4: Chalmers Techniska Högskola - 5: Roma la Sapienza - 6: University of Wollongong - 7: AIST



Network 3: Europe in the international battery network



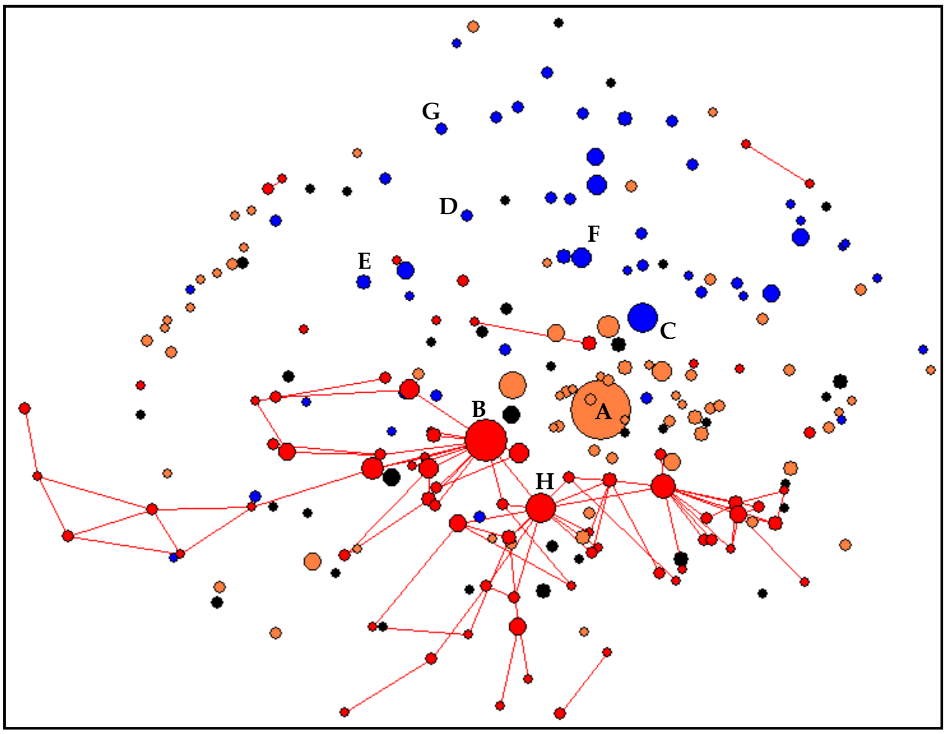
Network 4: The international battery network

Asia North America Europe Private Company

1 : DOE - 2 : CAS - 3 : Tsinghua University - 4: Chalmers Techniska Högskola - 5: Roma la Sapienza - 6: University of Wollongong - 7: AIST

On the other hand, the fuel-cell network is a lot denser and contains more middle-sized actors than the battery one. It also has a completely different structure (Network 8a). In order to analyze the regional characteristics of the network, we proceeded as previously by building three partial networks.

The Asian network is represented in Network 5. Unlike in the battery network, the structure of international cooperation between institutions is not shaped by Asian activities. Also, Europe seems to have much more impact on fuel-cell related technologies, thanks to a relatively wide network and the Helmholtz-Gemeinschaft institute, which plays the role of a regional hub (Network 6). North America plays here again a central thanks to the U.S. Department of Energy (DOE) (Network 7).

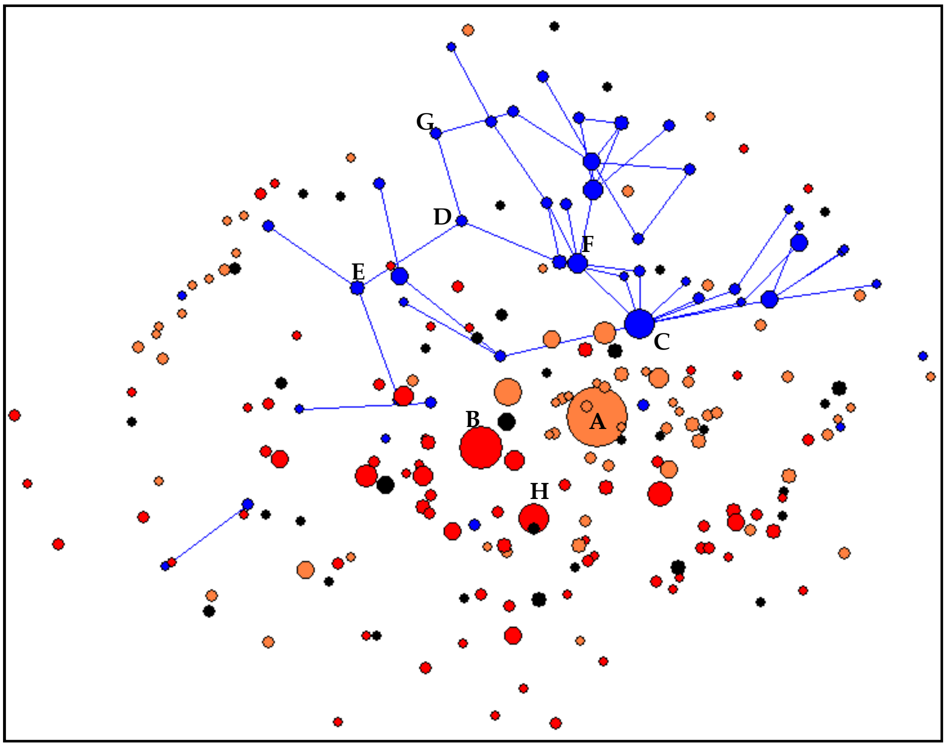


Network 5: Asia in the international fuel cell network

Asia North America Europe Private Company

1 : DOE - 2 : CAS - 3 : Tsinghua University - 4: Chalmers Techniska Högskola - 5: Roma la Sapienza - 6: University of Wollongong - 7: AIST

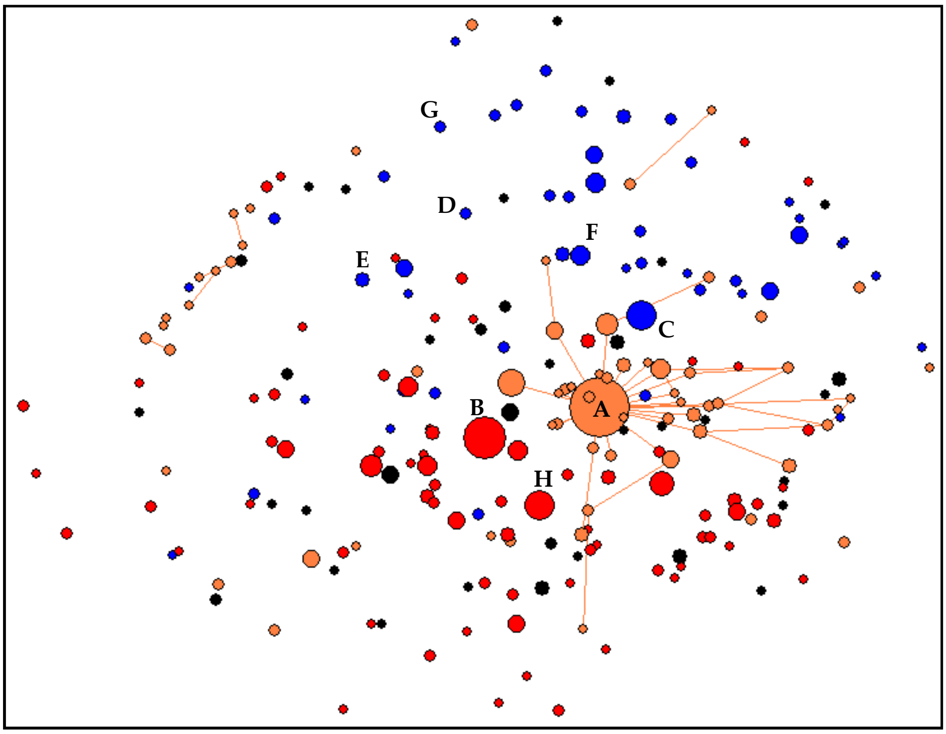
In order to study more in detail the role of North America and the positioning of Europe, we built a network where only the strongest links (more than 15 shared collaborations) were represented. Network 8b not only shows that the U.S is a major central actor of the fuel-cell research but also that the role of Europe is weaker than it previously seemed. Indeed, national hubs such as the French Conseil National de la Recherche Scientifique (CNRS) and the German Helmholtz-Gemeinschaft appear to have very week bilateral interactions. This leads Europe to be an important but outlying actor, belonging to three distinct branches.
On the other hand, Asian institutions are aggregated around the Chinese Academy of Sciences and DOE, building a strong and central collaboration network.



Network 6: Europe in the international fuel cell network

Asia North America Europe Private Company

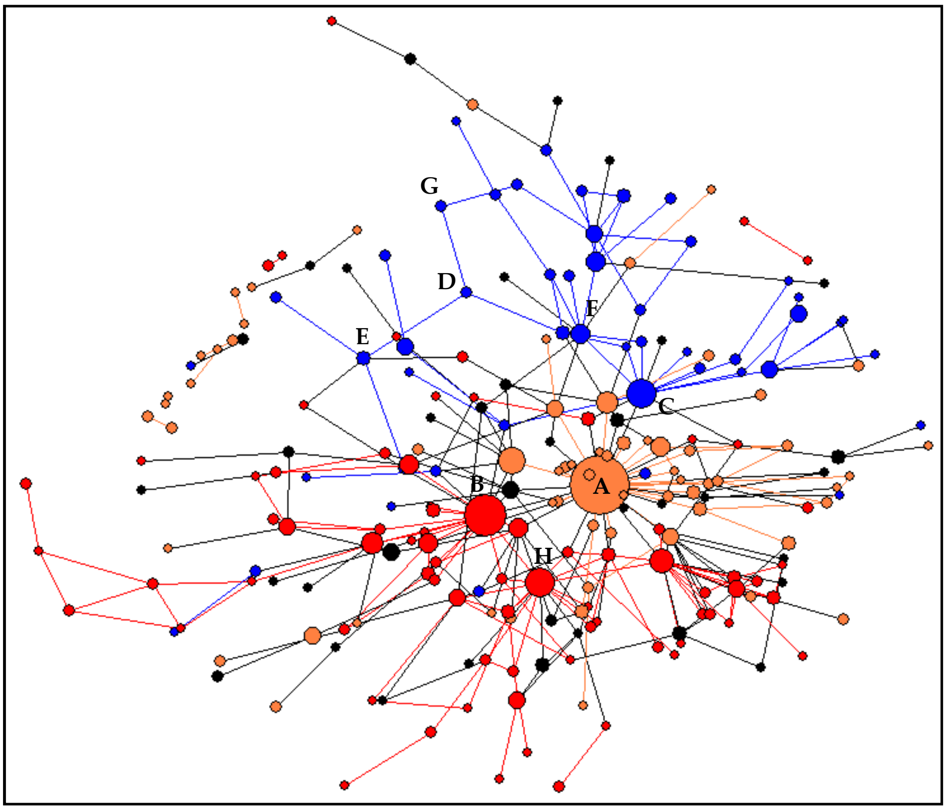
A : DOE - B : CAS - C : Helmholtz-Gemeinschaft e.V. - D: Chalmers Techniska Högskola - E: Russian Academy of Science - F: KTH - G: Roma la Sapienza - H:AIST



