Dynamic of Publication Network in German Photovoltaic Industry

Vasaf, Esmaeil and Sanatkhani, Mahboobeh

Friedrich-Schiller-University of Jena, Friedrich-Schiller-University of Jena

10 November 2014

ABSTRACT

Besides high policy-induced motivations for development of research activities in photovoltaic industry, there have been a few social network studies concentrating on the scientific publication in this field. This study tried to shed light on the structure and evolution of publication network in German PV industry from 1988 to 2013. For this purpose, using the centrality indices, I realized the most influential actors as potential source of knowledge and actors who play the central role in knowledge production and diffusion.

In next step, I investigated the dynamic of co-authorship network of scientists. Results showed that against the downward trend of network's cohesion, overall compared to the same size random generated network, German PV co-authorship network is characterized as a small world network which emphasizes the efficient diffusion of knowledge compare to other type of network.

Finally, to disclose the drivers behind the evolution of co-authorship network, I hypothesized two different scenarios. First, using descriptive analysis, the existence of preferential attachment mechanism is investigated. Fitting power law distribution over degree of nodes rejected our hypothesis for all investigating time windows. Therefore, preferential attachment mechanism cannot significantly explain the evolution of the network and reveals that network is robust in response to removal of large nodes. Second, looking at the composition of knowledge on map of science provided strong evidence in support of interdisciplinarity nature of German PV industry. Our descriptive analysis shows that along with existence of leading macro-disciplines such as Materials Science and Physics Applied, new subject categories of science have found a significant position over the existing knowledge domain during the observed period.

1. Introduction

1.1. Introduction and Background

Rising concerns about the environmental issues in one hand and the volatility of energy prices in recent decades in another hand brought a lot of attention to other source of energies. Due to accessibility of solar energy and easy installment of solar panels in the form of distributed power generation, the photovoltaic industry (PV) has been considered as a reliable environmentally benign source of energy. Despite the higher production cost of electricity by PV technology, large investments have been promoted and usually subsidized by governments to enhance the industry and firms to conduct research projects for the purpose of low production cost. It is noteworthy that, since 2001, the PV manufacturing has grown even more than all optimistic trend estimation leading to a significant reduction of production cost[1](2013). The situation for Germany is not different of the whole world direction (Wulf, Missner et al. 2010):

"Events like the United Nation's climate conference in Copenhagen in December 2009, which largely focused on CO2 reduction targets, make it clear that the renewable energy sector in general and the German photovoltaic industry in particular possess tremendous growth opportunities. Experts believe that due to the increasing global awareness for climate change on the one hand and decreasing prices for solar modules on the other, the photovoltaic industry will grow at an annual rate of 11 percent until 2020."

Evan though there are high policy-induced motivations for development of research activities in photovoltaic industry, there has been a few social network studies concentrating on the scientific publication in this field. Scientific production can be analyzed in different aspects. One may focus on the collaborative patterns of scientist: a group of scientist who make a network and share their own knowledge to develop and produce new knowledge. Also, it may be of interest to look at the institutions performance in scientific publishing; research institutes, universities and firms usually conduct joint research projects which connect scientists to each other. Moreover, co-authoring is a common form of scientific activities which can connect two or more institutions to each other. Therefore, the collaboration pattern of research institutions, universities and firms

[1]. www.irena.org/Publications

based on scientific papers can be analyzed in a social network framework. Hence, we can obviously observe the prominent actors in knowledge production as well as find who controls the transfer of knowledge. This actor initially can be a scientist who has a gate-keeper role and controls the import and export of knowledge into a community, institute, industry or even a country. In other view, an institution can obtain the role of knowledge transfer. In a social network perspective, a central actor can connect system (in its wide definition) to outside knowledge sources(Graf 2011). In this research I try to depict those interactions and shed light on the basic knowledge production side of German PV industry via a Social Network Analysis method.

1.2. Justification of Research Topic

Social Network Analysis is defined as a universal view of relationship between actors: "each relationship refers to a particular type of resource exchange. The actors who exchange these resources may be individuals, but also may be organization or institutions […] or concepts such as subjects linked in a hypertest document" (Haythornthwaite, 1998). In this context, the publication network can be defined as a relation between authors and institutions (considered as actors) or scientific field which constructs the knowledge based in a specific technological field. This research aims to analyze the society of scientist in PV industry based on the co-authorship network and investigate the evolution of social network of institutions linked to each other by means of scientists affiliated with them. The collaboration between scientists facilitates the interdisciplinary studies and thus they can utilize different source of knowledge as well as complementary fields. Therefore, it would be helpful to look at the relationship between scientific subject categories and scope of basic researches in PV industry.

2. Theoretical Background

2.1. Scientific collaboration, objectives and limitations

The idea of collaboration comes from a simple observation of human's social behavior. In scientific collaboration, there are numerous initiatives which gather scientists in a group or organization. Willingness to do interdisciplinary studies, lowering cost of experiments, desire for innovation and product development are the most important derivers for research collaboration. As a result, “frequent communication between collaborators is often associated with greater trust, increased output (i.e. scientific publications) and greater value for money” (Bellanca 2009). Collaborators share their resources to gain in the form of wealth or reputation.

In many countries, fostering collaboration (particularly in the form of university-industry collaboration) is becoming a significant part of policy intervention. There are two preliminary assumptions for analyzing such policy implementation. In one hand, it is explicitly assumed that all participants have the same understanding of the concept of collaboration. In other hand, if the collaboration is fostering by policy makers, we must be able to measure its strengths and weaknesses as well as its dynamic over years in response to policy intervention (Katz. and Martin 1997). These assumptions seem to be strong, at least in a straightforward measuring point of view.

To sum up, beside strong social policy motivation, scientific collaborations neither have the same initiatives nor the same expected effects in diffusion of knowledge. In current research, I concentrate on the dynamic of coauthoring as a consequence of scientific collaboration in one technological field. Knowing the limitations, we need a precise approach to give us a good picture of status quo and dynamic of cooperation between scientists and the institutions. For this purpose, we apply the Social Network Analysis (SNA) as a novel approach in economic analyzing.

A pairs of scientists who collaborate with each other and publish a paper base on their joint research, make a co-authorship link. In all scientific fields and technological research theme we observe such relationship and therefore, we can illustrate a co-authorship network.

Babliometrics ascertains that modern science has been faced by a fascinating expansion of collaboration between scientists in local environment (Melin and Persson 1996), at international level (Wagner and Leydesdorff 2005, Leydesdorff and Wagner 2008), or in a specific research field (González-Alcaide, Aleixandre-Benavent et al. 2008, Erfanmanesh., Rohani. et al. 2012). As Melin and Perssonl(1996) emphasize, the strength and expansion of scientific interactions change the concept of production functions in a way that we are experiencing an emergent form of *production unit*: “If more than half of the papers produced by the scientists at a given university are co-authored with scientists at other universities or research institutions it is no longer meaningful to talk about the university as a sole producer of knowledge. It is rather the network of interacting scientists that is the critical production unit". This definition of collaboration network has an important implication; we can explain the dynamic of institutional collaboration by looking at the underlying co-authorship network of scientists. In other words, the pattern of communication between scientists determines development path and growth of institutional network. In this context, theory has two alternative explanations for dynamic of co-authorship network: *preferential attachment* and *emergence of interdisciplinary sciences* which I discuss in following.

2.2. Preferential Attachment

There is a consensus that growth of scientific disciplines and diffusion of knowledge is crucially depend on the structure of its subordinated network (Newman 2001, Cowan and Jonard 2004). The theoretical basis of this notion is that science matures as a *collective effort*; new theories are rooted in the current theories, ideas and publications which diffuse in scientific communities by close collaboration of scholars. Therefore, it would be important to know what affect the structure of network and more important, what can explain the process of emerging scientific networks.

It is usually discussed in social network theory that nodes' degree distributions is associated with the image of the actors in the network and what he/she accumulated during the time. Put differently, the behavior of actors to join the network is not in a random way with equal probability for each node. In theory, this phenomenon is known as *preferential attachment*: "When [actors] choosing between two possible links, they will seek to connect to the more connected members. In other words, when someone is seeking a collaborator, they will seek someone who is already highly connected and therefore has access to resources and reputation" (Wagner, Roessner et al. 2011). In this mean, connection can be any features that give popularity to one specific node. In a publication network, a scholar might have several collaboration links or be a star reference in a subject of knowledge with many citations. Therefore, newcomers or isolate actors try to become attached to central actors. This process is widely known as preferential attachment which is proved that is responsible for evolving networks with power law degree distribution (Barabasi, Jeong et al. 2002). Back to the concept of *invisible college*, we can conclude that in a scientific network which follows power low distribution, more connections are belong to the *small world of scientists*. Therefore, in current research:

Structural change and growth of German PV scientific network can be characterized by the preferential attachment process.

2.3. Interdisciplinarity and Integration of Knowledge Fields

What characterized today's modern economies is surely the dramatic growth of sciences-based technologies. "What differentiates a modern knowledge-based economy from a more traditional one is the process by means of which the knowledge used in production processes is generated" (Krafft, Quatraro et al. 2011). In particular, importance of knowledge associated with the national production highlights the role of universities, research institutions and R&D companies. Rafols et al. (2010) explain that specialization traditionally was the dominant feature of modern universities which guarantied success of the scientific development. Yet this approach toward scientific development has an intensive change, enlarging the role of interdisciplinary in scientific research. They cite from the Lenoir, T (1997) that:

> "Scientists at the research front do not perceive their goal as expanding a discipline. Indeed most novel research, particularly in contemporary science, is not confined within the scope of a single discipline, but draws upon work of several disciplines. If asked, most scientists would say that they work on problems. Almost no one thinks of her- or himself as working on a discipline".

Nowadays, the revolutionary definition of science has a viable policy implication aims to foster the interdisciplinary studies. Therefore, it is important to set up accurate methodologies to picture the local and national *map of science*. The preliminary step toward this goal is to have an unambiguous definition of interdisciplinary. One most cited definition of interdisciplinary research is defined by Porter et al. (2006):

"Interdisciplinary research (IDR) is a mode of research by teams or individuals that **integrates**

- perspectives/concepts/theories **and/or**
- tools/techniques **and/or**
- information/data

from two or more bodies of specialized knowledge or research practice."

This definition put more weight on the diversity aspect of IDR and distinguishes it from the transitions in body of knowledge. Meanwhile the *integration* indicates that an IDR is cohesive with some degree of relatedness between subfields. The process of integration and constitution of IDR, as Wagner, et al. (2011) discuss, is more difficult to be followed. Instead, we can observe the result of this process as a published paper or scientific report. Tracing and disseminating the IDR in publication is mainly done using babliometric analysis, but using different methods. One of the most recent approaches which applied the advancements in *information visualization* techniques is called *map of science*.Boyack(2004) summarized all commonly approaches in mapping of science and their applications. In short, there are two mainstream: first applies the network of different units and construct the semantic network such as co-authorship network, co-word network, co-occurrence etc. Second, babliometric metrics directly quantify the diversity or distribution of categories of science. Current research follows the first approach and in following, I review the basic concepts of mapping of science.

A map of science is a snapshot of disciplines or scientific concepts in which each are positioned based on the cognitive proximity. Therefore, homogenous fields are locally linked to each other while heterogeneous disciplines are positioned in far distance. The procedure is similar to making co-authorship networks discussed in previous section but changes the definition of nodes and ties based on scientific categories and co-relational structure between them. In fact, we look at the co-occurrence of disciplines in publications and make the network based on the overlapping fields.

Constructing map of science enables us to investigate how interdisciplinary the field of interest is by tracing the size of nodes (disciplines) and strength of ties (overlaps). If the observed network turn to have a complex structure with strong relation between different scientific categories, then it can be analyzed as a driver for evolving co-authorship network.

The interdisciplinary characteristic of German PV industry is served to be a source of emergent co-authorship network.

3. Data and Methodologies

3.1. Data Source

This study concentrates on the publication data as a source of *basic research* inphotovoltaic industry. For this purpose we need to collect all related published papers. Also, we have to limit our publications origin to Germany boundary to capture the local scientific efforts and research activities only in German PV industry. For this reason, I run a Boolean search in the Thomson Reuters Web of Science (WoS) Core Collection. WoS Core Collection provides a world wild coverage of over 12000 highest impact journals from 1900 to now[2]. Each paper is well indexed and includes all information that introduced the content, relevance and origin of papers as well as the reliable classification in Science categories. Similar to all studies in co-authorship network, data preparation starts with searching items in WoS.

3.2. Identification of Publications Dataset

The best way to search for relevant publications in data bases like WoS is to apply the lexical keywords method. This approach is usually accompanied by some errors in which precision and recall is a challenging step before start running the search. Finding appropriate search terms can diminish this problem. In fact, some search terms capture a wide range of publication within scientific research fields which most of them may not be relevant. In this context, Porter et al. (2007) introduced three key criteria for the purpose of adjusting search terms:

First, a search term should encompass a sizeable amount of relevant papers. Second, the search keywords must cover the topic of research with acceptable transparency. Third, it should be possible to add, remove or modify search terms to cover the new areas of inspected knowledge fields or being elastic in evolving knowledge.

According to above criteria, I evaluated a group of most significant candidate search terms in WoS and manually looked at the output. The final list is as follows:

Query: (SO=("photo*voltaic*") OR TS=("solar cell*" AND electricit*) OR TS=("*photo*voltaic*" OR "PV cell*" OR "pv module*" OR "concentrating pv" OR "pv panel*") OR TS=("solar module*" AND (electricit*)) OR TS=("*crystalline silicon" AND pv) OR TS=("thin* film" AND pv))

The output dataset covers all publications from 1989 to 2013. To be able to compare yearly changes of babliometric statistics and construct our co-authorship network, I excludedthe four first months of year 2014. Further refining conducted to “Germany” territory, “Article” publication typology and “German OR English” languages.

[2]. Reference: http://thomsonreuters.com/web-of-science-core-collection/

3.3. Preparing and Cleaning Strategies

A co-authorship network is constructed based on actors (institutions and authors) as nodes and their relationships. Each publication in WoS includes useful information such as name of authors, research fields, official address, publication year etc. Each type of co-authorship network requires different information set. For example, collaboration network of scientists is made using authors' name. Regional co-authorship network is highly relied on the address of authors. Therefore, the objective of analysis determines type of information that would be extracted from the dataset. However, there are serious limitations in working with publication data and we need to overcome them before starting our statistical and network analysis. A quick search for specific writer reveals that a single author is appeared with different names in his papers. For example, Brabec C J may reports BrabecChristoph J., Christoph J Brabec or CJ Brabec. But, in a specific technological field such as PV, it is more likely that all of them refer to one person. Also, there may be two authors with exact same names in our dataset. Neglecting this problem may cause our network to become more fragmented or extremely denser.

This research uses the Bibexcel toolbox (Persson, Danell et al. 2009), a strong babliometric program developed by OllePersson. Bibexcel directly accepts text files downloaded from WoS. The output files can be read by MS. Excel or any text reading softwares. Also, we can use R programming tool for more data manipulation along with the importing them directly into the well-known networking packages such as Pajek[3] or UCINET[4]. The starting point in current research is extracting authors' name and institutions attached by the year of publication. Cleaning the names is done by Bibexcel. At first, it converts upper case to lower (for example BRABEC CJ to Brabec CJ) and remove hyphens, periods, etc. from the author's name. Second, it removes extra characters from the right side of names. Final list can be saved in Excel or txt filesto be used for the network analysis in Pajek or any other softwares.

Working with institutions names is associated with the same problem as authors'. However, some writers put the German name of their affiliated institutions on the papers or use different versions of one single institution. Therefore, before working with network, we must clean the data precisely. The procedure that I applied is as follows:

---

[3]. http://pajek.imfm.si/doku.php
[4]. https://sites.google.com/site/ucinetsoftware/home

1. Extracting the institution names in one column while attached a number (as code) of each paper in second column
2. Lowering the case of names
3. Sorting the names alphabetically
4. Finding similar names as well as the accordant German names
5. Searching in the internet for most frequent institutions names and making a table for different version of single names
6. Replacing the longer names (from step 5) with the shortest one
7. Searching again for some frequent keywords (such as Max Planck or Fraunhofer) in output list
8. Repeat step 6
9. Final cleaned list of institution names

Now, we can use the completed list for further statistical analysis or making the collaboration network. In addition, I categorized the institutions based on their organizational type to Universities, Research institutions and others (i.e. firms).

3.4. Methodology

Besides babliometric analysis of the publication data set, constructing and analyzing the co-authorship networks is the main goal of this research. Following the data preparation, I used R programming language (Network and SNA Packages (Butts 2008)) for calculating network statistics. Next, I appliedPajek(Batagelj and Mrvar 2004) to illustrate the International co-authorship network and co-authorship network of Institutions and Scientists separately.

3.4.1. Power-Law Distributions

It is discussed in literature that some events such as collaboration patternsmay follow the preferential attachment mechanism. Therefore, the subsequent network may turn out to be the scale free network and the authors' degree distribution follows the power law. In continue, I review the concept of power law distribution and the statistical estimation of its parameters.
Statistically, a random variable x follows power law if we observe a probability distribution which satisfies the following statement:

$$p(x) \sim Cx^{-\beta}$$

where$p(x)$ is probability distribution of x, C is distribution's constant parameter and β characterizing*exponent* or *scaling parameter*. The exponent parameter typically lies in the range of $2 < \beta < 3$ for a random variable following power law distribution (Clauset, Shalizi et al. 2009). Clauset, et al. (2009) investigate that in real empirical studies, power law phenomena only satisfied with those x more than a threshold, x_{min} and we roughly say that "the tail of the distribution follows a power law". Therefore, we need a precise estimation of x_{min} to avoid miss specification of interested distribution and exponent parameter β. For too small and too large assumed quantity of x_{min}, we might fit a distribution on those part of data which do not really obey power laws and consequently calculate a wrong β. What we need is to calculate the lower-boundary of the interested variable x in our empirical data set. Then, it is possible to estimate the exponent parameter β. Clauset, et al.,(2009) proved that the best method to fit the power-law distribution on empirical distribution such that $x \geq x_{min}$, is the Maximum Likelihood Estimation (MLE). The ML estimator of β for *continues* datasets is:

$$\hat{\beta} = 1 + n \left[\sum_{i=1}^{n} ln \frac{x_i}{x_{min}} \right]^{-1} ,$$

assuming that our data is $x_1, \dots, x_n$ for all $x \geq x_{min}$.

In this research, the estimation of parameters and following statistical tests will be done using the poweRlaw package (Gillespie 2014) in R programming language.

3.4.2. Overlay Map of Science

An alternative explanation for the growth of co-authorship networks is the interdisciplinary needs for the existing technological field. In this manner, actors (scholars) generate new connections with others from outside of their *small world* of collaborators. The subsequent change is developing new ties between older disciplines and new fields of knowledge, a new structure that we discussed to be an interdisciplinary research. I am going to describe the structure of science in German PV industry to investigate whether it proposes such multidisciplinary approach in its publication network. The procedure that I apply here is the Overlay Map of Science introduced by Rafols et al. (2010):

“In addition to capturing disciplinary diversity, they [overlay maps] can also help to clarify the relative location of disciplines and thereby enable us to gain insights of another of the aspects of interdisciplinary research, namely their position in between or central (or marginal) to other research areas”.

According to them, the procedure to draw an overlay map can be summarized as follows:

1. At the ANALYZE tab of web of science for the selected papers, export the list of Web of Science Categories.
2. Run the Windows DOS mini-program WC10.exe and create a .vec file to be uploaded into Pajek. This vector includes weights different science categories.
3. Upload Pajek project file which would be used as a base map to visualize the interactions between science categories.
4. Draw the network in Pajek and set the size of vertices and colors for cluster mapping.

The output map illustrates the interactions between subfields and the position of each in the current technological field. The distance between nodes and the extent of links (co-occurring in scientific papers) can be a fortified base for identifying potential complementary collaboration.

4. Arguments and results

The purpose of current analysis is to understand the structure of collaboration network and its evolution over years. At first step, I will review the general statistics of publication data in the field. Using network metrics and network illustration, in second section, I try to shed light on the role of distinguished institutions in developing the inspected field of science in Germany. Section three constructs my particular investigation of network dynamics and characterization by means of co-authorship network of scientists. Following that, I propose my finding about the interdisciplinary nature of PV industry, before I conclude in last section.

4.1. General statistics

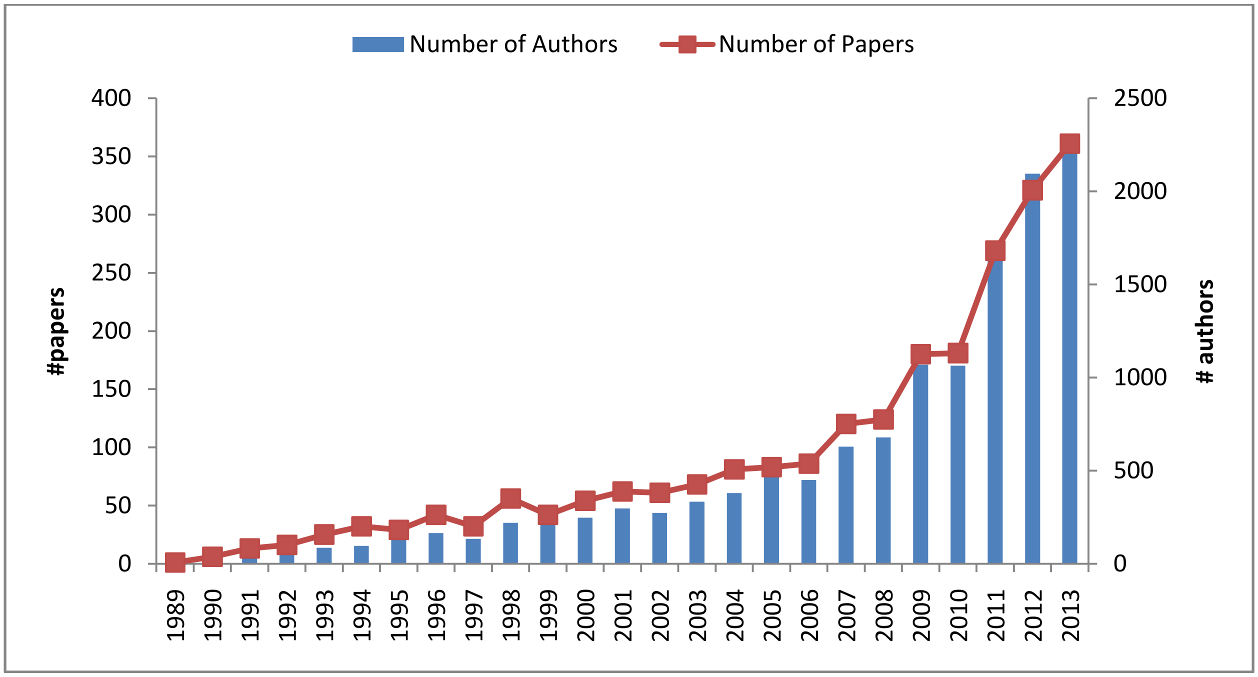
Running the search query in Web of Science (WoS) Core Collection using specified search terms for German photovoltaic industry, resulted totally 2345 papers for the period of 1989 to 2013. Table 1 shows the number of publications, number of authors (see figure 1) and average number of authors per paper. We see that there is only one paper for starting year and 361 papers for last year. Average number of authors per paper is calculated by taking the average number of collaborators in each paper in each year and represents a range of three to (more than) six during the considering period. The rising number of collaborators in each paper, as we discussed in the theoretical background section, can be either a sign of increasing clustering among scientists in the network or rising tendency for interdisciplinary studies. We will back to these two hypothesis later in our social network analysis and science map of knowledge and look at it with more details.