Network 7: North America in the international fuel cell network

Asia North America Europe Private Company

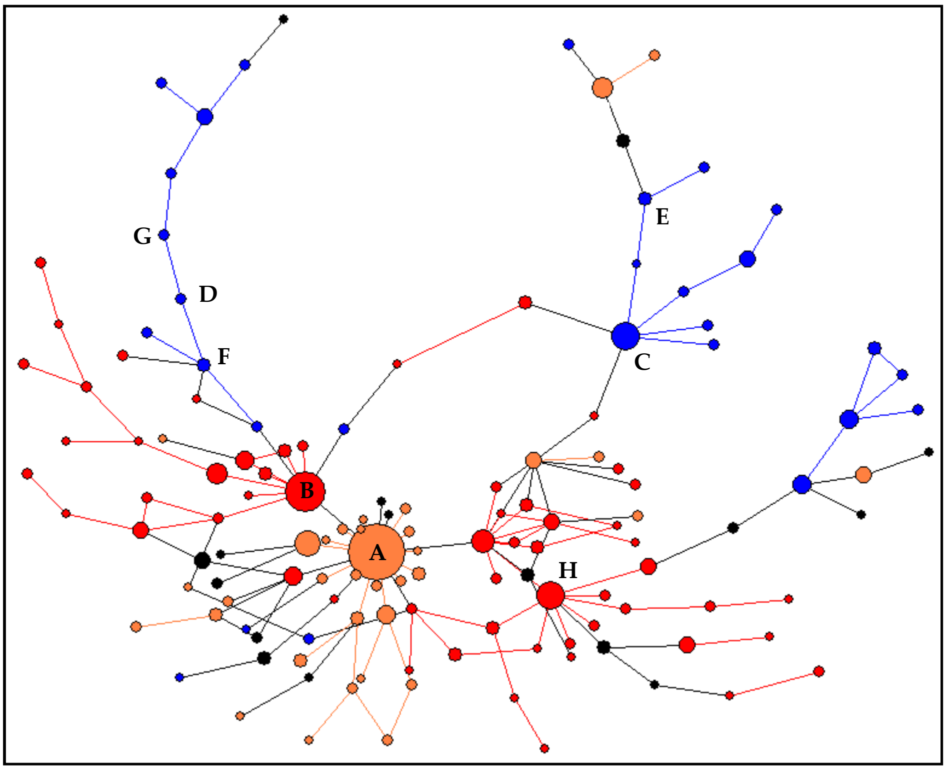
A : DOE - B : CAS - C : Helmholtz-Gemeinschaft e.V. - D: Chalmers Techniska Högskola - E: Russian Academy of Science - F: KTH - G: Roma la Sapienza - H:AIST



Network 8a: The international fuel cell network

Asia North America Europe Private Company

A : DOE - B : CAS - C : Helmholtz-Gemeinschaft e.V. - D: Chalmers Techniska Högskola - E: Russian Academy of Science - F: KTH - G: Roma la Sapienza - H:AIST



Network 8b: The international fuel cell network – Strongest links

Asia North America Europe Private Company

A : DOE - B : CAS - C : Helmholtz-Gemeinschaft e.V. - D: Chalmers Techniska Högskola - E: Russian Academy of Science - F: KTH - G: Roma la Sapienza - H:AIST

2. *Main types of Fuel Cells*

When it comes to the comparison of the scientific activity around fuel-cells and batteries, it appears that the tendency has changed since 2004 when the number of publications on fuel-cell related technology over-passed those on batteries (Figure 15). The figure shows the exponential trend line which fits the data from 1985 up to now, with a R^2 coefficient of 0.98.

Because fuel-cells are a more prolific field than batteries when it comes to scientific publications, we decided to study more in detail the characteristics of this growth using the tools introduces in section I.

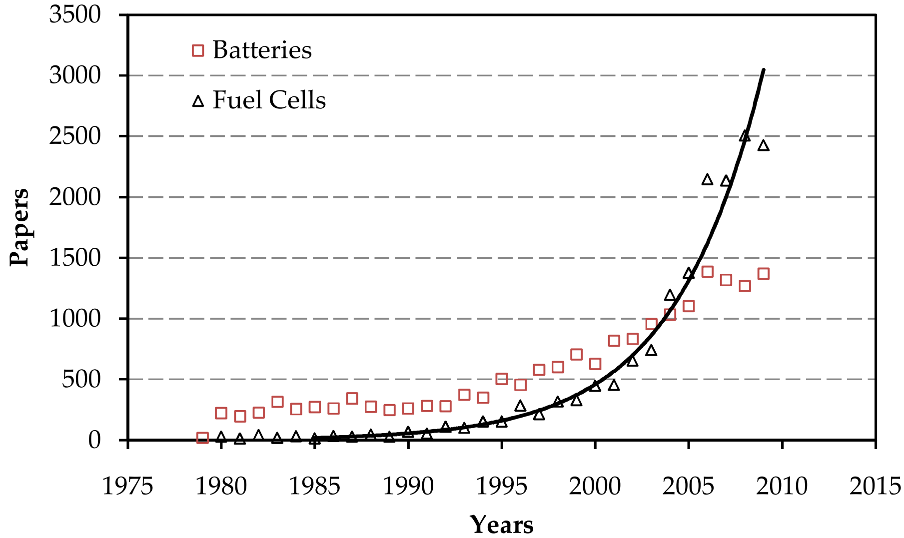


Figure 15: Evolution of the world scientific production on batteries and fuel cells

a. Geographical comparison

The first aspect of this growth is based on geographical considerations. Because scientific research is dependent on national policies and funding agencies orientation, we divided the evolution presented in Figure 15 into that of the top 5 producers of science in the fuel-cell field; namely the United States, China, Japan, South Korea and Germany. We will try to analyze these evolutions by taking in account the network studies of section II.1.b.

Figure 16 provides a focus on the period of rapid growth, ranging from 1995 to today. The first noticeable feature of this period is that the U.S.A and Japan are probably responsible for the dynamics of the scientific research on this field. Reasons for this rapid evolution of the world scientific production should therefore be looked for in United States or Japan national research policies.

Examples of such important policies in the United States can be found in the U.S. Department Of Energy (DOE) "Policy Act". That of 1992 gives a major kick to all FC programs by authorizing the development of fuel cell vehicles with special focus on cars and busses (U.S. Department of Energy, 1995). Even if it has been largely criticized for being "a smokescreen for the administration's rejection of environmentally friendly policies" (Pegg, 2003), President George Bush's 2003 Hydrogen Fuel Initiative (five-year $1.2 billion hydrogen development program) might have been a helping factor for the development of fuel cells in the United-States after 2005 but has failed to maintain the U.S on their rapid growth. On the other hand, DOE's Energy Efficiency and Renewable Energy program (EERE) has also been a major player in the development of the field after 2002, thanks to the *National Hydrogen Energy Roadmap* (EERE, 2002).

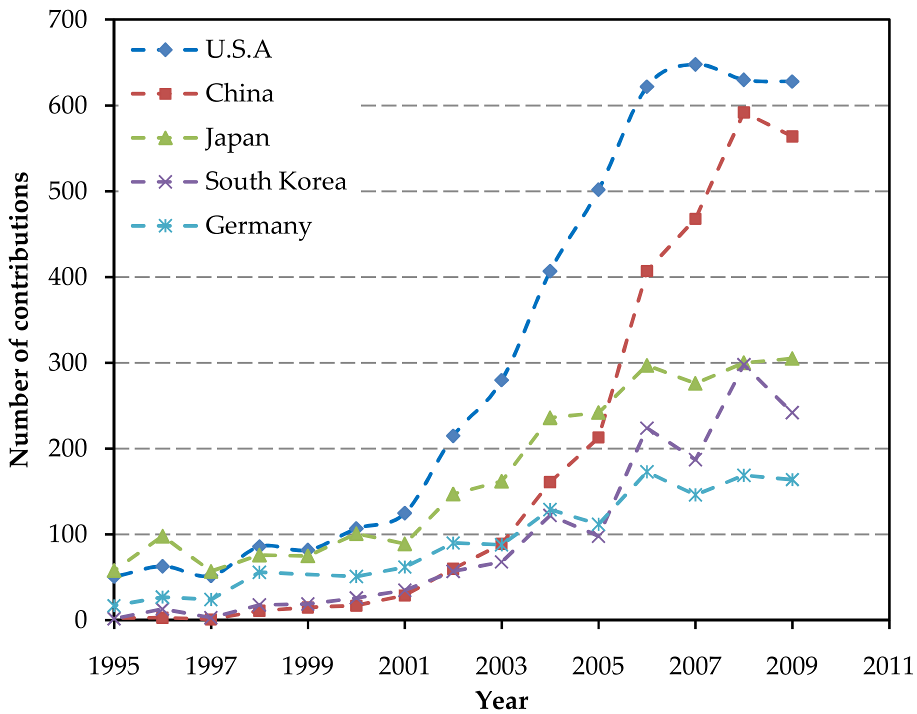


Figure 16: Number of papers by leading countries, 1995–2008

These considerations highlight again the major role played by regional hubs in the orientation of scientific research. Moreover, the strong network connections between the U.S. and Asia visible in Network 8b explain the explosion in fuel-cell research which followed in China (Figure 16). Even if Chinese contributions to research on fuel-cell technologies were inexistent before 1990 (Figure 17) it managed to take Japan's position as number 2 and eventually challenge the U.S.A as the most active country on FC research. Figure 17 shows that even if the increase in the number of American publications is significant after 1999, it was especially profitable to China which benefited from a transfer of knowledge enjoying its key position as America's first collaborator.

On the other hand, Europe didn't take advantage of this activity started by the U.S. as Germany, the major European actor in the field, stayed at a relatively constant share of the world scientific production on fuel cells (Figure 17).

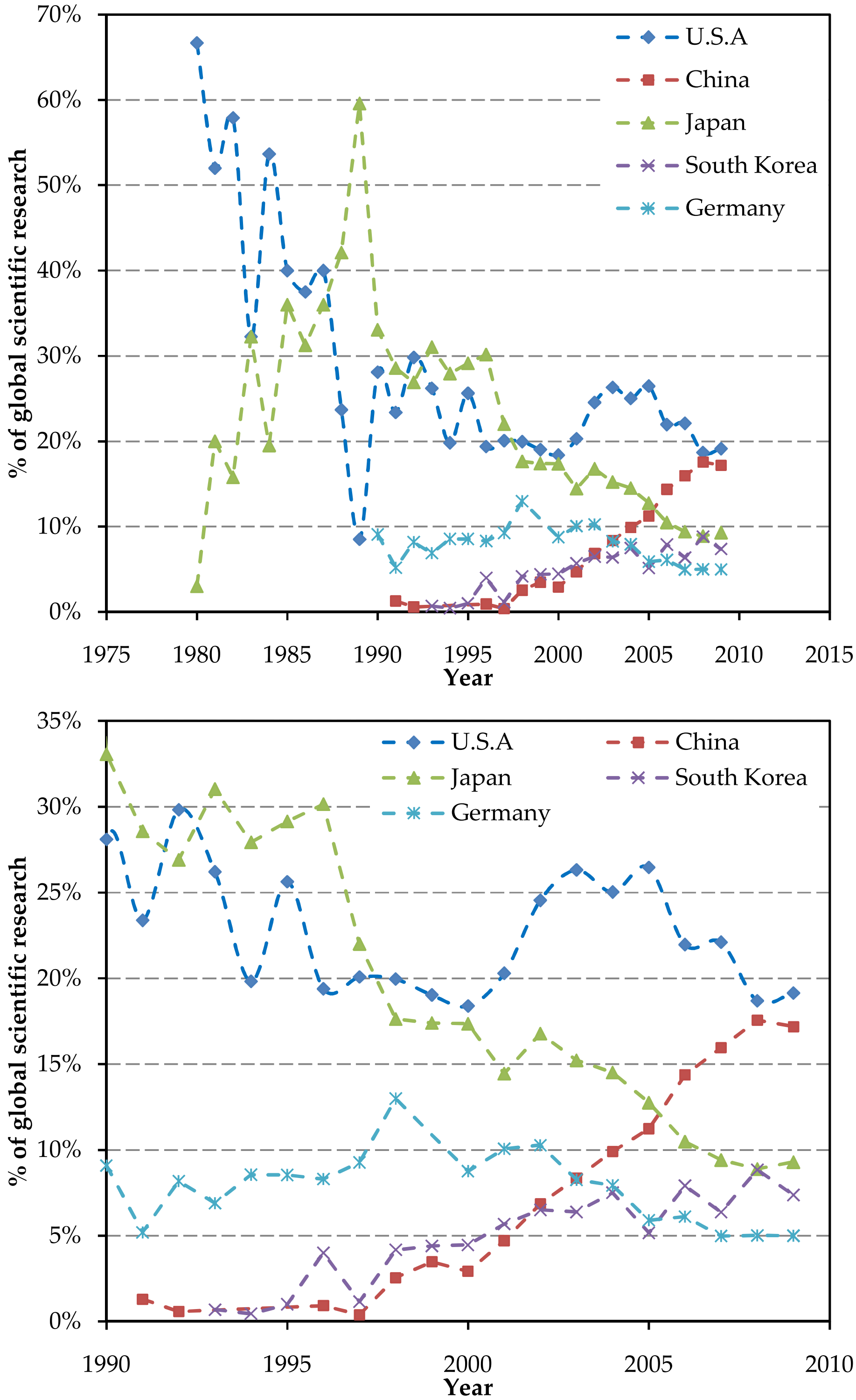


Figure 17: Top 5 countries - Share of the world scientific production on fuel cells

b. Growing and dying technologies

The second part of this study on fuel cells will use bibliometrics to focus on a more technical issue. The idea is to identify the fuel cell technologies which are most studied today, and those who are gradually disappearing. For this purpose we focused on both scientific peer reviewed papers and American patents.

Since the number of papers and patents has been rapidly increasing since 1999 (Figure 15 and Figure 18), the specific evolution of major fuel cell types was represented as a percentage of the total scientific production on FCs. As can be seen in Figure 21[6], certain types of fuel cells such as alkaline fuel cells (AFC), phosphoric acid fuel cells (PAFC) and molten carbonate fuel cells (MCFC) have been progressively abandoned and were not objects of the rapid increase observed in the field after 1999.

Alkaline fuel cells (AFC) are the first promising type of fuel cell developed and widely used in the U.S space program. They were chosen for their high performance and a 60% efficiency in space applications to send men to the moon (Department of Energy, 2009). However, they can be easily poisoned by carbon dioxide which makes them difficult to use in most common applications. Intensely developed during the fifties, research on AFC has gradually decreasing in the seventies, which results in a very low share of AFC articles and patents in Figure 19 and Figure 21.

On the other hand, phosphoric acid fuel cells (PAFC) are the first commercially developed fuel cell. They reach their higher efficiency (85%) when used for the co-generation of electricity and are quite robust against polluting agents. Because of their usually important size and weight, they are traditionally used in stationary applications. As shown on Figure 21, patents on PAFC have been decreasing in the eighties because of the previous decrease of scientific publications on the topic and the focus on lighter types of fuel cells for mobile applications. The regain of attention which can be observed in the early nineties could be due to the new use of PAFC in city buses and the increase in the number of papers published in the late eighties.

Finally, the most noticeable decrease in relative scientific attention is found for molten carbonate fuel cells (MCFC). This type of cell was originally studied for its promising characteristics: an efficiency neighboring 85%, high resistance to carbon monoxide and carbon dioxide, and a possibility to perform internal reforming. Moreover, because they operate at very high temperatures (650 °C), they don't need expensive metals to be used as catalysts; which considerably reduces the cost of a stack. Research on MCFC focuses today on the increase of the life time of the cell by looking for new corrosion-resistant

[6] Note that the interpretation of the curves is made much easier thanks to the use of smoothing splines, which even out the noise in the data. In all the curves of this section we used a smoothing parameter $\alpha = 20$ (refer to I.3.b for more information on smoothing splines).

materials (Department of Energy, 2009). MCFC still carry great expectations for the production of electricity in stationary plants and are preferred over PAFC for their much higher efficiency (Department of Energy, 2009). However, with the increasing need of light fuel cells operating at low temperature for mobile or portable applications, the scientific focus has gradually drifted away from MCFC towards smaller types of cells. The continuous decrease visible in Figure 19 and Figure 21 in the number of patents and articles focusing on this type of fuel cell reflects the shift of research toward fuel cell for applications such as cell phones, computers and cars.

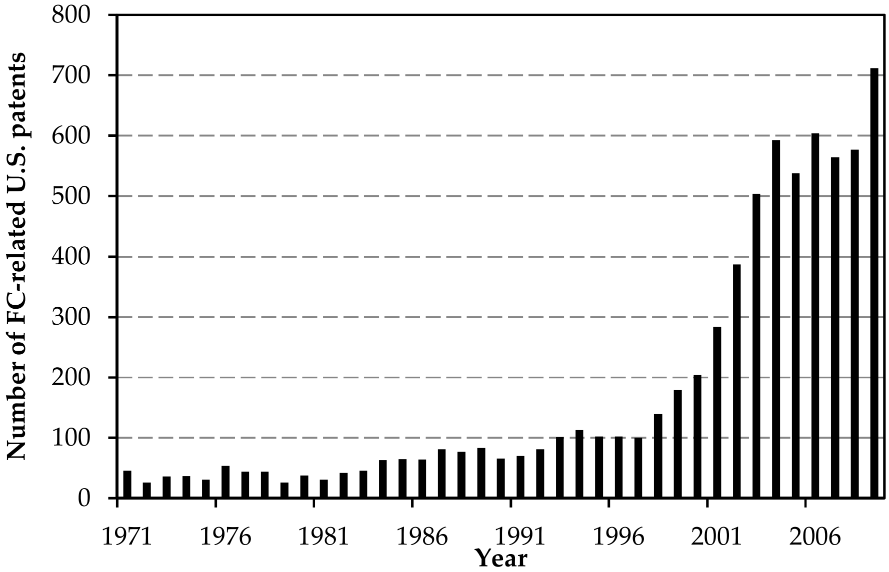


Figure 18: Number of FC-related U.S. patents per year

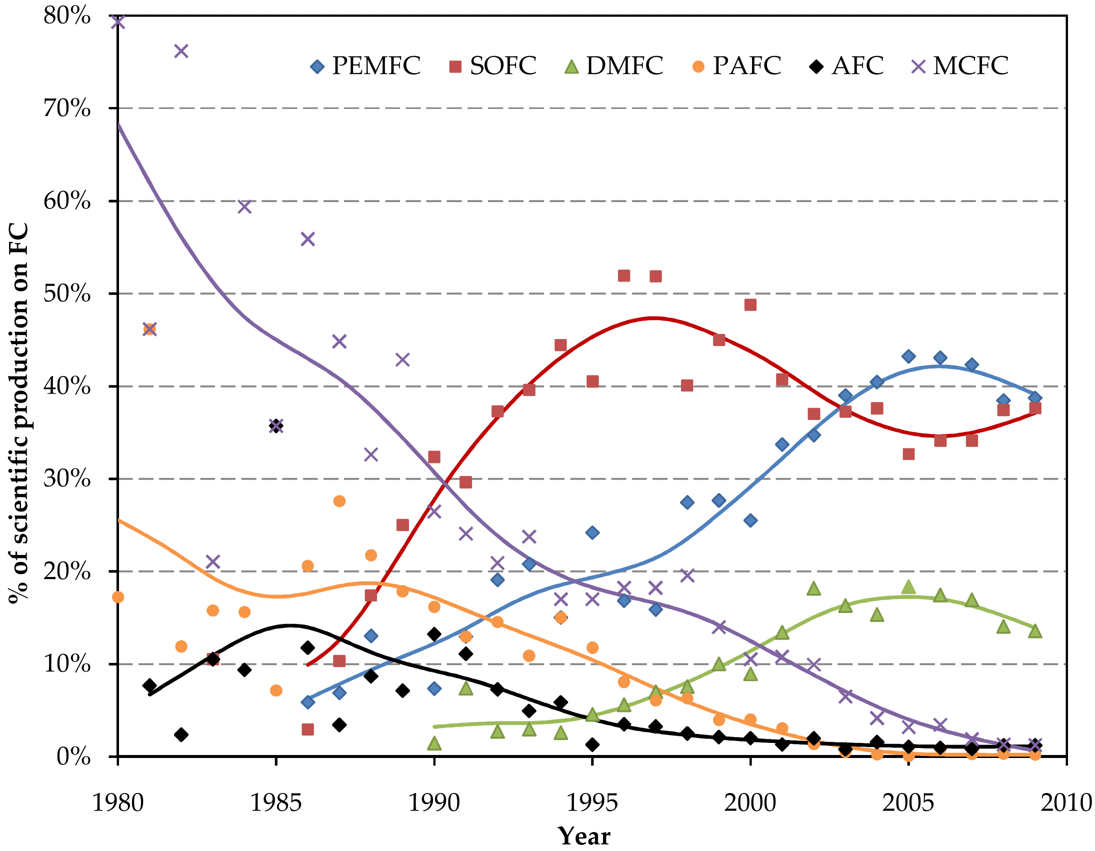


Figure 19: Repartition of fuel cell related articles

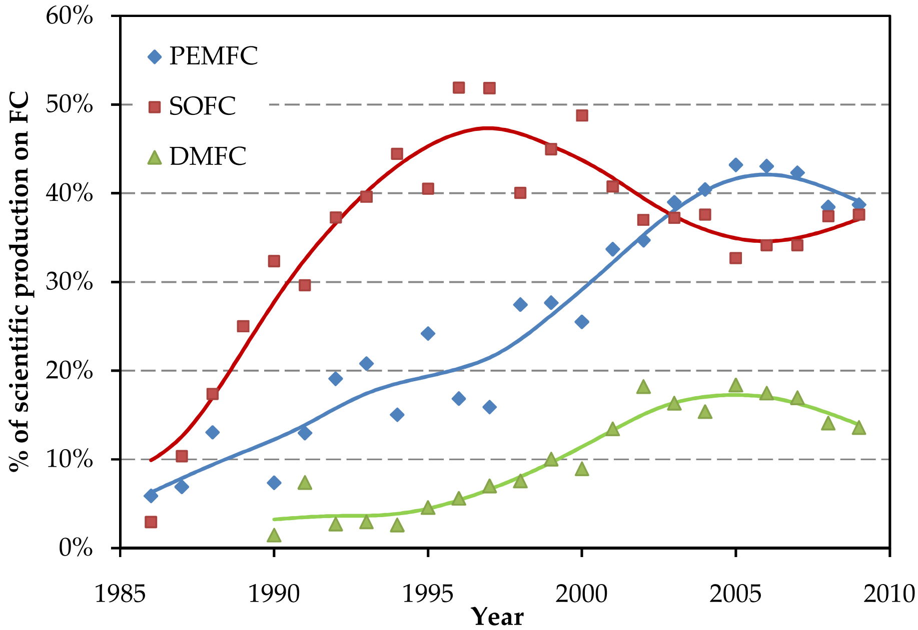


Figure 20: Repartition SOFC, DMFC and PEMFC articles

On the other side, the focus on three other fuel cell types has particularly increased: Solid oxide fuel cells (SOFC), proton exchange membrane fuel cell (PEMFC), also known as polymer electrolyte membrane (PEM), and direct methanol fuel cells (DMFC). It is interesting to note that the evolution of the number of patents (Figure 22) and articles (Figure 20) present the same characteristics. In both cases, we observe a maximum in the activity regarding SOFC around 1995 with a following decrease. After this period, it seems that SOFC are becoming again the most studied fuel cell type. Indeed, the importance of PEMFC, the most promising fuel cell type for the car industry, has been constantly decreasing since 2005; maybe due to increased interest in batteries.