Table 1: General statistics in German PV industry publications (1989-2013)

Year	Number of Papers	Number of Authors	Average number of Author per Paper
1989	1	1	1.0
1990	6	19	3.2
1991	13	38	2.9
1992	16	75	4.7
1993	25	86	3.4
1994	32	95	3.0
1995	29	139	4.8
1996	42	165	3.9
1997	32	133	4.2
1998	56	220	3.9
1999	42	215	5.1
2000	54	246	4.6
2001	62	297	4.8
2002	61	272	4.5
2003	68	332	4.9
2004	81	379	4.7
2005	83	472	5.7
2006	86	450	5.2
2007	120	628	5.2
2008	124	678	5.5
2009	180	1068	5.9
2010	181	1064	5.9
2011	269	1628	6.1
2012	321	2094	6.5
2013	361	2271	6.3

Source: Web of Science after cleaning names and own calculations

Figure 1: Annual number of articles and authors (1989-2013)



Source: Web of Science

Extracting the "address" field from the publication data enables us to carefully explore the main publishing institutions in basic research of PV industry. It is not surprising that most productive institutions in our network are German based universities and research institutes. Table 2 shows that in considering period, the "Fraunhofer Institute Wind Energy & Energy System Technology IWES[5]" locates at top of our list with 136 papers and "University of Erlangen Nurnberg" with 120 papers is the second most productive institutes in German PV industry. Special case in our observation is the Helmholtz Zentrum Berlin (HZB) which established on January 2009 by joint of "Hahn Meitner Institute (HMI)" and "Berliner Elektronenspeicherring-GesellschaftfürSynchrotronstrahlung"[6]. Hence, our data for HMI only covers the publications to the end of 2008 and subsequent data for HZB just includes publications after January 2009. Therefore, the sum of all publications by HMI and HZB results the highest number of papers among institutions during the whole period.

[5]. Fraunhofer IWST was established on 2009 from the former Fraunhofer Center for Wind Energy and Maritime Engineering CWMT and the Institute for Solar Energy Supply Technology ISET.
Reference: http://www.kassel.de/wirtschaft/institutionen/infos/12159/
[6]. reference: http://en.wikipedia.org/wiki/Helmholtz-Zentrum_Berlin

Table 2: High productive institutions based on authors' affiliation (1989-2013)

Institution name	# of times appeared in separate papers
Fraunhofer Institute Wind Energy & Energy System Technology (the former ISET)	136
University of Erlangen Nurnberg	120
Hahn Meitner Institute (HMI)	112
Max Planck Inst Polymer Research	110
University Stuttgart	94
FraunhoferInst Solar Energy Syst ISE	90
Technical University Dresden	87
ForschungszentrumJulich	73
University Oldenburg	72
Karlsruhe Institute of Technology (KIT)	69
Helmholtz Zentrum Berlin (HZB)	65
University of Jena	63
University of Freiburg	60
University of Wurzburg	57
University of ULM	54
University of Munich	51
Technical University of Ilmenau	50
Technical University of Munich	50
University of Bayreuth	49
ZAE Bayern	47

Source: Web of Science after cleaning addresses

4.2. Co-Authorship Network of Institutions

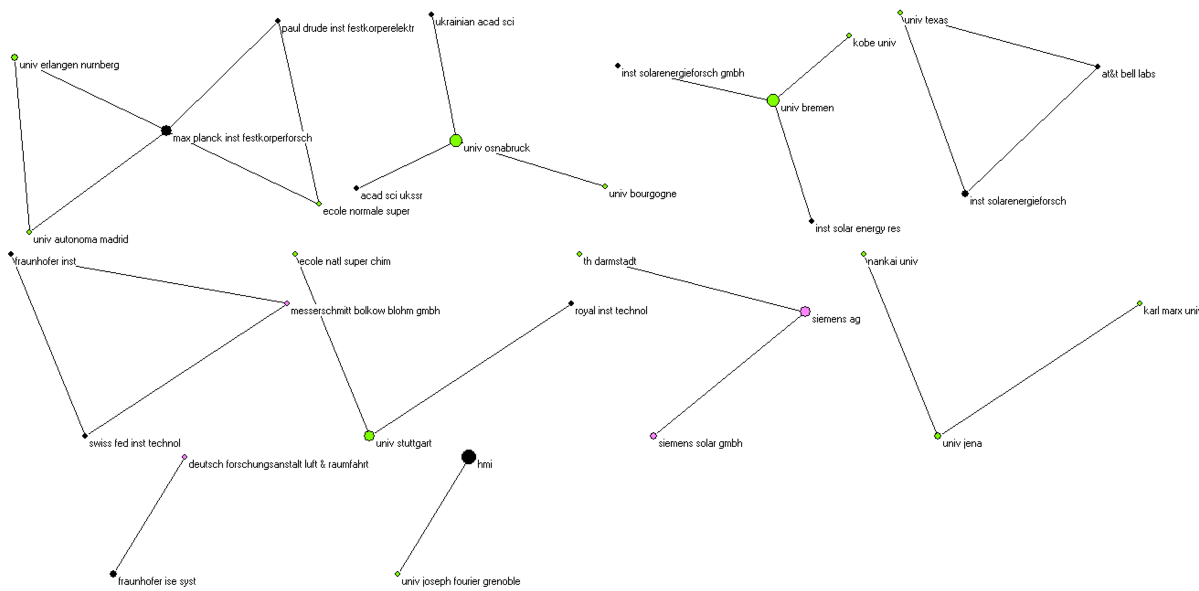
To analyze the collaborative research in science, we can construct network either based on the authors or institutions. Even though the motivation for joint authoring between scientists is affected by social and cognitive proximities, the institutional collaboration can be induced by national or organizational strategies. National research foundations such as universities and institutions try to enter into new field of knowledge by conducting official cooperation with other universities. Furthermore, they aim to share the excessive cost of fundamental research and their existing research labs. Companies may follow different objectives. They can overwhelm new industry challenges and penetrate to new markets by engaging in research activities with universities and research institutions. They may also share their R&D resources with other companies looking for innovations and taking advantage of external R&D human resources.

In this section, I try to analyze such interactions between institutions. First of all, I classified institutions in three different groups: universities, research institutions and firms. These are three important layers of organizational research units in Germany. Universities are involved in basic science, companies are interested in more purely applied research, and research institutes located in the middle of two other groups and usually have collaboration with both of them. The co-authorship network is constructed using the reported address of scientists on papers. Two organizations are collaborated if there is at least a joint publication whose writers are affiliated with them. Number of times that each institute appears in their address field, determines size of nodes and the frequency of co-publications (co-occurrence) between different institutions denotes the size of ties or simply, how strongly they are connected.

To analyze the dynamic of publication network in German PV industry, I represent the growing network in different time windows. Figure 2 illustrates the infant network of collaboration during the years 1989 to 1993. Research institutions are denoted in black and universities in green nodes while firms pictured in purple. We immediately, find that universities and research institutions were strongly involved in formation of the network. The exceptions are Siemens and its sub-company that had a few publications jointly with technical university of Darmstadt. More important, the initial attempts in this field were done in corporation with the foreign institutions. In particular, "Max Planck Institute of Solid state research (Festkorperforsch)" had a link with two foreign universities from Spain and France. Same collaborations were launched by university of Stuttgart, University of Bremen and university of Osnabruck. At the bottom of the picture we can see the "HMI" which had a weak link with university of Joseph Fourier in France, although it had more publications compare to other actors.

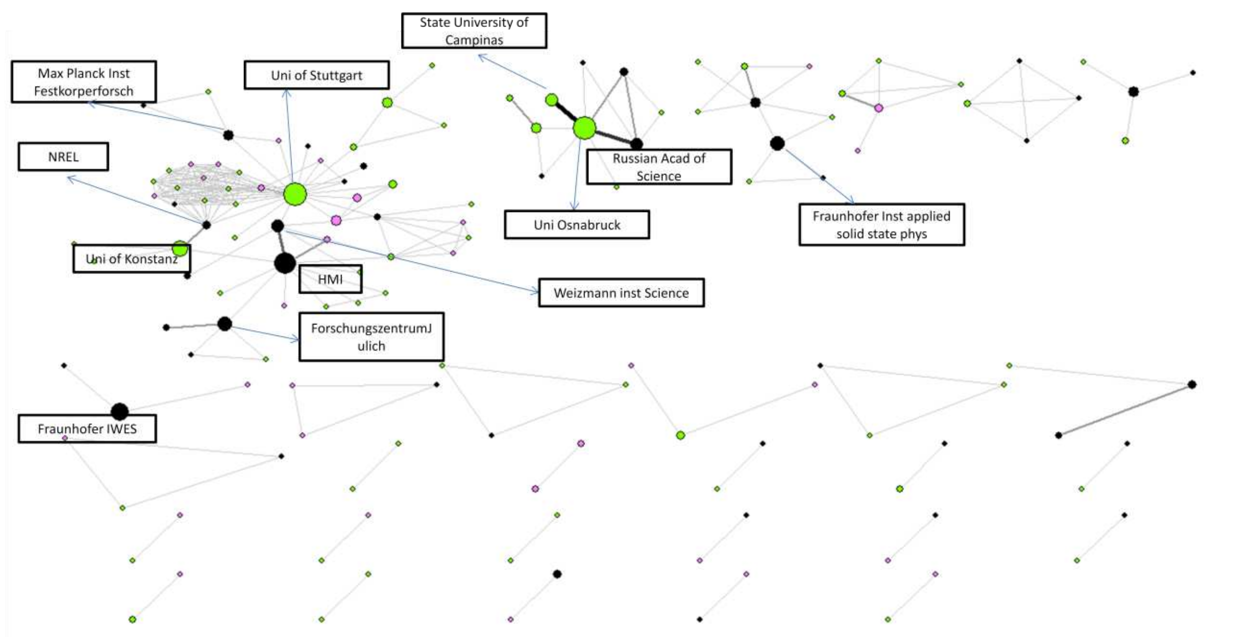
The co-authorship network of institutions grew in next period from 1994 to 1998 and we can easily observe the structured main component of network. University of Stuttgart, University of Konstanz and HMI were the most productive actors within the main component witch connected other actors to each other. We can see the National Renewable Energy Laboratory (NREL) of USA that worked as a cut point or bridge in main component at the left side of network which had a strong tie with university of Konstanz and a cluster of less publishing institutes in the network. University of Osnabruck was in the core of second largest component with strong relation with several international universities. The interesting result is that Fraunhofer institutes with its branches at right and left sides of picture had not specific strong relations with other institutions and did not integrate into the main component of network during that period which supposed to be the result of official policy or individual preferences of working scientists to had inter-organizational joint research during that period.

Figure 2: Co-authorship network of institutions in German PV industry (1989-1993)



Produced by Pajek- All institutions included
Black nodes: research institutions- Green nodes: Universities- Purple nodes: Others

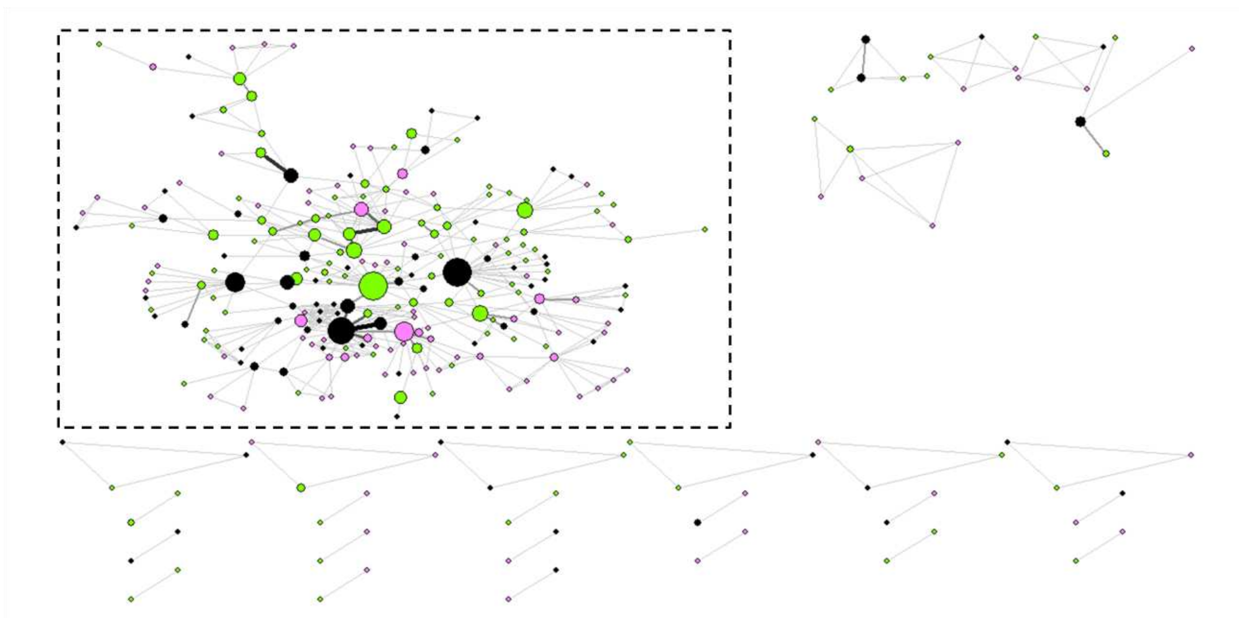
Figure 3: Co-authorship network of institutions in German PV industry (1994-1998)

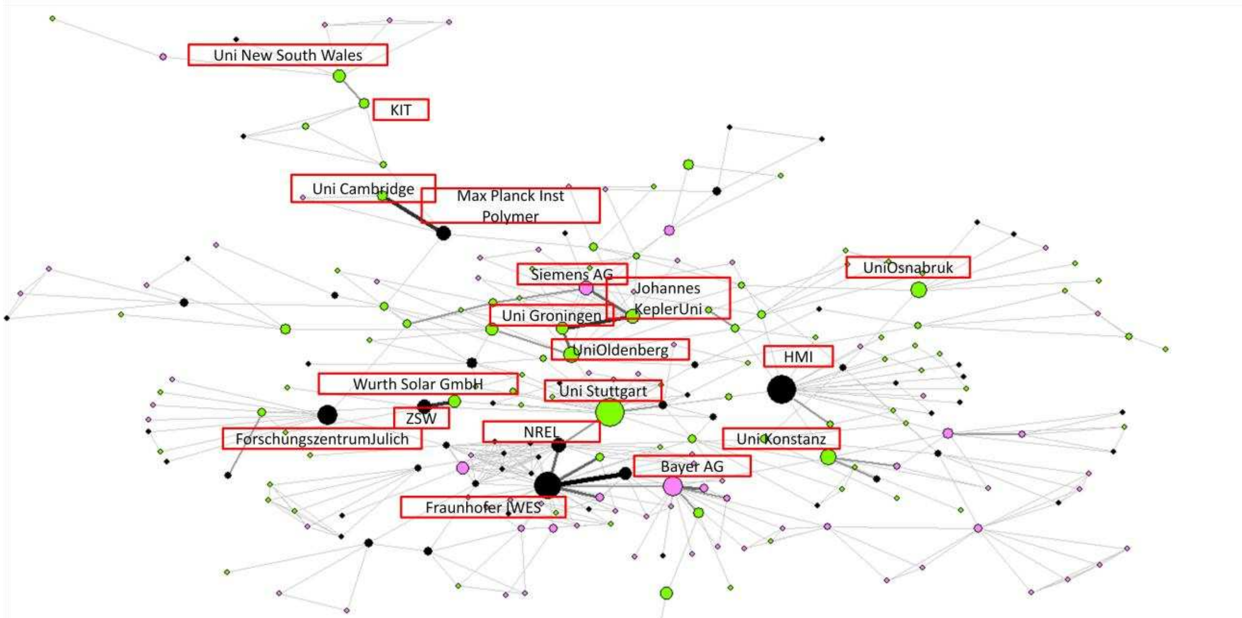


Compare to previous time window, the institution network for period 1999 to 2003, became denser with a giant component encompassing most of the active important nodes in the field. Concentrating on the main component subgraph, reveals that Fraunhofer IWES, at the bottom of picture, turned to find a prominent role in the network and developed many connections with a range of research institutions (black nodes) and firms (purple nodes). Moreover, Fraunhofer IWES and its collaborators are linked to the main component by means of NREL of USA. However, we can see the Max Planck institute at top of picture with limited connections to other institutions (Notice that there was a strong tie with the Cambridge University on that period).

To sum up, principle characteristic of the institutions network in considering period is the existence of several large nodes which possessed the role of transmitting knowledge in the network. Most of these universities and research institutions deserve to be considered as bridges in the network which at the same time controlled the knowledge diffusion.

Figure 4: Co-authorship network of institutions in German PV industry (1999-2003)





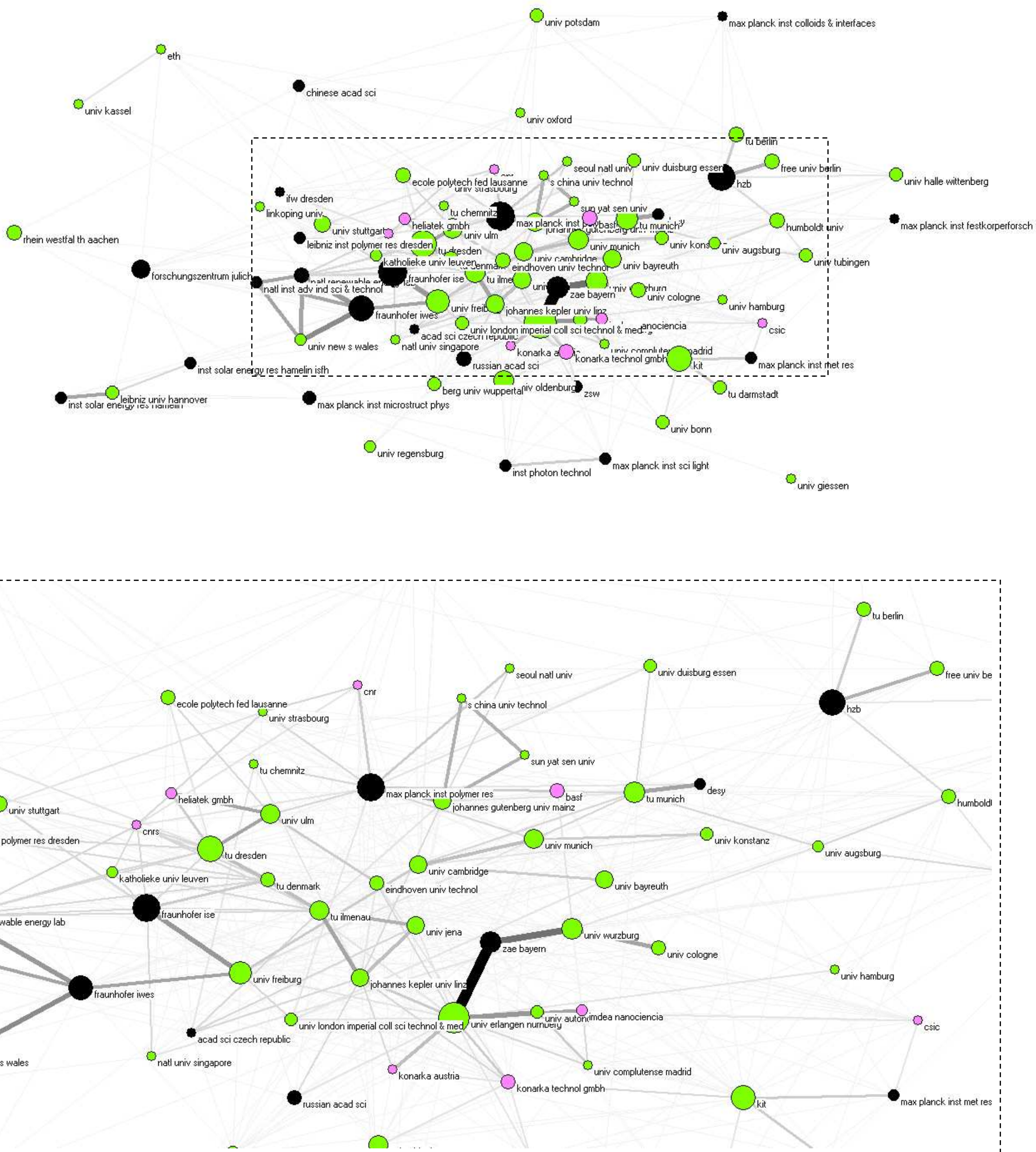
In the last period[7], starting 2009 to 2013, we observed the dominant position of national universities including technical universities in the network. Again, research institutions such as Max Planck institute of polymer research, Fraunhofer IWES, Fraunhofer ISE and Helmholtz Zentrum Berlin (the former HMI) are the most productive research institutions. Fraunhofer IWES has strong linkages with University New South Wales and two American research institutions highlighting the organizational and personal tendency to have collaboration with foreign institutes. At the center of the network we can see a noticeable linkage between ZAE Bayern and university of Erlangen Nurnberg. Looking at the website of this institute shows that ZAE Bayern comprises three divisions. The division of “Photovoltaics and Thermosensorics", is located in Erlangen which headed by “Prof. Dr. Christoph J. Brabec” from the university of Erlangen Nurnberg together with “Richard Auer (Dipl.-Ing.)” from ZAE Bayern[8]. Therefore such a strong collaboration created between these two institutes in the network.

Because of the complexity of the network and difficulty to analyze the exact position of each node in scientific research, and therefore transmit of knowledge, in next step I propose a detailed explanation using network metrics. The objective is to have a better understanding about current scientific collaboration network of German PV industry.

[7]. I skip here the middle period (2004-2008).

[8]. reference: http://www.zae-bayern.de/en/the-zae-bayern/chartered-objectives.html

Figure 4: Co-authorship network of institutions in German PV industry (2009-2013)



Produced by Pajek- For simplicity, institutions with at least 8 papers in the considering period (83 nodes among total 1027 nodes) are displayed.