By comparing Figure 20 and Figure 22 we can also note the delay with which American patents are submitted after scientific research: The scientific focus on PEMFC starts to be visible around 1986 but its industrial consequences are only visible starting from 1990.
Because the technology of DMFC is quite similar to that of PEMFC, the trend observed in Figure 20 for both fuel cell types is relatively identical. However, thanks to the interest of DMFC for portable applications, the number of patents submitted on this specific type of cell has not decreased yet.

If a projection in the future had to be made here, we would say that we will witness in the coming years a quasi-equal distribution of research efforts between SOFC and PEMFC technologies. Depending on the external conditions such as the price of oil and the will of automotive companies to develop fuel cell technologies, research on SOFC could soon lead again the research in the field. Also, depending on the strategies of mobile devices manufacturers, research on DMFC could start to increase independently from that on PEMFC. This tendency has already been visible in patents publication since year 2000 when patents on DMFC didn't follow the same trend as those on PEMFC (Figure 22). Again, public strategies may lead to the focus on one or the other of the technologies thanks to programs such as the U.S. Department of Energy/Solid State Energy Conversion Alliance Solid Oxide Fuel Cell Program (Gaz Technology Institute, 2009).

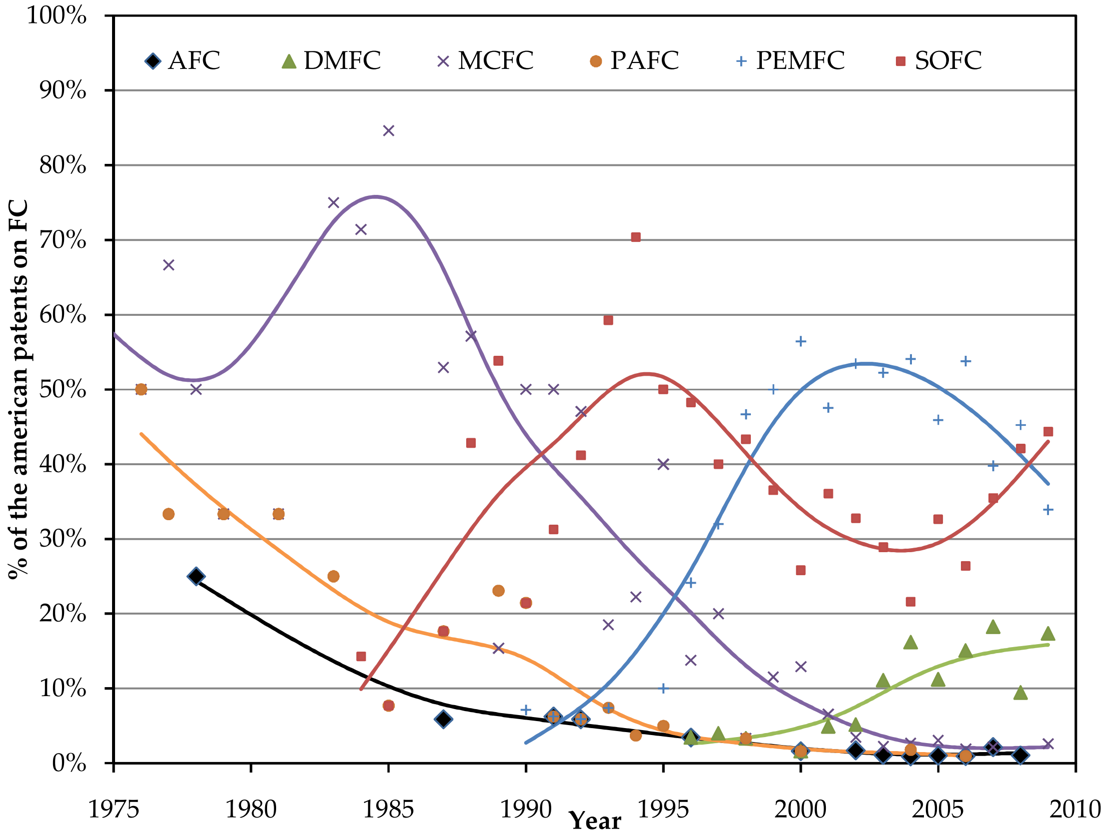


Figure 21: Repartition of fuel cell related patents

c. S-curve comparison

Another feature we can extract from the data presented in Figure 21[6] is the so called S-curve in diffusion of innovation (Rogers, 1962). The TRIZ concepts allow evaluating the maturity of a technology given for example the evolution in the number of patents submitted (Slocum, 1998). Since MCFC and SOFC are two alternatives for stationary production of electricity, it might be of interest to study their comparative evolution.

As can be seen in Figure 21, the activity on SOFC started to become noticeable when research on MCFC was at its top, around year 1985. This is also visible on Figure 19 where the drop in scientific publications on MCFC is strongly correlated with the increase of that on SOFC.

This consideration shows that the scientific focus is not likely to move back on MCFC: with the decrease of patent applications and scientific papers on this type of fuel cells, the key challenges once faced (i.e. durability) have less chance to be solved. In the mean time, with the investment put on SOFC, this type of cells will go on attracting researches and publications. However, contrarily to the systems TRIZ is usually used on, none of these technologies have reached their maturity yet because none was implemented on a large scale.

With the increase in the number of patents after 2004 on SOFC and the absence of challenging technologies all long the drop following 1995, solid oxide fuel cells are likely to become the preferred type of fuel cell for electric power plants. As explained earlier, we expect the recent increase in SOFC activity to go on in the coming years until the technology becomes fully operational and economically competitive.

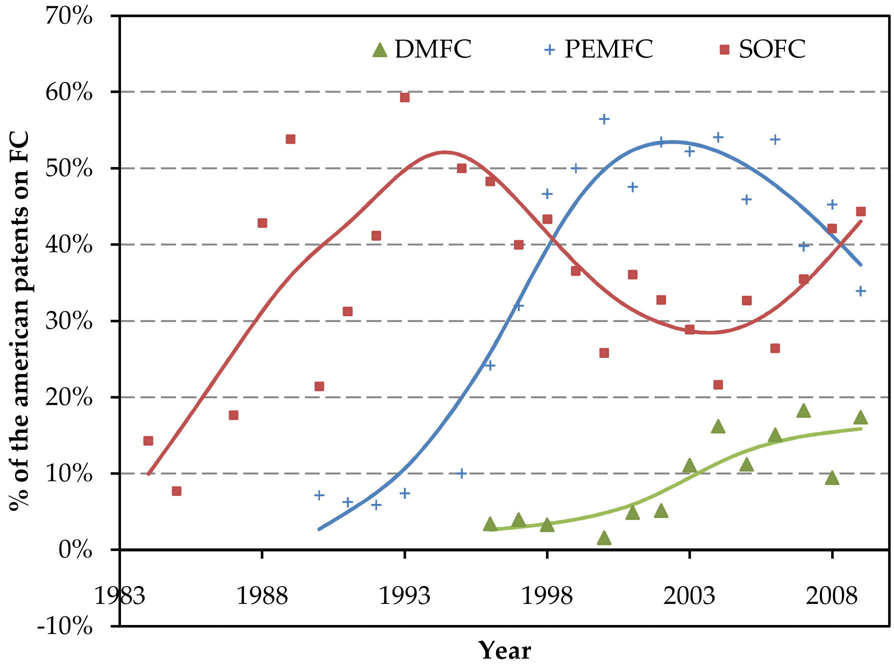


Figure 22: Repartition of SOFC, DMFC and PEMFC patents

3. *Time-lag before scientific impact*

The idea in this section is to evaluate the time span between the publication of a work and the year in which it receives maximum citations, i.e. the time-lag before maximum impact. Even if we already had a look at the time lag between scientific research and patents publications in the last sections, we can use bibliometrics here to evaluate the delay with which a finding regarding fuel cells and batteries starts to spread among the scientific community, and when it can be considered as old. In other words, we determined the citation peak for all the articles published in a certain year[7].
In order to do so, we have used the method described in section I-3, interpolating the data with splines. However, since both fields are in a constant growth (Figure 15), the calculated citation peak has to be investigated using a scale free indicator. This was done by taking the overall activity per year as a scale indicator. The citations received by a paper in a given year were therefore divided by the total number of articles published in that same year.

The results are shown here for all the articles published between 1996 and 2000, on both batteries (Figure 23) and fuel cells (Figure 24) related technologies.

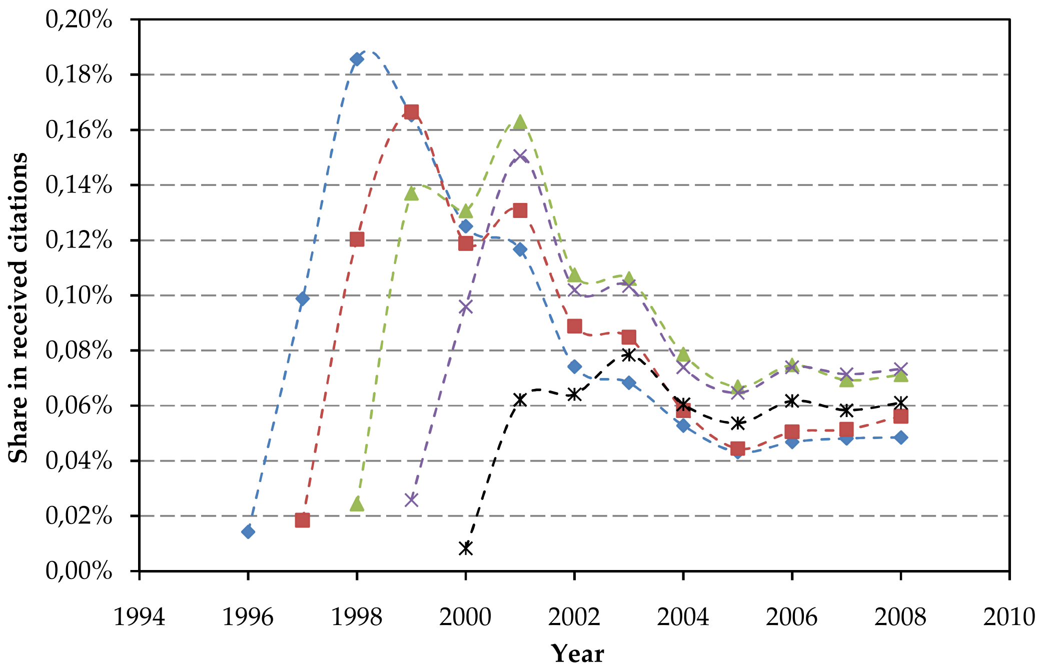


Figure 23: Share of citation received for articles on batteries published between 1996 and 2000

[7] The citation peak is often used to calculate the impact factor of journals. We used this technique in our article presented at the STI 2010 conference in Leiden: *Spline fitting tool for scientometric applications: estimation of citation peaks and publication time-lags.*

Our first conclusion is that citations follow the same trend, displaying a quick increase in the first years followed by a slower decrease in the share of received citations. Moreover we can conclude that research on both batteries and fuel cells has a quick impact on the community. Thanks to our method using splines, we were able to precisely calculate the time-lag before scientific impact[8] (Table 5). Even if this difference is not significant, we can note that the publication time-lag is smaller in the fuel cells field, which is the one with the fastest growth. When compared to other journals and fields, we realize how quick the spreading of knowledge is here: when the overall estimated time-lag seems to be around 3.5 years (Figure 25), some fields have a time lag before scientific impact longer than 12 years (Sabir, Campbell, & Archambault, 2010).