4.2.1. Degree of Nodes in Institutions' Network

Degree of node is defined as the number of immediate neighbors of that node. One node with higher degree has more close connection (collaboration) in the network and can be considered as potential resources of knowledge for a node in the network. Table 2 shows the top institutions in the co-authorship network of institutions during the years 2009 to 2013. University of Erlangen Nurnberg has more collaborators in the network by 103 co-authorship connections, followed by Max Planck Institute of Polymer Research and Fraunhofer institute of polymer research with degree equal to 92 and 90, respectively. Sixteen institutions among the top 20 high degree actors are German ones and other are originated from other countries. Most central foreign institute is National Renewable Energy Lab of USA with 62 co-authors in German scientific network of PV industry.

Table 2: Degree of institutions in co-authorship network (2009-2013)

Institution	Degree
University of Erlangen Nurnberg	103
Max Planck Institute of Polymer Research	92
Fraunhofer ISE	90
Technical University of Dresden	87
Helmholtz Zentrum Berlin (HZB)	81
Technical University of Ilmenau	76
Karlsruhe Institute of Technology (KIT)	71
National Renewable Energy Lab	62
Technical University of Denmark	61
University of Freiburg	60
University of Jena	57
Fraunhofer IWES	56
University Of Wurzburg	52
Katholieke University Leuven	51
Johannes Kepler University of Linz	49
ZAE Bayern	48
University of Cambridge	46
Konarka Technology GmbH	46
University of Oldenburg	44
Berg University of Wuppertal	42

Calculated usingPajek

4.2.2. Closeness Centrality in Institutions Network

This measure identifies the speed of conveying knowledge between nodes and increases by decreasing distance relative to other nodes. Therefore, higher closeness centrality guaranties the faster transmission of knowledge to other actors and from whole network to that specific node. Calculation output presented in table 3 shows that University of Erlangen Nurnberg and Karlsruhe Institute of Technology (KIT) possess the most central role based on the average accessibility to other actors. Also, we see that Max Planck institute of Polymer research and Fraunhofer ISE are among the most central research institutions in PV industry in which other institutions can quickly access them in the network. More interesting is emergent of the Chinese Academy of Science in the list of top central institutes (closeness centrality) in the network of institutions network.

Table 3: Closeness Centrality for institutions network (2009-2013)

Institution	Closeness Centrality
University of Erlangen Nurnberg	0.38006
Karlsruhe Institute of Technology (KIT)	0.37900
University of Freiburg	0.37760
Technical University of Ilmenau	0.37329
Fraunhofer ISE	0.36743
Technical University of Dresden	0.36710
Max Planck Institute of Polymer Research	0.36677
Konarka Technology GmbH	0.35499
Helmholtz Zentrum Berlin (HZB)	0.35422
University of Cambridge	0.35103
University of Munich	0.35027
University of Jena	0.34982
University Of Wurzburg	0.34923
Fraunhofer IWES	0.34923
Katholieke University Leuven	0.34804
Berg University of Wuppertal	0.34700
National Renewable Energy Lab	0.34539
University of Stuttgart	0.34525
Chinese Academy of Science	0.34452
Imperial College London	0.34336

Calculation with Pajek for 20 first top institutions

4.2.3. Betweenness Centrality in Institutions Network

If we want to find how likely it is that a node be located among a communication path between other pairs of nodes (or the shortest path between them), we must focus on the betweenness centrality measure. Comparing the top 20 list in table 4 with two above tables reveals that there are some new institutions in the list while some others disappeared. In particular, different to previous lists, we see that the most central actors based on betweenness centrality are German based institutions (19 among 20 first institutes where University of Cambridge is located at last position). Back to the figure 4, it turned that these institutes usually connect sub-graphs to each other and thus to the giant component of network. Therefore, mostly they are German institutions that have the role of bridges in the network and ease the transmission of knowledge in the field.

Table 4: Betweenness Centrality measures for Institutions network (2009-2013)

Institution	Betweenness Centrality
University of Erlangen Nurnberg	0.086
Max Planck Institute of Polymer Research	0.084
Fraunhofer ISE	0.083
Helmholtz Zentrum Berlin (HZB)	0.082
Karlsruhe Institute of Technology (KIT)	0.064
Technical University of Dresden	0.061
Fraunhofer IWES	0.049
University of Freiburg	0.048
University of Jena	0.044
Technical University of Ilmenau	0.039
University of Oldenburg	0.039
ForschungszentrumJulich	0.032
University Bayreuth	0.031
University Munich	0.030
Free University of Berlin	0.030
Technical University of Munich	0.030
University Of Wurzburg	0.029
Technical University of Berlin	0.028
Johannes Gutenberg University of Mainz	0.025
University of Cambridge	0.025

Calculation usingPajek for 20 first institutions

Comparing the collaboration network of institutes in Germany with other countries apparently give a better understanding about the structural characteristics of network. However, current narrow study does not aim to evaluate other countries'. Instead we refer to the study of Guo et al.,(2009) which analyzes the "nanotechnology enhanced thin-film solar cells". Neglecting the difference in technological domain, the comparison of the local institutions network in US and Germany exhibits completely distinct structures. During the years 2001 to 2006, the US institutions network possessed some main actors with slightly different number of publications but with fragmented cooperation. In other words, there was not a specific leading actor in the network of "nanotechnology thin-film solar cells in US" along with a weak interconnection between institutes in scientific publishing. In contrast, for Germany the HMI institute had a leading position in the network. At the same time, all important actors are locally interconnected and reveal an emergent network in the "area of nano thin-film solar cells research in Germany". Our analysis for whole photovoltaic research field approves the great advantages of major research institutes and universities in the knowledge diffusion and development of German PV science.

To conclude, we discussed that the institutional collaboration network in PV industry is growing with some strong interactions between prominent institutes. We observed that the performance of different institutes have changed dramatically over years. Yet, we have not analyzed the reasons that contribute to the formation and evolution of investigating scientific network. It is discussed that the position of an institution in the network is widely affected by the collaborative works and reputation (popularity) of individuals affiliated there. In other words, changes in the situation of an institute within a network (measured by degree centrality, closeness centrality and betweenness centrality) can be explained by the collaboration network of its scientists. Therefore, in the context of current research it is necessary to look at the appearance of co-authoring linkages among scientists. We discussed in theory section that growth of network can be either the result of preferential attachment phenomenon or the necessity for developing new streams in existing knowledge base or revelation of the interdisciplinary. In following, at first step I try to shed light on the evolution of co-authorship network of scientists. At next step, I apply a network visualization method to investigate the complex of knowledge categories (global map of knowledge) in German PV industry.

4.3. Co-authorship Network of Scientists

4.3.1. Structural Characteristics

One important question in social network analysis is that how the structure of network influence the diffusion of knowledge. In this context, I follow Cowen and Jonard's explanation (Cowan and Jonard 2004) in which "knowledge diffuses through barter exchange among pairs of agents, and aggregate performance is measured as the mean knowledge level over all agents". Put differently, scientists share their idea at their local environment (clusters) in the network with face-to-face contacts. Therefore, we hope to find a pattern of relationship between scientists in which choosing a collaborator among all existing candidates is not a randomly drawn mechanism. We observed in previous section that co-authorship network of institutions is growing in a way that new actors attach to main component of network. Also, prominent universities and research institutes revealed higher involvement in knowledge diffusion in the network by collaborating with each other. However, we did not discuss whether the underlying network is efficiently working in diffusion of knowledge or not. Network analysts agree that structure of network in the form of small world can facilitate transmission of knowledge (Newman 2001, Cowan and Jonard 2004). There is a strong policy implication resulting from a small world network observation: If a network turned to be small world network, the knowledge spillovers happen in a more efficient way and the whole network would be less exposed to the removal of star scientists (Fleming, King et al. 2007).

In social network analysis, there are two critical measurements that help to determine the structure of network: mean shortest path (average distance) that shows how close authors are in the network; and clustering coefficient that specifies the likelihood of collaboration among two authors, when they have a collaborator in common. A network is characterized as small world network if its clustering coefficient is higher than the corresponding random network while simultaneously exhibits low shortest path length similar to random network (Cowan and Jonard 2004). These two attributes implies that scientists besides the actors in the vicinity, sometime look for in far distance within the network to find a prospective research collaborator with fresh ideas. At the same time, more clustered collaborations bring trust to the group and promotes transmission of knowledge in the local environment usually with face-to-face contacts. Therefore, this structure hopefully is a "good sign for science" (Newman 2001).

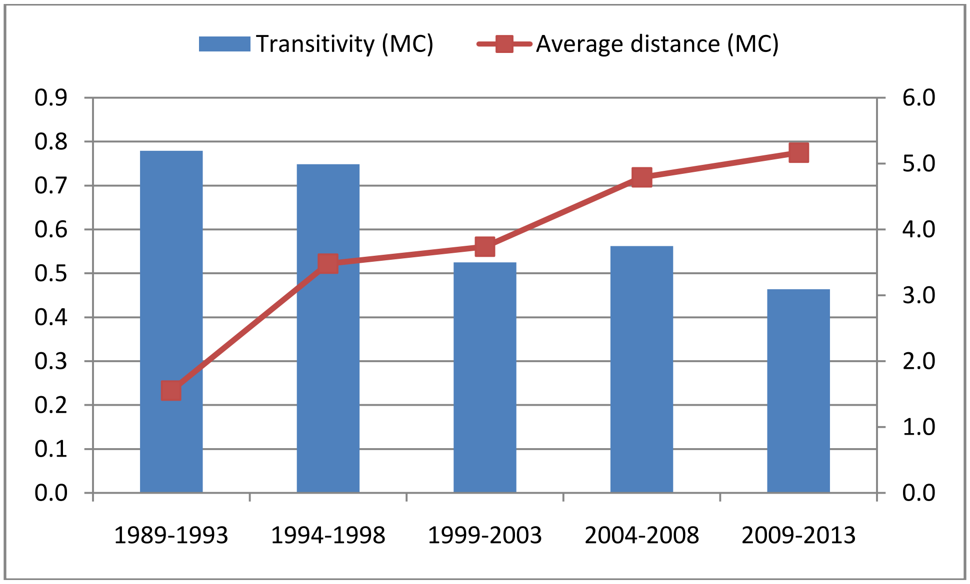
Table 5: Structural measures of co-authorship network

Measure	1989-1993	1994-1998	1999-2003	2004-2008	2009-2013
Number of Nodes	171	557	962	1587	4712
Number of Edges	371	1561	3084	6717	23761
Network Density	0.0255	0.0101	0.0067	0.0053	0.0021
Size of Main Component (MC)	15	107	165	859	3395
Share of Main component (%)	8.8	19.2	17.2	54.1	72.1
Share of Isolates (%)	5.3	2.7	1.2	0.7	0.3
Average distance (MC)	1.556	3.484	3.740	4.793	5.167
Transitivity (MC)	0.779	0.748	0.525	0.562	0.464
Average distance in Random network* (MC)	3.607	3.851	3.897	3.684	3.908
Transitivity in Random network (MC)	0.020	0.014	0.008	0.005	0.002

Calculated using Network, SNA, igraph and intergraph packages in R

*Random graph is generated with the same size and edges of original network.