Field	Time-lag
Batteries	3.0 years
Fuel cells	2.6 years
General	3.6 years

Table 5: Time-lag in battery and fuel cell fields

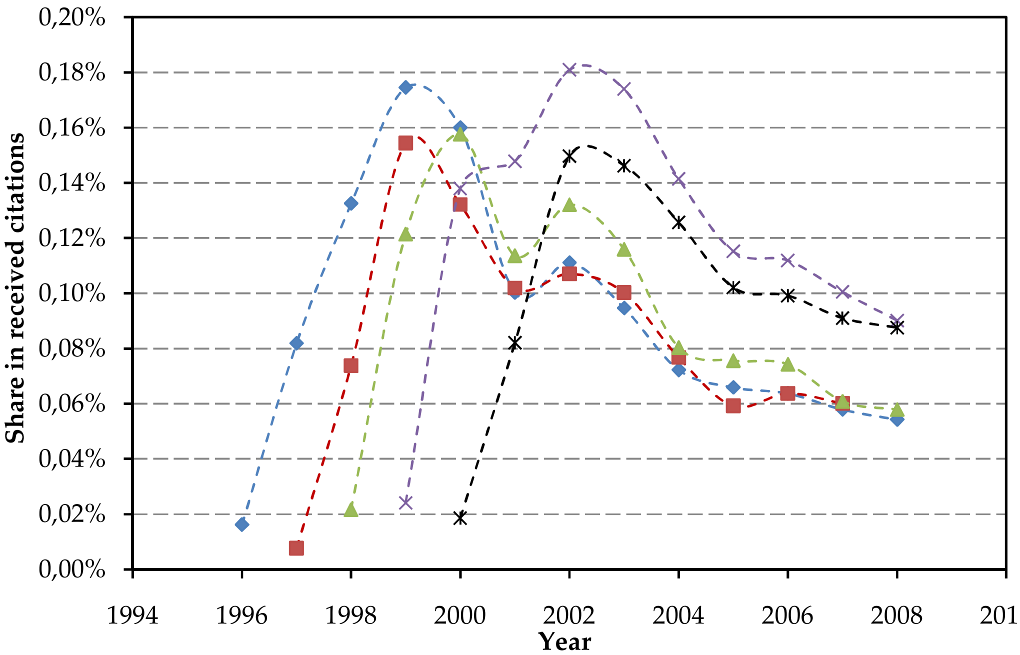


Figure 24: Share of citation received for articles on fuel cells published between 1996 and 2000

[8] The interest of our method based on Splines is to reduce the impact of holes in the data. If the citations received by the journal Blood in 1999 were missing, the peak would still be reached after 3.4 years. Without splines smoothing, the estimate would move from 3 to 4 years.

Also, this study highlights the key years in the evolution of the field. As can be seen in Figure 24, the articles published in 1999 have played a key role and have received more relative citations than any of the years before. This type of consideration can be associated with Figure 15 page 51 to obtain the time evolution in both the number of papers published and their quality or impact.

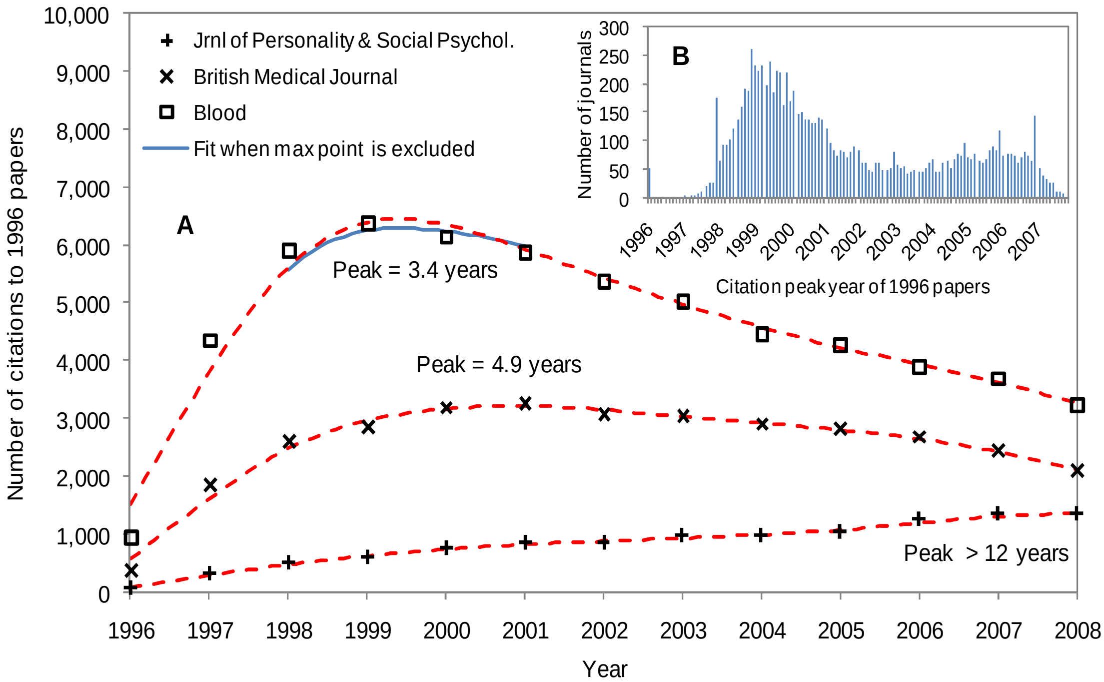


Figure 25: Number of citations per year for three journals (Sabir, Campbell, & Archambault, 2010)

Conclusion

The aim of this report was to present scientometric tools and methodology using mainly bibliometrics, a discipline focusing on the study of scientific publications. The mathematical tools described here are either already used by research analysts or subject to ongoing research: the work on splines fitting of bibliometric data was presented at the international conference on science and technology conference in Leiden, Netherlands (STI 2010), and the one on the probabilistic collaborative index has been submitted for publication.

The particular application of bibliometric indicators to batteries and fuel cells intended to present some of the key concepts behind science and technology evaluation. Similar methods are used to rank, map and evaluate scientific institutions in order to provide decision makers with the information they need. It allowed highlighting the importance of regional hubs in the development of a coherent research strategy, and the absence of European scientific hubs. As can be seen in the collaboration networks we built, Asia is the most active region, thanks to a very strong interaction between China, Korea and Japan, which stimulates other international partners' activity. A special focus was put on the weak position of Europe.

On the other hand, we used bibliometrics as a tool for innovation by evaluation the research activities around the different types of fuel cells. We correlated these observations with a patent analysis and the industrial focus on specific applications. Also, we characterized the progression speed of knowledge in these specific fields and compared it to the overall research. It appeared that both battery and fuel cell related research have a particularly quick impact, allowing both fields to grow faster than the overall scientific system.

The conclusions we drew in this specific case study should be taken as examples of what can be done today with bibliometrics and what sort of information we might be able to obtain. A complete evaluation would require further investigations and the use of other scientometric tools, such as surveys and environmental scans to identify the socio-political drivers of the observed phenomena.