Figure 5: Average distance and clustering coefficient in main component of co-authorship network



To statistically analyze the structure of co-authorship network of scientists, at first step I generate a random graph with the same size of the empirical network using "igraph" package (Csardi and Nepusz 2006) in R and convert it to network format using "intergraph" package (Bojanowski 2013). To be comparable, I set the size and number of edges exactly identical to original network. Moreover, because the shortest path length only can be calculated for connected network, following Fleming et al., (2007) and Andrade et al., (2014) I extract the main component for both networks and compare the small world measures of them. This method ensures that there is at least one path between two scholars within the network under investigation.

Table 5 presents two cohesion measures as well as some other statistics for co-authorship network of scientists in different 5-year time windows. Notice that the clustering coefficient is calculated based on the transitivity definition introduced in(Newman 2001). We see that along with increasing number of actors and connections in the network, the main component grew dramatically. Largest component at first period only encompassed about 9 percent of total scientists but at the last period more than 72 percent of all scholars were attached to giant component. Although the network grew quickly during the investigating period, it becomes less dense denoting that the rate of creating collaboration links was lower than rate of entering new scholars into the network.

Figure 5 illustrates that the co-authorship network became less cohesive over years, in which the clustering coefficient decreased and the average distance slightly increased. The static result for clustering coefficient in last period is compatible with the study of Newman[9](2001). In a different study with US patent data, Fleming et al.,(2007) report a decreasing shortest path length and increasing clustering coefficient. One possible explanation for such decreasing clustering coefficients in our publication network is that (as we observed in the institution network) a majority of high productive authors in the network are affiliated with main institutions and universities who are responsible for official research projects in the form of supervisors or head of research teams. Therefore, it may affect new triad collaboration formation at lower hierarchy of research teams and raise the role of bridges in the network. However, comparing the cohesion measurements with those of the generated random network reveals that the co-authorship network of scientists can be precisely expressed as a small world network with almost similar mean shortest path and substantially higher clustering coefficients.

Co-authorship network for the period 2009 to 2013 is illustrated in figure 6. It shows that the collaboration network is constructed by numerous distributed clusters each of them highlighted by two or three prominent authors in the core and severalones in the periphery. Existing strong linkages among some authors emphasizes that there are kind of hierarchical relationship between scientists in each cluster (as a sample, I highlighted three strongest linkages in the network with their corresponding affiliations).

To sum up, our co-authoring data proves that the whole network resembles the most efficient type of social networks in diffusion of knowledge. Even though the overall evolution path tended to produce a less cohesive network with higher mean distance and less clustering coefficient, as Cowan and Jonard(2004) emphasize, “there is an identifiable region of the space of structures in which diffusion is much more complete than elsewhere.” In current research I do not aim to identify such region for PV industry, instead I tried to picture the evolution of scientific network structure.

Next objective is to explore the drivers of such development in co-authoring network. We observed that most of the publications in the field are accompanied by co-authoring with prominent actors (scholars with more publications). This behavior is usually addressed in theory as preferential attachment and is defined to be responsible for network evolution, the hypothesis that I try to examine in next section.

[9]. Notice that time period, field of knowledge and size of the network are not the same.

Figure 6: Co-Authorship Network of scientists (2009-20013)

Produced by Pajek- Only nodes with at least two papers are displayed. Highlighted nodes are affiliated with corresponding institutes.

4.3.2. Power Law distribution

In this section, I investigate the existence of preferential attachment in the co-authorship network in PV industry. Looking forward for new resources and state of the art knowledge are the driving power for scholars to attach to those with high reputation. This process is responsible for development of social networks and is empirically examined in many studies.

The authors' degree centrality is the critical variable in our estimation and I try to fit a model over its distribution. Knowing that the degree distribution of nodes is statistically a discrete distribution, I use following approximation for estimation exponent parameter of model:

$$\hat{\beta} \approx 1 + n\left[\sum_{i=1}^{n} ln \frac{x_i}{x_{min} - \frac{1}{2}}\right]^{-1}$$

Applying MLE estimation and fitting the power law distribution on the data set for degree distribution of nodes, yields $\hat{\beta}$ and x_{min} parameters reported in table 6. First period in our data set includes few observations and thus is excluded from the analysis. We explained that for power law distributed quantities, the exponent parameter lies between 2 and 3 ($2 < \beta < 3$). The estimated x_{min} guaranties that the estimation is done for those quantities that best suited the power law distribution. Our estimation reveals that the preferential attachment mechanism seems to be held only for last period 2009 to 2013 and for other period, the estimated parameters are out of accepted range. Figure 7 illustrates Cumulative Distribution Function (CDF) of nodes' degrees during years from 2009 to 2013 and red line is the best fit of model. To summarize, for last period, the number of collaboration that a scholar received may be depends on the visibility of that node in the science network. In our co-authorship network, we measured the visibility by degree of nodes which can be associated by higher access rate to resources and consequently, brings reputation in knowledge field. However, we need to statistically test whether the distribution significantly obey the power law distribution. According to (Clauset, Shalizi et al. 2009) and using poweRlaw package (Gillespie 2014), the goodness of fit can be tested for following hypothesis:

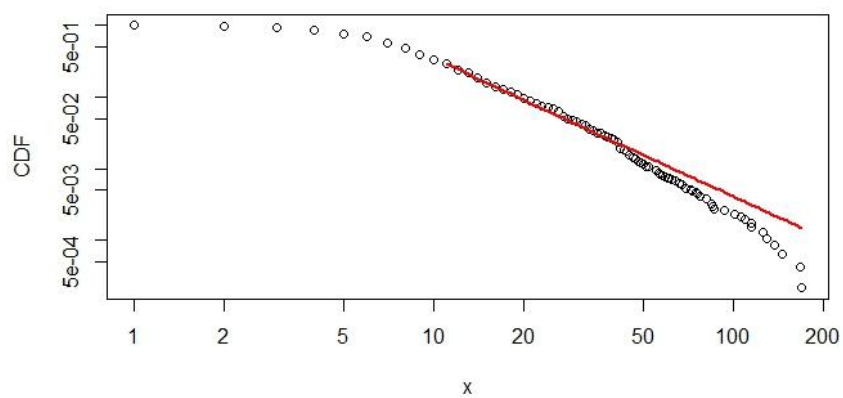
H_0 : data is generated from a power law distribution.
H_1 : data is not generated from a power law distribution.

I run a bootstrapping for 200 iterations over the empirical degree distribution of nodes and calculate the p-value for above hypotheses. If the resulted p-value is not large enough (p-value≈0), then the estimated parameters do not plausibly fit the expected distribution in favor of other type of distributions. Our estimation shows that the mean p-value is equal to 0 indicates that the model does not statistically fit the power law model or simply, the data is generated from a distribution other than power law. Figure 8 shows the cumulative distribution for mean and variance of parameters together with estimated p-value from the bootstrapping procedure.

Table 6: Power law estimation for degree distribution in co-authorship network of scientists

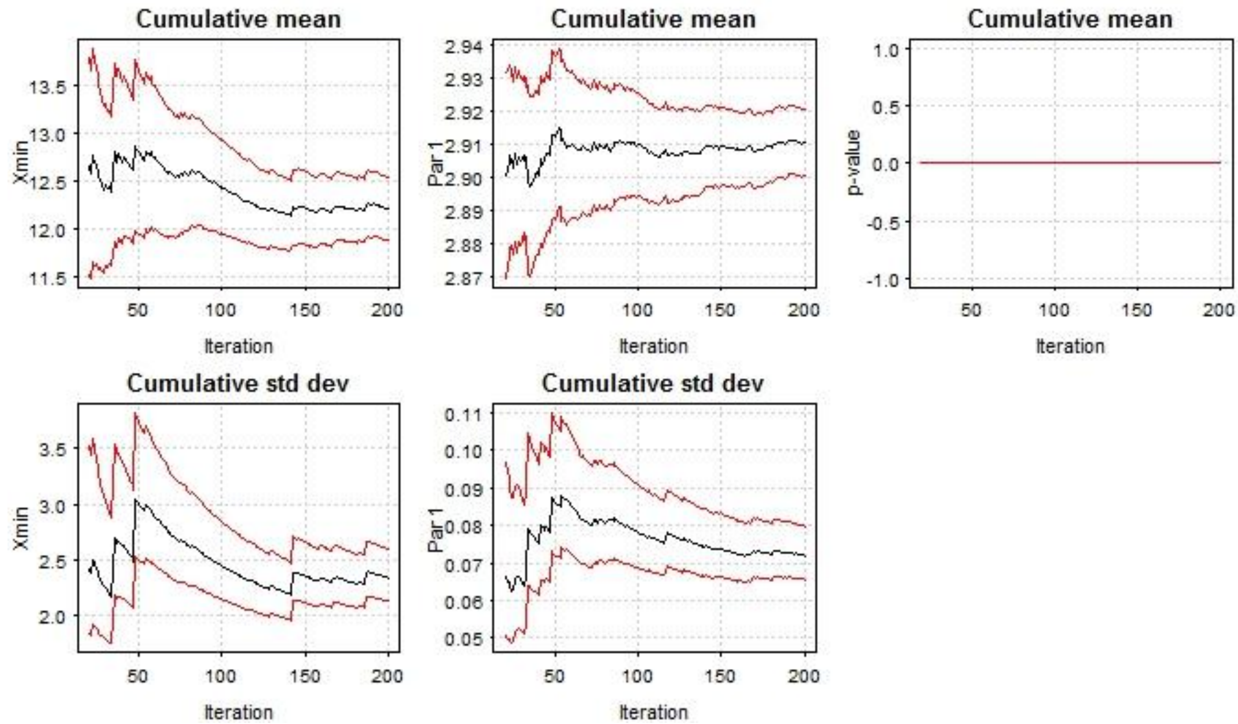
Parameter	1994-1998	1999-2003	2004-2008	2009-2013
x_{min}	7	9	18	11
$\hat{\beta}$	3.34	4.74	4.12	2.91
Kolmogorov-Smirnov (KS)	0.07	0.07	0.05	0.04

Figure 7: Cumulative Distribution Function (CDF) for degree centrality of nodes (2009-13)



Estimated using poweRlaw package (Gillespie 2014) in R. Dot plot shows the degree distribution and red line is the best fit of model.

Figure 8: Bootstrapping the empirical data using a goodness of fit



Estimated using poweRlaw package (Gillespie 2014) in R. Two first columns respectively are cumulative estimations of mean and standard deviations of x_{min} and $\hat{\beta}$. The plot in right hand side provides the mean estimate of p-value. Red lines show the 95% confidence interval.

The power law analysis does not support our hypothesis about the evolution of co-authorship network in German PV industry using preferential attachment mechanism. As Kas et al., (2012) emphasize, “for preferential attachment, what is more important for a new comer is the cumulative earnings of the existing nodes, not how close they are to the newcomer node in the network”, something that is not fulfilled for German PV industry. This property makes the underlying network robust in occurrence of any important elimination of actors and preserves the scientific network against violation of transmission path of knowledge. Furthermore, this result supports our previous observation about the small world structure of co-authoring network. In such a network the local interactions within clusters provide a strong motivation of future collaborations while for preferential attachment neither geodesic nor longitudinal distances play an important role. In preferential attachment the cumulative advantage of existing nodes performs as an important indicator for future path of scientific collaborations of new actors, something that is not compatible with our observation in German PV industry.