Annexes

```
-- SQL Request BATTERIES

DROP TABLE bd_hicham.dbo.batteries

SELECT DISTINCT id
INTO bd_hicham.dbo.batteries
FROM Scopus_2009_v3.dbo.fts

WHERE
(
contains (title,'("Batteries" OR "Battery")
AND NOT ("Battery of test*" OR "test battery" OR "battery operated"
             OR "ability batter*" OR "assault and batter*" OR "assay
batter*" OR "assessment batter*"
             OR "battery awareness*" OR "battery factory worker*" OR
"battery factory worker*" OR "behavioral Batter*"
             OR "cognitive" OR "behavioral" OR "Functional
observational batter*" OR "Western Aphasia Batter" OR "Workaholism"
            )')
OR
contains (keyword_concat,'("Batteries" OR "Battery")
AND NOT ("Battery of test*" OR "test battery" OR "battery operated"
             OR "ability batter*" OR "assault and batter*" OR "assay
batter*" OR "assessment batter*"
             OR "battery awareness*" OR "battery factory worker*" OR
"battery factory worker*" OR "behavioral Batter*"
             OR "cognitive" OR "behavioral" OR "Functional
observational batter*" OR "Western Aphasia Batter" OR "Workaholism"
            )')

OR
contains (all_fields, '
"Alkaline battery" OR
"Aluminium battery" OR
"Atomic battery" OR
"Chromic acid cell" OR
"Clark cell" OR
"Daniell cell" OR
"Dry cell battery" OR
"Earth battery" OR
"Galvanic cell" OR
"Leclanché cell" OR
"Lemon battery" OR
"Lithium battery" OR
"Mercury battery" OR
"Molten salt battery" OR
"Nickel oxyhydroxide battery" OR
"Optoelectric nuclear battery" OR
```

```
"Organic radical battery" OR
"Oxyride battery" OR
"Paper battery" OR
"Reserve battery" OR
"Silver-oxide battery" OR
"Voltaic pile" OR
"Penny battery" OR
"Trough battery" OR
"Water-activated battery" OR
"Weston cell" OR
"Zinc-air battery" OR
"Zinc-carbon battery" OR
"Zinc chloride battery" OR
"Dr30 battery" OR
"Flow battery" OR
"Vanadium redox battery" OR
"Zinc-bromine flow battery" OR
"Lead-acid battery" OR
"Deep cycle battery" OR
"Marinised battery" OR
"VRLA battery" OR
"AGM battery" OR
"Gel battery" OR
"Lithium-ion battery" OR
"Air-fueled lithium-ion battery" OR
"Lithium ion polymer battery" OR
"Lithium iron phosphate battery" OR
"Lithium-sulfur battery" OR
"Lithium-titanate battery" OR
"Molten salt battery" OR
"Nickel-cadmium battery" OR
"Nickel-cadmium battery vented cell type" OR
"Nickel-iron battery" OR
"Nickel hydrogen battery" OR
"Nickel metal hydride battery" OR
"NiMH battery" OR
"Nickel-zinc battery" OR
"Organic radical battery" OR
"Polymer-based battery" OR
"Polysulfide bromide battery" OR
"Rechargeable alkaline battery" OR
"Smart battery system" OR
"Sodium-sulfur battery" OR
"Super iron battery" OR
"Super charge ion battery" OR
"Zinc matrix battery" OR
"Atomic battery" OR
"Optoelectric nuclear battery" OR
"Nuclear micro-battery" OR
"Backup battery" OR
"Battery vacuum tube" OR
"Battery pack" OR
"Biobattery" OR
"Button cell" OR
"CMOS battery" OR
"Commodity cell" OR
```

```
"Flow batteries" OR
"Lantern battery" OR
"Nanobatteries" OR
"Local battery" OR
"Photoflash battery" OR
"Traction battery" OR
"Watch battery" OR
"Zamboni pile"
')

)


-------------------------------------------------------
-- Step 2

INSERT INTO bd_hicham.dbo.batteries
SELECT DISTINCT id
FROM Scopus_2009_v3.dbo.fts WHERE
(
contains (all_fields,'
"Deep cycle battery" OR
"Deep cycle battery" OR
"Battery electricity" OR
"12 volt battery" OR
"2CR5" OR
"4SR44 battery" OR
"Aluminium battery" OR
"Atomic battery" OR
"Avcon" OR
"Battery Directive" OR
"Battery holder" OR
"Battery indicator" OR
"Battery room" OR
"Betavoltaics" OR
"Call2Recycle" OR
"Clean Energy Project" OR
"Cross Battery Assessment" OR
"Electric vehicle battery" OR
"Enercell" OR
"Evolta" OR
"Exide" OR
"Exide Industries" OR
"Float voltage" OR
"Gold Peak" OR
"Grove cell" OR
"Wilhelm Hellesen" OR
"In cell charge control" OR
"Leclanché cell" OR
"Mallory and Co Inc" OR
"lazy battery effect" OR "battery memory" OR
"Metal-air electrochemical cell" OR
"Optoelectric nuclear battery" OR
"Organic radical battery" OR
"Peukert''s" OR
"Power paper" OR
```

```
"Radioisotope piezoelectric generator" OR
"Radioisotope thermoelectric generator" OR
"Ragone chart" OR
("Rechargeable battery" AND NOT "requires a rechargeable battery") OR
"Reserve battery" OR
"Super Charge Ion Battery" OR
"Smart Battery Data" OR
"Smart Battery System" OR
"Solid-state battery" OR
"Car battery" OR
"Thermionic converter" OR
"Tough Solar" OR
"Traction battery" OR
"Trickle charging" OR
"Trough battery" OR
"Uniross"
')

OR  contains(all_fields,  '("cathode"  OR  "anode"  OR  "electrode"  OR
"electrolyte") AND "battery" AND NOT "battery powered"')
)
 AND id NOT IN (SELECT id FROM bd_hicham.dbo.batteries)

------------------------------------------------------
-- By journal

INSERT INTO bd_hicham.dbo.batteries
SELECT DISTINCT id
FROM Scopus_2009_v3.dbo.article AS art WHERE
(art.source_title='Proceedings  of  the  Annual  Battery  Conference  on
Applications and Advances')
 AND id NOT IN (SELECT id FROM bd_hicham.dbo.batteries)



------------------------------------------------------
-- Elimination

DELETE FROM bd_hicham.dbo.batteries
FROM bd_hicham.dbo.batteries AS h
      INNER JOIN Scopus_2009_v3.dbo.fts AS fts
      ON h.id = fts.id
WHERE (
contains(all_fields,'
"patient*" OR
"treatment" OR
"dementia" OR
"medical*" OR
"microorganism" OR
"fermentation" OR
"sucrose" OR
"glucose" OR
"neuro*" OR "rehab*" OR
"health" OR
"men" OR "women" OR "human" OR "adult" OR
"psycho*" OR
```

```
"managemenr" OR
"therapy" OR
"environment*" OR
"organ" OR "heart" OR
"microcontrol*" OR
"surgery" OR "respiration" OR "hospital" OR "injur*" OR "Bed" OR
"Halstead" OR
"Society of Automotive Engineers" OR
"geo*"
')

OR contains(title,'"Fuel cell*" OR "solar"')
)

DELETE FROM bd_hicham.dbo.batteries
FROM bd_hicham.dbo.batteries AS h
      INNER JOIN Scopus_2009_v3.dbo.article AS fts
      ON h.id = fts.id
WHERE (subject like '%health%'
OR subject='Earth Science and Geography'
OR subject='Life Sciences'
OR subject like '%social%'
OR source_title like '%medicine%'
OR source_title='Journal of Microbiology and Biotechnology'
OR source_title like '%dentistry%'

)
```

Annex 1: SQL Code for the construction of the battery envelop

```
-----------------------------------------------------------------------
--------------
-------------------Creation of the original envelop

DROP TABLE bd_hicham.dbo.fuel_cell

--Creating envelope v_01
--Step 1
SELECT DISTINCT fts.id
INTO bd_hicham.dbo.fuel_cell
FROM Scopus_2009_v3.dbo.fts AS fts
WHERE
(
contains (all_fields,'
("fuel cell" OR "fuel cells" OR "fuel-cell" OR "fuel-cells" OR fuelcell
OR fuelcells)
OR
((SOFC OR SOFCs OR "Solid Oxide FC*")
AND NOT (SOFSMC OR "Self-Organi Fuzzy*" OR "Self Organi Fuzzy*" OR
"shallow water optical fib cable*" OR "Sagnac optical fib current
sensor*"))
OR
((MCFC OR MCFCs OR "molten carbon FC*") AND NOT ("connection fuzzy*" OR
"fuzzyconnection*" OR "fuzzy cluster*" OR "micropylar canal forming
```

```
cell*"))')
)

--Step 2

INSERT INTO bd_hicham.dbo.fuel_cell
SELECT DISTINCT fts.id
FROM Scopus_2009_v3.dbo.fts AS fts WHERE
(
contains (all_fields,'
((DMFC OR DMFCs OR "Direct methanol FC*")AND NOT ("ligand" OR
"polyhalid*" OR "Discrete-time Memoryless Fading Channel" OR
"dorsomedial frontal cortex*" OR "dorsal medial frontal cortex*" OR
"dexterity measure for force condition*" OR "diabete mellitus female
control*" OR "driving mechanism with flexible connector*" OR
decamethylferrocene OR dimethylferrocene OR "decamethyl ferrocene*" OR
"dimethyl ferrocene*" OR "decamethylferrocenium*" OR "digital microflow
controller*" OR "DAMATO Multifixation Campimeter*"))')
)
 AND id NOT IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)

--Step 3
INSERT INTO bd_hicham.dbo.fuel_cell
SELECT DISTINCT fts.id
FROM Scopus_2009_v3.dbo.fts AS fts WHERE
(
contains (all_fields,'
((PAFC OR "phosphoric acid FC*") AND NOT ("ferric chloride*" OR "phase
automatic frequency control*" OR "pairwise additive function
counterpoise*" OR "site function counterpoise*" OR "proteasome
accessory factor*" OR "MBBR" OR "PAFC dosage" OR "coagula*" OR
"polyferric silicate chloride*" OR "wastewater" OR "pulmonary artery
flotation*" OR "percentage of eaten leaf area*" OR "flow cytometry*" OR
"Panic Attack Frequency Calendar*"))
')
)
 AND id NOT IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
--Step 4
INSERT INTO bd_hicham.dbo.fuel_cell
SELECT DISTINCT fts.id
FROM Scopus_2009_v3.dbo.fts AS fts WHERE
(
contains (all_fields,'
((PEFC OR PEFCs OR "Polymer electrolyte FC*") AND NOT (((paper OR
PEFC*) AND (Certifi* OR zertifi*)) OR "PEFC Zerti*" OR "PEFC certi*" OR
"PEFC certificate" OR "PEFC scheme" OR "forest*")
OR (PEFC OR PEFCs OR "Polymer electrolyte FC*") AND NOT (forest* OR
Sawmill* OR timber OR FSC OR "Recycled Paper*" OR "plasmid encoded*" OR
"PE fimbr*" OR "Peak expiratory flow*" OR FEV1 OR "flow volume curve*"
OR "embryo forming*" OR "elliptic Fourier coefficient*"))
')
)
 AND id NOT IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)

--Step 5
INSERT INTO bd_hicham.dbo.fuel_cell
```

```sql
SELECT DISTINCT fts.id
FROM Scopus_2009_v3.dbo.fts AS fts WHERE
(
contains (all_fields,'
(DFAFC OR "direct formic acid FC*")
OR
((PEMFC OR "Proton exchange membrane FC*") AND NOT "pulse electromagnetic field*")
OR
((DEFC OR "direct ethanol FC*") AND NOT "deficient control")
')
)
 AND id NOT IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)

--Step 6
INSERT INTO bd_hicham.dbo.fuel_cell
SELECT DISTINCT fts.id
FROM Scopus_2009_v3.dbo.fts AS fts WHERE
contains (all_fields,'"Gas Diffusion Layer*"')
 AND id NOT IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)

--Step 7
INSERT INTO bd_hicham.dbo.fuel_cell
SELECT DISTINCT fts.id
FROM Scopus_2009_v3.dbo.fts AS fts WHERE
contains (all_fields,'"Internal reforming*" OR ("Anode-supported*" AND NOT OLEDs) OR "microfuel cell*" OR "micro fuel cell*"')
 AND id NOT IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)

--Step 8
INSERT INTO bd_hicham.dbo.fuel_cell
SELECT DISTINCT fts.id
FROM Scopus_2009_v3.dbo.fts AS fts WHERE
contains (all_fields,'"Cathode catalyst layer*" and not "Solid Polymer Electrolyte Reactor*"')
 AND id NOT IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)


--Step 9
INSERT INTO bd_hicham.dbo.fuel_cell
SELECT DISTINCT article.id
FROM Scopus_2009_v3.dbo.article
WHERE source_title like '%fuel cell%'
AND article.id NOT IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)


-- Elimination

DROP FROM bd_hicham.dbo.fuel_cell
FROM bd_hicham.dbo.fuel_cell AS hic
      INNER JOIN Scopus_2009_v3.dbo.article AS art
      On art.id=hic.id
WHERE (source_title LIKE '%British_Plastics_and_Rubber%')
```

```
----------------------------------------
---- Creating the DB containing the sub-envelopes


----------------------------------------
-------------INSERTION of Alkaline Fuel Cells

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'AFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'AFC OR AFCS OR "alkaline Fuel*"')
)



----------------------------------------
-------------INSERTION of Direct Borohydride Fuel Cells

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'DBFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'DBFC OR DBFCs OR "direct borohydride F*"')
)


----------------------------------------
-------------INSERTION of Direct Ethanol Fuel Cells

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'DEFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'DEFC OR DEFCs OR "direct ethanol F*"')
)


----------------------------------------
-------------INSERTION of Direct Formic Acid Fuel Cells


INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'DFAFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'DFAFC OR DFAFCs OR "direct formic acid F*"')
)
```

```
-----------------------------------------
-------------INSERTION of Direct Methanol Fuel Cells

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'DMFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'DMFC OR DMFCs OR "direct methanol F*"')
)


-----------------------------------------
-------------INSERTION of Molten Carbon Fuel Cells

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'MCFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'MCFC OR MCFCs OR "molten carbon F*"')
)



-----------------------------------------
-------------INSERTION of Metal Hydride Fuel Cells

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'MHFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'MHFC OR MHFCs OR "metal hydride F*"')
)

-----------------------------------------
-------------INSERTION of Phosphoric Acid Fuel Cells

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'PAFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'PAFC OR PAFCs OR "phosphoric acid F*"')
)


-----------------------------------------
-------------INSERTION of Protonic Ceramic Fuel Cells
```

```sql
INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'PCFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'PCFC or PCFCs OR "proton ceramic F*"')
)

----------------------------------------
-------------INSERTION of Polymer Electrolyte Fuel Cells

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'PEFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'PEFC OR PEFCs OR "polymer electrolyte F*"')
)

----------------------------------------
-------------INSERTION of Proton Exchange Membrane Fuel Cells

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'PEMFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
(
contains (all_fields,'PEMFC OR PEMFCs')
OR
contains (all_fields,'"PEM F*"')
OR
contains (all_fields,'"proton exchange membrane F*"')
OR
contains (all_fields,'"PEM" NEAR "fuel cell*"')
OR
contains (all_fields,'"polymer electrolyte membrane f*"')
)
)


----------------------------------------
-------------INSERTION of Reformed Methanol Fuel Cells

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'RMFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'RMFC OR RMFCs OR "reform methanol F*"')
)
```

```sql
----------------------------------------
-------------INSERTION of Planar Solid Oxide Fuel Cells

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'PSOFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'PSOFC OR PSOFCs OR "planar solid oxid f*"')
)

----------------------------------------
-------------INSERTION of Solid Oxide Fuel Cells

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'SOFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'SOFC OR SOFCs OR "solid oxide F*"')
)
----------------------------------------
-------------INSERTION of Biological Fuel Cells
--- Also contains "Microbial Fuel Cells" and "Upflow Microbial Fuel
Cells"

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'BFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'BFC OR BFCs
OR UMFC OR UMFCs OR MFC OR MFCs OR "microbial F*"
OR "biofuel cell*" OR "bio-fuel cell*"
OR "biological Fuel*" OR "biological FC*"
OR "biochemical fuel*" OR "biochemical FC*"')
AND NOT (Contains (all_fields, '"Biocarbon fuel cell*"'))
)


----------------------------------------
-------------INSERTION of Solid Polymer Fuel Cells

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'SPFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'SPFC OR SPFCs OR "solid polymer F*"')
)
```

```
----------------------------------------
-------------INSERTION of Zinc Air Fuel Cells

INSERT INTO bd_hicham.dbo.fc_subcat
SELECT id,'ZAFC' FROM Scopus_2009_v3.dbo.fts

WHERE (
id IN (SELECT ID FROM bd_hicham.dbo.fuel_cell)
AND
contains (all_fields,'ZAFC OR ZAFCs OR "zinc air F*"')
)



-----------------------------------------------------------------------
---------
-- Affichage du contenu des differentes catégories


SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='AFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='DBFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='DEFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='DFAFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='DMFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='MCFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='MHFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='PAFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='PCFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='PEFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='PEMFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='RMFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='PSOFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='SOFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='BFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='SPFC'
SELECT count(*) FROM bd_hicham.dbo.fc_subcat WHERE fc_type='ZAFC'

SELECT count(*) FROM bd_hicham.dbo.fuel_cell WHERE (
      id NOT IN (SELECT id FROM bd_hicham.dbo.fc_subcat)
      )
```