Reviewing similar studies in network evolution analysis shows that the existence of preferential attachment is not a universal feature of all co-authorship networks. For example, Abbasi et al., (2012) found a positive significant correlation between creation of new connections and degree centrality of existing authors. In another study, Verspagen and Werker(2004) estimated a log-log model over the betweenness statistic for different sets of collaboration data in Evolutionary Economics field. The estimated model shows a fat tail distribution with an imperfect estimate of exponential parameter for their empirical degree distribution. They conclude that the network under investigation is evolving by new linkages to existing scholars with higher visibility. In contrast, Kas et al., (2012) did not observe such mechanism in their high-energy physics data set. In fact, their estimation for exponent parameter lies beyond the expected range for power law distribution.

4.4. Map of Science in German PV Industry

One alternative for tracing the network dynamics is the emergence of interdisciplinary in considering filed. In fact, new advancements in technological knowledge demands absorption of other proficiencies and qualifications. As a result, it is expected that an emerging network become more interdisciplinary over years. Diversified science categories in a technological field, foster the interactions between scientists and hopefully make the network to grow rapidly.

In this section, I try to shed light on the composition of scientific categories in German PV industry. I draw an overlay map of science in PV industry based on the methodology provided in (Rafols, Porter et al. 2010). As they emphasize, one application of overlay map of science is that we can capture the connections among different categories and their positions in global map of science by visualization the overlapping disciplines.

Table 7 shows most frequent Web of Science (WOS) *subject categories* in the publication data from 1989 to 2013. It is obvious that the percentage column does not add up to 100 percent, because most of papers are classified in more than one category. In considering period, “Physics Applied” and “Materials Science Multidisciplinary” were the most frequent science categories which separately cover about 50 percent of all papers. Also, “Energy Fuels”, “Physics Condensed Matter” and “Chemistry Physical” appeared separately in almost 20 percent of all papers.

Table 7: distribution of publications based on Web of Science, Science categories (1989-2013)

Web of Science Categories	Records	% of 2639
PHYSICS APPLIED	1262	47.82
MATERIALS SCIENCE MULTIDISCIPLINARY	1236	46.84
ENERGY FUELS	649	24.59
PHYSICS CONDENSED MATTER	608	23.04
CHEMISTRY PHYSICAL	502	19.02
CHEMISTRY MULTIDISCIPLINARY	318	12.05
NANOSCIENCE NANOTECHNOLOGY	301	11.41
MATERIALS SCIENCE COATINGS FILMS	187	7.09
POLYMER SCIENCE	128	4.85
ENGINEERING ELECTRICAL ELECTRONIC	125	4.74
OPTICS	124	4.70
PHYSICS ATOMIC MOLECULAR CHEMICAL	96	3.64
ELECTROCHEMISTRY	71	2.69
ENGINEERING CHEMICAL	60	2.27
ENVIRONMENTAL SCIENCES	51	1.93
PHYSICS MULTIDISCIPLINARY	47	1.78
CRYSTALLOGRAPHY	38	1.44
MULTIDISCIPLINARY SCIENCES	29	1.10
MATERIALS SCIENCE CERAMICS	26	0.99
INSTRUMENTS INSTRUMENTATION	25	0.95

Reference: Web of Science

Figure 9-11 disclose the dynamic of scientific integration of micro-disciplines in PV industry for different periods[10] in different periods. It shows that PV related publications started with high concentration on Physics Applied and Material science Multidisciplinary. During the middle period from 1999 to 2003 the share of other categories significantly increased. Specifically, Physics Condensed Material, Energy Fuels and Material Science Coating Films experienced near doubled shares as a science subjects in published papers. More interesting is the emergent of Engineering Electrical Electronics in that period with a sizeable share as subject of scientific writing. Also, Polymer Science and NanoscienceNanothechnology appeared in almost 5 percents of papers revealing the role of new technologies in PV industry. Optics did not have important changes over these years.

Figure 11 illustrates the composition of science for the last 5-year window. We observe rising involvements of Chemistry Multidisciplinary and Chemistry Physical highlighting the importance of chemistry as a macro-discipline in recent scientific researches in PV technology. Also, material Science Coating Film lost its share in recent papers even less than its share on starting period 1989-1993.

Figure 9: Distribution of Science Categories in German PV industry (1989 to 1993)

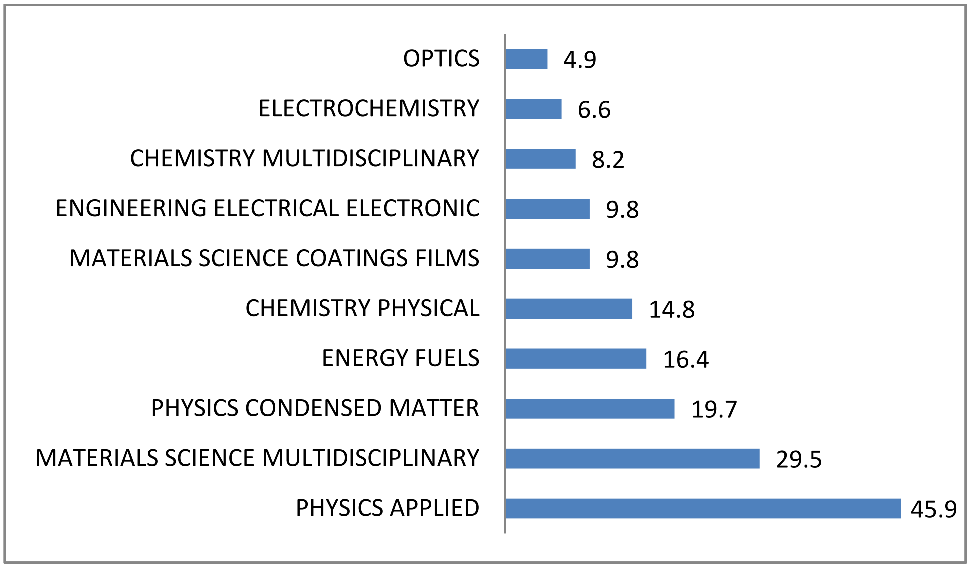


Figure 10: Distribution of Science Categories in German PV industry (1999 to 2003)

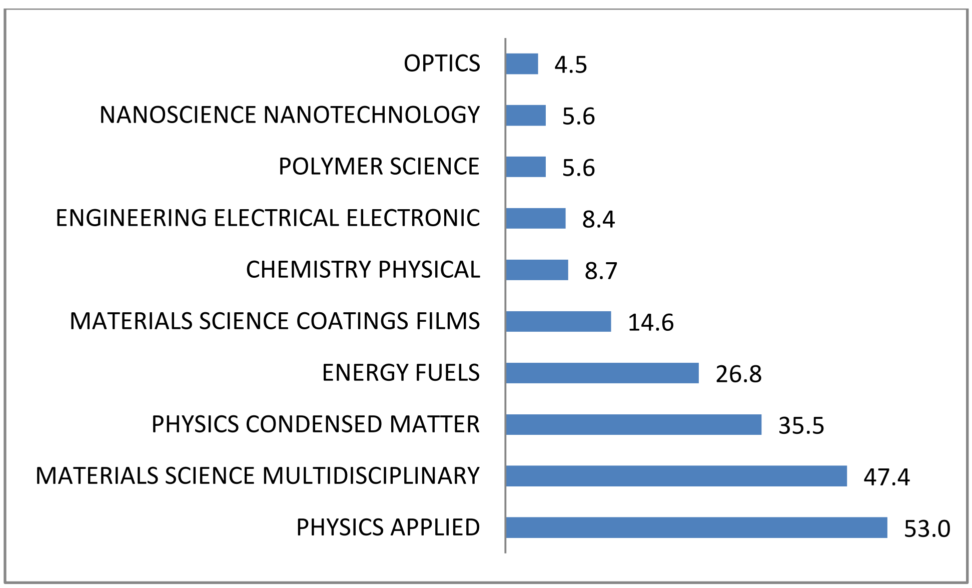
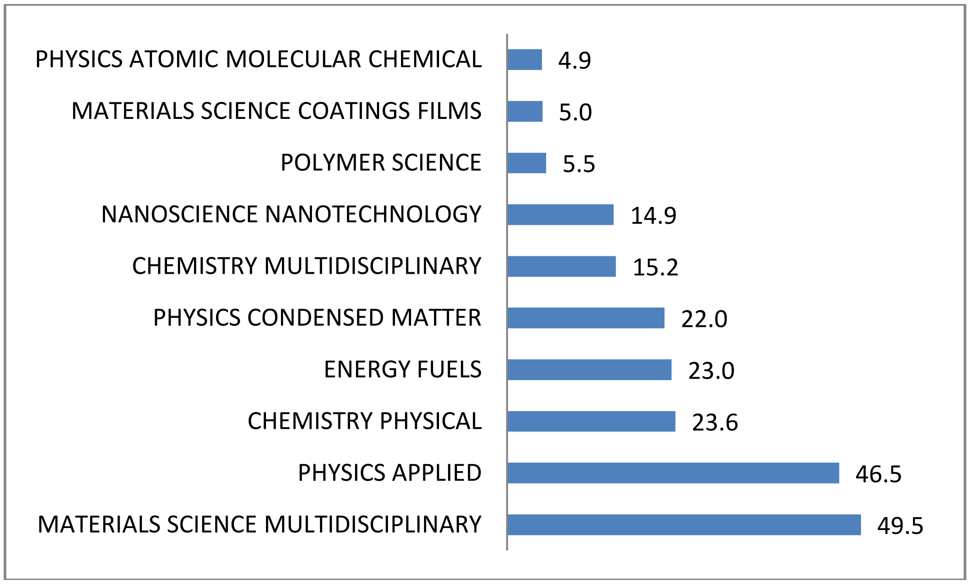


Figure 11: Distribution of Science Categories in German PV industry (2009 to 2013)



Along with above statistical review, we need to assess the relation between each category in the overlay map. Figure 12 illustrates the overlay network in which each nodes represents one WoS subject category. Size of nodes denotes frequency of the category in publication data and size of ties (edges) is representative for similarities of two subject categories (Rafols, Porter et al. 2010). The colures on the map denote the 19 macro-science fields based on WoS classification (also denoted by the colored box on the map).

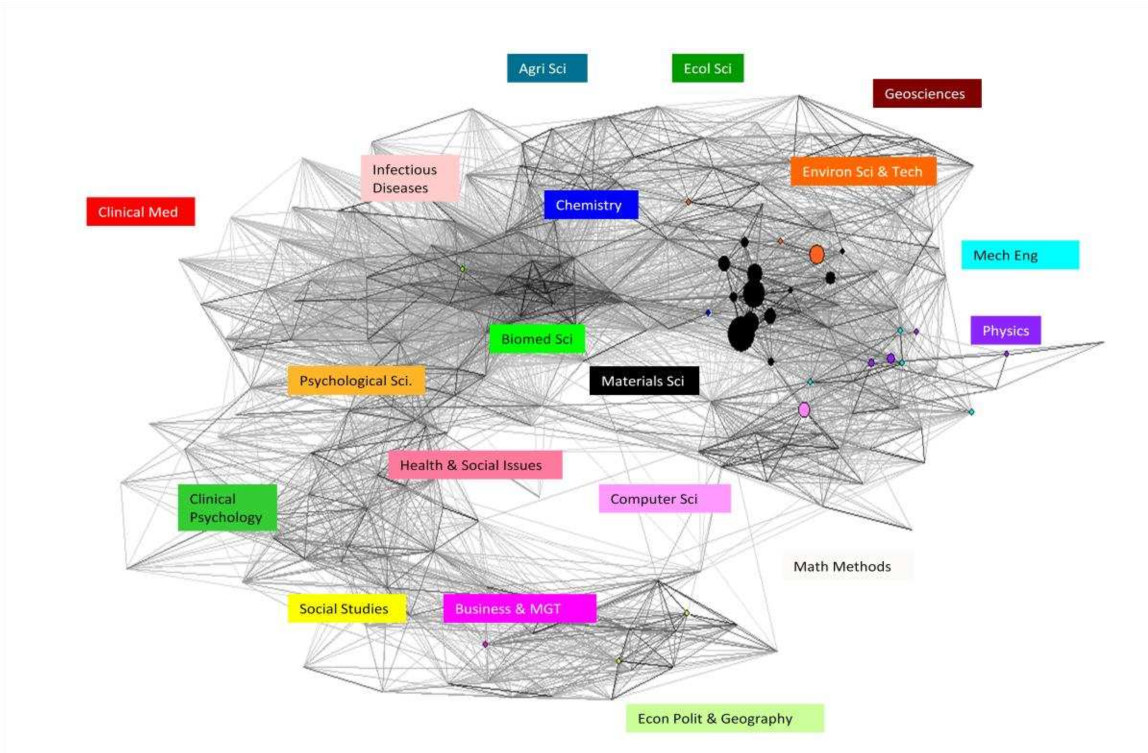
Figure 12 shows the composition of PV science in first period of our study (1989-1993). We see that large nodes are located at the left part of network including “Material Science” and “Environmental Science & Technologies”. Other fields such as “Physics” and “Computer science” located in short geodesic distance of them, indicating the short cognitive distance of fields.