Annex 2: SQL code for the construction of the fuel cell envelop

Country	Size	ScIR	Country	Size	ScIR
United States	4727616	**1.03**	Tunisia	18533	**0.93**
China	1363291	**0.57**	Belarus	17767	**0.82**
United Kingdom	1286866	**1.36**	Venezuela	17693	**0.96**
Japan	1254956	**0.72**	Yugoslavia	17199	**0.63**
Germany	1191925	**1.43**	Morocco	16334	**1.17**
France	853298	**1.43**	Cuba	15415	**0.77**

Canada	647254	**1.33**	Colombia	14946	**1.22**
Italy	626155	**1.19**	Lithuania	12993	**0.84**
Spain	466782	**1.10**	Algeria	12061	**1.15**
Australia	417356	**1.20**	Jordan	10981	**0.74**
Russia	413857	**1.00**	Estonia	10899	**0.99**
India	410907	**0.57**	Kenya	10223	**1.44**
Netherlands	358036	**1.39**	Bangladesh	9720	**1.08**
Rep. of Korea	330609	**0.80**	Indonesia	9402	**1.56**
Switzerland	257749	**1.66**	United Arab Emirates	9128	**1.08**
Sweden	256729	**1.38**	Kuwait	8851	**0.73**
Brazil	246782	**0.87**	Lebanon	7742	**0.91**
Taiwan	239861	**0.52**	Philippines	7659	**1.28**
Poland	218709	**0.98**	Viet Nam	7650	**1.52**
Belgium	194056	**1.45**	Puerto Rico	7016	**1.18**
Turkey	178243	**0.48**	Iceland	6085	**1.20**
Israel	157321	**1.13**	Armenia	5748	**0.94**
Austria	135318	**1.33**	Uruguay	5704	**1.20**
Denmark	134616	**1.37**	Serbia Montenegro	5677	**0.71**
Finland	126996	**1.15**	Latvia	5551	**0.97**
Greece	112483	**0.96**	Uzbekistan	4884	**0.74**
Mexico	99499	**1.10**	Sri Lanka	4700	**1.02**
Norway	98574	**1.21**	Georgia	4676	**1.02**
Czech Rep.	97562	**1.04**	Tanzania	4666	**1.41**
Singapore	88458	**1.00**	Peru	4600	**1.46**
New Zealand	82003	**1.14**	Cyprus	4507	**1.15**
Ukraine	77433	**0.91**	Cameroon	4382	**1.30**
Portugal	75326	**1.22**	Ethiopia	4347	**1.20**
Argentina	75280	**1.01**	Oman	4332	**0.88**
Hungary	74352	**1.20**	Costa Rica	4020	**1.27**
South Africa	73678	**1.03**	Azerbaijan	4014	**0.55**
Iran	71069	**0.57**	Zimbabwe	3868	**1.14**
Ireland	60819	**1.16**	Uganda	3805	**1.38**
Egypt	48764	**0.83**	Ghana	3468	**1.15**
Romania	43346	**1.06**	Kazakhstan	3375	**0.75**
Thailand	42953	**1.11**	Nepal	3375	**1.06**
Chile	37857	**1.25**	Senegal	3132	**1.19**
Slovakia	36251	**1.05**	Moldova	3078	**1.06**
Croatia	32909	**0.61**	Ecuador	2503	**1.25**
Slovenia	30511	**0.83**	Luxembourg	2494	**1.21**
Bulgaria	30377	**1.08**	Jamaica	2323	**0.75**
Malaysia	29942	**0.87**	Macedonia FYR	2244	**0.76**
Saudi Arabia	27193	**0.74**	Syria	2229	**0.98**
Pakistan	25426	**0.69**	Cote d'Ivoire	2214	**1.10**

Nigeria	20630	**0.57**	Qatar	2196	**0.88**

Table 6: ScIR for the 100 top producers of science

Size Ranking	Institution	Size (papers)	PCI
1	**DOE - US Department of Energy**	**1348**	**0.88**
2	**Chinese Academy of Sciences**	**821**	**0.93**
3	**Helmholtz-Gemeinschaft e.V.**	**611**	**0.73**
4	**AIST - Advanced Industrial Science and Technology**	**546**	**0.77**
5	**Penn State - Pennsylvania State University**	**434**	**1.02**
6	**University of Science and Technology (UST)**	**427**	**1.22**
7	**Tsinghua University**	**340**	**0.87**
8	**ETH - Eidgenössische Technische Hochschule**	**314**	**0.93**
9	**University of Connecticut**	**303**	**0.90**
10	**CNRS - Centre national de la recherche scientifique**	**294**	**1.19**
11	**NRC Canada - National Research Council**	**290**	**1.03**
12	**Harbin Institute of Technology**	**275**	**1.00**
13	**CNR - Consiglio Nazionale delle Ricerche**	**269**	**0.91**
14	**Georgia Tech - Georgia Institute of Technology**	**264**	**0.80**
15	**Shanghai Jiao Tong University**	**260**	**0.85**
16	**University of London, Imperial College London**	**257**	**1.05**
17	**Nanyang Technological University (NTU)**	**233**	**1.10**
18	**Consejo Superior de Investigaciones Científicas**	**232**	**0.81**
19	**Tokyo Institute of Technology**	**232**	**1.00**
20	**Seoul National University**	**231**	**1.22**
21	**Max-Planck-Gesellschaft**	**224**	**0.95**
22	**DTU - Danmarks Tekniske Universitet**	**221**	**0.88**
23	**University of South Carolina**	**217**	**1.04**
24	**FCRC - Queen's-RMC Fuel Cell Research Centre**	**213**	**1.32**
25	**IEEE - Institute of Electrical and Electronics Engineers**	**205**	**1.27**
26	**Virginia Polytechnic Institute and State University**	**204**	**1.11**
27	**Kyushu University**	**197**	**1.12**
28	**KTH - Kungliga Tekniska högskolan**	**190**	**0.94**
29	**University of Michigan**	**182**	**1.07**
30	**University of Tokyo**	**180**	**1.25**
31	**University of Science and Technology of China**	**180**	**0.79**
32	**Case Western Reserve University**	**177**	**0.92**
33	**Kyoto University**	**172**	**1.24**
34	**Tohoku University**	**170**	**1.14**
35	**General Motors Corp.**	**169**	**0.89**

36	**University of Illinois At Urbana-Champaign**	**163**	**1.08**
37	**Yonsei University**	**163**	**1.10**
38	**Russian Academy of Sciences**	**162**	**0.88**
39	**Council of Scientific and Industrial Research (CSIR)**	**157**	**0.67**
40	**Jilin University**	**154**	**1.18**
41	**Korea Advanced Institute of Science and Technology**	**154**	**1.03**
42	**US Navy**	**152**	**0.99**
43	**Mitsubishi Group**	**149**	**1.13**
44	**Newcastle University (UK)**	**146**	**0.92**
45	**Samsung Group**	**139**	**1.03**
46	**University of British Columbia**	**137**	**1.28**
47	**Yuan Ze University**	**137**	**1.00**
48	**Korea University**	**136**	**1.09**
49	**Yokohama National University**	**136**	**1.01**
50	**University of Yamanashi**	**135**	**1.01**

Table 7a: FC - Ranking by size of the 50 largest institutions

Size Ranking	Institution	Size (papers)	PCI
1	**Huafan University**	**40**	**2.08**
2	**Doshisha University**	**32**	**1.66**
3	**Simon Fraser University**	**62**	**1.59**
4	**University of Southern California**	**46**	**1.57**
5	**Mie University**	**47**	**1.54**
6	**McGill University**	**34**	**1.53**
7	**National Chiao Tung University**	**58**	**1.50**
8	**American Institute of Aeronautics and Astronautics**	**53**	**1.49**
9	**National Taiwan University**	**107**	**1.48**
10	**Sun Yat-sen University**	**95**	**1.47**
11	**Gwangju Institute of Science and Technology**	**47**	**1.45**
12	**Michigan State University**	**64**	**1.45**
13	**University of Puerto Rico**	**32**	**1.45**
14	**Kansai Electric Power Co. Inc.**	**44**	**1.45**
15	**Industrial Technology Research Institute**	**51**	**1.45**
16	**Oita University**	**47**	**1.42**
17	**Chongqing University**	**54**	**1.40**
18	**Feng Chia University**	**44**	**1.40**
19	**KMUTT - King Mongkut's University of Technology Thonburi**	**38**	**1.40**
20	**Drexel University**	**30**	**1.40**
21	**JST - Japan Science and Technology Agency**	**87**	**1.39**

22	**Florida State University**	**32**	**1.38**
23	**University of Maryland College Park**	**41**	**1.36**
24	**Chulalongkorn University**	**58**	**1.34**
25	**FCRC - Queen's-RMC Fuel Cell Research Centre**	**213**	**1.32**
26	**NIST - National Institute of Standards and Technology**	**48**	**1.32**
27	**HUST - Huazhong University of Science and Technology**	**34**	**1.32**
28	**KIST - Korea Institute of Science and Technology**	**60**	**1.31**
29	**Toho Gas Co. Ltd.**	**30**	**1.29**
30	**Inha University**	**58**	**1.28**
31	**University of Tennessee at Knoxville**	**38**	**1.28**
32	**University of British Columbia**	**137**	**1.28**
33	**Kogakuin University**	**43**	**1.28**
34	**IEEE - Institute of Electrical and Electronics Engineers**	**205**	**1.27**
35	**Queen's University**	**53**	**1.27**
36	**Moscow State University**	**34**	**1.27**
37	**Hyundai Group**	**35**	**1.26**
38	**University of Tokyo**	**180**	**1.25**
39	**University of California, Berkeley**	**105**	**1.25**
40	**Université de Poitiers**	**74**	**1.25**
41	**Kyoto University**	**172**	**1.24**
42	**University of Dayton**	**38**	**1.24**
43	**UIMC - Universidade de Aveiro**	**54**	**1.23**
44	**Dalian University of Technology**	**96**	**1.23**
45	**University of Illinois at Chicago**	**41**	**1.23**
46	**A-STAR Agency for Science, Technology and Research**	**37**	**1.22**
47	**Universidad de La Laguna**	**43**	**1.22**
48	**Seoul National University**	**231**	**1.22**
49	**University of Science and Technology (UST)**	**427**	**1.22**
50	**University of Science and Technology Beijing**	**78**	**1.21**

Table 7b: FC - Ranking by PCI of the top institutions with more than 30 papers

Size Ranking	Institution	Size (papers)	PCI
1	**DOE - US Department of Energy**	**567**	**1.07**
2	**Chinese Academy of Sciences**	**371**	**0.95**
3	**Tsinghua University**	**270**	**0.83**
4	**Central South University**	**229**	**0.74**
5	**Hanyang University**	**196**	**1.18**
6	**Zhejiang University**	**193**	**0.81**
7	**AIST - Advanced Industrial Science and Technology**	**188**	**0.91**
8	**Harbin Institute of Technology**	**180**	**0.72**
9	**Wuhan University**	**161**	**0.72**
10	**Council of Scientific and Industrial Research (CSIR)**	**147**	**0.82**
11	*IEEE - Institute of Electrical and Electronics Engineers*	**143**	**1.31**
12	**University of Wollongong**	**142**	**0.96**
13	**Kyoto University**	**141**	**1.14**
14	**Fudan University**	**141**	**1.06**
15	**University of Science and Technology (UST)**	**129**	**1.25**
16	**Beijing Institute of Technology**	**126**	**0.82**
17	**Tokyo Institute of Technology**	**124**	**0.98**
18	**Seoul National University**	**124**	**0.93**
19	**Korea Advanced Institute of Science and Technology**	**124**	**0.92**
20	**CNRS - Centre national de la recherche scientifique**	**116**	**1.10**
21	**University of California, Berkeley**	**107**	**1.29**
22	**ETH - Eidgenössische Technische Hochschule**	**107**	**0.85**
23	**Shanghai Jiao Tong University**	**103**	**1.07**
24	**UNIROMA1 - Università degli studi di Roma la Sapienza**	**103**	**0.99**
25	**Tokyo University of Science**	**103**	**0.75**
26	**Saga University**	**102**	**0.97**
27	**NASA**	**98**	**1.06**
28	**MIT - Massachusetts Institute of Technology**	**95**	**1.05**
29	**University of Science and Technology of China**	**89**	**0.79**
30	**US Army**	**89**	**0.53**
31	**Samsung Group**	**88**	**1.06**
32	**Gyeongsang National University**	**87**	**0.94**
33	**Osaka Prefecture University**	**87**	**0.75**
34	**Kyushu University**	**86**	**1.07**
35	**Bar-Ilan University**	**86**	**1.05**
36	**Nankai University**	**84**	**0.80**
37	**National Taiwan University**	**83**	**1.33**
38	**University of South Carolina**	**83**	**1.04**

39	Tianjin University	83	1.00
40	Tohoku University	81	1.19
41	Illinois Institute of Technology	78	1.08
42	National University of Singapore (NUS)	78	1.04
43	Xiamen University	77	1.18
44	JST - Japan Science and Technology Agency	75	1.49
45	Mitsubishi Group	73	1.05
46	NTT - Nippon Telegraph and Telephone Corp.	73	0.46
47	University of Texas at Austin	71	0.79
48	Chonnam National University	70	1.04
49	UPMC - Université Pierre et Marie Curie	69	1.14
50	Dalhousie University	67	1.12

Table 8a: Batteries - Ranking by size of the 50 largest institutions

PCI Ranking	Institution	Size (papers)	PCI
1	JST - Japan Science and Technology Agency	75	1.49
2	Konkuk University	31	1.41
3	University of Florida	33	1.39
4	Tokyo Metropolitan University	62	1.39
5	Osaka University	58	1.39
6	Industrial Technology Research Institute	56	1.36
7	National Taiwan University	83	1.33
8	Yokohama National University	32	1.31
9	IEEE - Institute of Electrical and Electronics Engineers	143	1.31
10	University of California, Berkeley	107	1.29
11	Tel Aviv University	37	1.28
12	Mie University	59	1.26
13	University of Science and Technology (UST)	129	1.25
14	Korea University	58	1.20
15	Tohoku University	81	1.19
16	Hanyang University	196	1.18
17	Yonsei University	45	1.18
18	Xiamen University	77	1.18
19	National Central University	58	1.16
20	Yamaguchi University	63	1.15
21	UPMC - Université Pierre et Marie Curie	69	1.14
22	Kyoto University	141	1.14
23	University of Tokyo	52	1.12
24	Tokyo University of Agriculture and Technology	39	1.12
25	Dalhousie University	67	1.12

26	**University of St. Andrews**	**42**	**1.11**
27	**CNRS - Centre national de la recherche scientifique**	**116**	**1.10**
28	**Université de Picardie Jules Verne**	**47**	**1.09**
29	**Uppsala universitet**	**47**	**1.09**
30	**Hydro-Québec**	**52**	**1.09**
31	**Université Bordeaux 1**	**43**	**1.08**
32	**Illinois Institute of Technology**	**78**	**1.08**
33	**Shanghai Jiao Tong University**	**103**	**1.07**
34	**Kyushu University**	**86**	**1.07**
35	**DOE - US Department of Energy**	**567**	**1.07**
36	**SCUT - South China University of Technology**	**47**	**1.07**
37	**Fudan University**	**141**	**1.06**
38	**University of New South Wales**	**38**	**1.06**
39	**NASA**	**98**	**1.06**
40	**Samsung Group**	**88**	**1.06**
41	**Mitsubishi Group**	**73**	**1.05**
42	**MIT - Massachusetts Institute of Technology**	**95**	**1.05**
43	**Bar-Ilan University**	**86**	**1.05**
44	**Chonnam National University**	**70**	**1.04**
45	**University of South Carolina**	**83**	**1.04**
46	**National University of Singapore (NUS)**	**78**	**1.04**
47	**University of Hawaii at Manoa**	**32**	**1.04**
48	**OCU - Osaka City University**	**49**	**1.04**
49	**National Tsing Hua University**	**65**	**1.03**
50	**Henan Normal University**	**37**	**1.02**

Table 8b: Batteries - Ranking by PCI of the top institutions with more than 30 papers

References

Allison, P. D. (1980). Inequality and Scientific Productivity. *Social Studies of Science , 10* (163).

Borgatti, S., Everett, M., & Freeman, L. *Ucinet for Windows: Software for Social Network Analysis.* Harvard, MA: Analytic Technologies.

Department of Energy. (2009, June 23). *Fuel Cell Technologies Program - Types of Fuel Cells*. Accessed on June 16, 2010, on EERE - Energy Efficiency & Renewable Energy: http://www1.eere.energy.gov/hydrogenandfuelcells/fuelcells/fc_types.html

Department of Energy. (2009, August 5). *Press releases*. Accessed on June 7, 2010, on U.S. Department Of Energy : http://www.energy.gov/7749.htm

EERE. (2002). *National Hydrogene Energy Roadmap.* EERE - Energy Efficiency & Renewable Energy, United States Department of Energy, Washigton, DC.

Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science , 11* (2), 89-121.

Gaz Technology Institute. (2009). *DOE/SECA SOFC Program*. Accessed on June 2010, 17, on GTI corporate website: http://www.gastechnology.org/webroot/app/xn/xd.aspx?it=enweb&xd=1researchcap\1_2energysyst\1_2_4_majorprojects\doesecasofcprogram.xml

Green, P., & Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models.* London: CHAPMAN & HALL/CRC.

Hubert, J. J. (1976). On the Naranan interpretation of the Bradford's Law. *Journal of the American Society for Information Science* , 339.

Katz, J. S. (2000). Scale-independent indicators and research evaluation. *Science and Public Policy , 27*, 23-26.

Katz, J. S. (1999). The self-similar science system. *Research Policy , 28*, 501–517.

Mandelbrot, B. B. (1997). *Fractals and Scaling In Finance.* New-York: Springer.

Miquel, J. F., & Okubo, Y. (1994). Structure of international collaboration in science-part II: Comparisons of profiles in countries using a link indicator. *Scientometrics* , 271-297.

Naranan, S. (1970). Bradford's Law of Bibliography of Science: an Interpretation. *Nature , 227*, 631 - 632.

Naranan, S. (1971). Power law relations in science bibliography - a self-consistent interpretation. *Journal of Documentation , 27* (2), 83-97.

Narvaez-Berthelemot, N. (1990). An index to measure the international collaboration of developing countries based on the participation of national institutions: The case of Latin America. *Scientometrics* , 37-44.

Pegg, J. (2003, February 6). *Bush Hydrogen Initiative Faces Many Obstacles*. Accessed on June 10, 2010, on Environment News Service: http://www.ens-newswire.com/ens/feb2003/2003-02-06-10.html

Reinsch, C. H. (1967). Smoothing by spline functions. (S. B. Heidelberg, Éd.) *Numerische Mathematik* , 177-183.

Rogers, E. M. (1962). *Diffusion of Innovation.* Free Press of Glencoe.

Sabir, H., Campbell, D., & Archambault, E. (2010). Spline fitting tool for scientometric applications: estimation of citation peaks and publication time-lags. *Proceedings of the 11th conference on science and technology indicators* .

Schroeder, M. (1991). *Fractals, Chaos, Power Laws.* Freeman.

Science-Metrix Corp. (2010). *Methods - Scientometrics.* Accessed on June 19, 2010, on Science-Metrix corporate website:
http://www.science-metrix.com/eng/methods_scientometrics_t.htm

Slocum, M. S. (1998). *Technology Maturity Using S-curve Descriptors.* Accessed on June 17, 2010, on TRIZ Journal:
http://www.triz-journal.com/archives/1998/12/a/index.htm

Taussky, O. (1949). A recurring theorem on determinants. *American Mathematical Monthly* , *56*, 672–676.

U.S. Department of Energy. (1995). *DOE Automotive Fuel Cell Development Program.* Washington.

Ueberhuber, C. W. (1995). *Numerical computation: methods, software, and analysis, Volume 1.* Berlin: Springer.

Yamashita, Y., & Okubo, Y. (2006). Patterns of scientific collaboration between Japan and France: Inter-sectoral analysis using Probabilistic Partnership Index (PPI). *Scientometrics* , *68*, 303–324.

Zitt, M., Bassecoulard, E., & Okubo, Y. (2000). Shadows of the Past in International Cooperation: Collaboration Profiles of the Top Five Producers of Science. *Scientometrics* , *47* (3), 627-657.