In middle investigating period (1999-2003) we observe the rising involvement of more sub-fields of material science in PV research with significant growth of environmental Science (mostly Energy Fuels), Computer Science and Physics. The interesting result is the association of various science subjects on recent years (2009-2013) in which about 8 macro-disciplines actively have gotten involved in construction the publication network of German PV industry. Specifically it seems that scientists from the Biomedical Science, Chemistry, Mechanical Engineering and

Economic Policy & Geography found PV technology as an interesting filed for their scientific research.

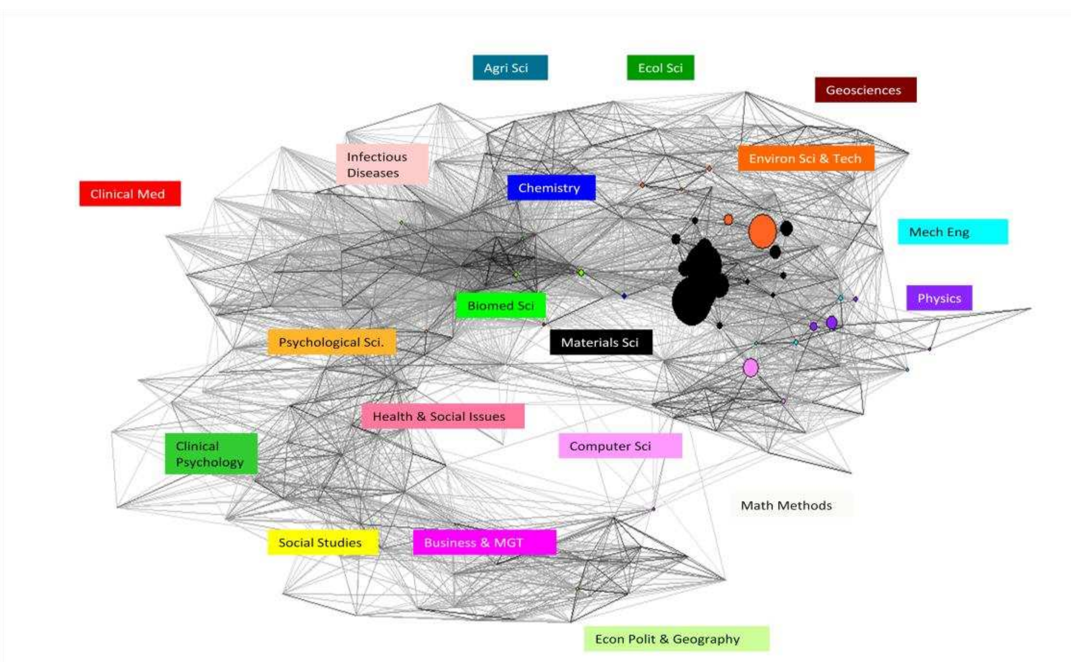
To sum up, PV industry using visual analysis of overlay map of science, turned out to be an interdisciplinary field of knowledge. This result is in line with the cross-country study of Ying et al.,(2009) in *Nono Thin-Film Publication*. Even though their study is limited to a specific technological branch in PV industry, the result is similar to our presentation of the interdisciplinarity nature of this scientific field. This interpretation provides evidence in support of our last hypothesis about the influence of interdisciplinary structure of knowledge and evolution of co-authorship network in German PV industry. Without more statistical analysis, we can conclude that growth of knowledge in this filed requires more complementary expertise which absorbs more disciplines even from the far distances within the overlapping map of science.

Figure 12: Overlay map of sciences in German PV industry (1989-1993)



Produced by Overlay toolkit map[11] presented by (Rafols, Porter et al. 2010) in Pajek.

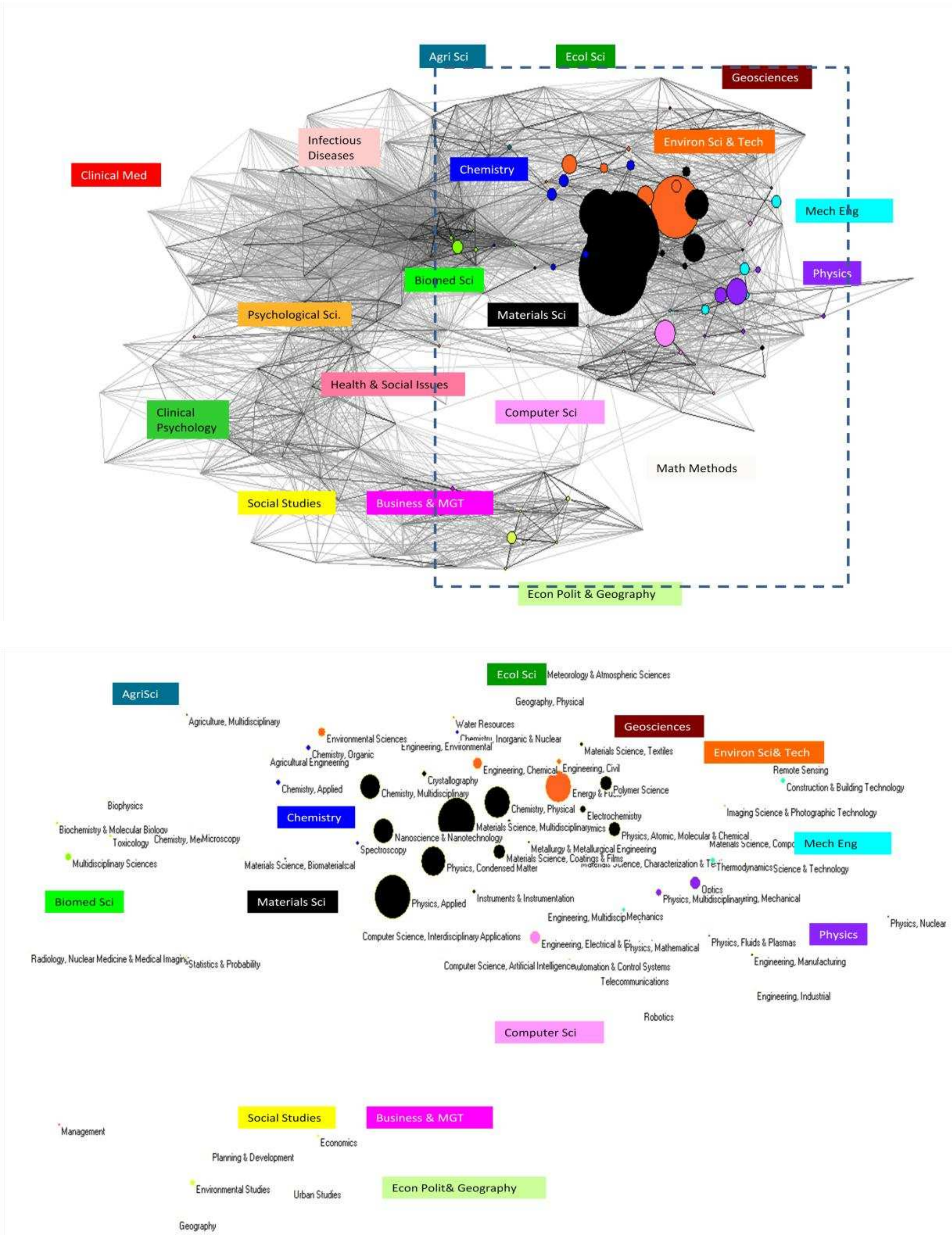
Figure 13: Overlay map of sciences in German PV industry (1999 – 2003)



Produced by Overlay toolkit map[12] presented by (Rafols, Porter et al. 2010) in Pajek

[11]. http://www.leydesdorff.net/overlaytoolkit/manual.riopelle.pdf

Figure 14: Overlay map of sciences in German PV industry (2009 – 2013)



Produced by Overlay toolkit map[13] presented by (Rafols, Porter et al. 2010) in Pajek.

[12]. http://www.leydesdorff.net/overlaytoolkit/manual.riopelle.pdf
[13]. http://www.leydesdorff.net/overlaytoolkit/manual.riopelle.pdf

5. Conclusion and Remarks

One of the important challenges in technological field analysis is to investigate the structural characteristics of the scientific network and explain the drivers behind the evolution of the network. In this study, I concentrated on the publication network of German PV industry and provided a social network analysis using the Web of Science Core Collection data.

In this research, several features of collaboration between institutions illustrated using the appropriate network statistics together with network visualizations in different time windows. Centrality indices helped us to find the most influential actors as potential source of knowledge, actors who have played the central role in knowledge production or acquired a bridge position in co-authorship network connecting other actors together. The result highlighted the importance of German large research institutes in conducting scientific research and publishing scientific papers in PV industry as well as their strong collaboration with foreign institutes. Looking at the evolution of co-authorship network revealed that the institution collaboration grew in a way that (along with emerging main component) assimilates a core-periphery structure in which large actors function as integrating point for smaller ones.

In next step, I investigated the dynamic of co-authorship network of scientists. I provided an empirical analysis to shed light on the status quo and dynamic of networks structure. Results showed that against the decreasing trend in cohesion of network, overall compared to the same size random generated network, German PV co-authoring network characterized as a small world network. Therefore, we can conclude that the network works efficiently in diffusion of knowledge.

To disclose the drivers behind the evolution of co-authorship network, I hypothesized two different scenarios:

First, I tested the existence of preferential attachment mechanism in our co-authorship network to find whether the newcomers try to attach to most significant scholars in the network or simply if the reputation plays as important roles in evolution of collaboration linkages. Fitting power law distribution over degree of nodes rejected our hypothesis in all investigating time windows. Therefore, preferential attachment mechanism cannot explain significantly the evolution of the network and reveals that network is robust in response to removal of large nodes. The result highlights the intensity of research clusters in the network and confirms the revealed small world characteristics of co-authorship network.

Second, looking at the composition of knowledge on map of science provided strong evidence in support of interdisciplinarity nature of German PV industry. Our descriptive analysis shows that along with existence of leading micro-disciplines such as Materials Science and Physics Applied the role of new subject categories of science over the existing knowledge domain was growing over years. As a result, changing the composition of knowledge in this filed have required more complementary expertise and absorbed more disciplines even from far distances (cognitive distance) within the overlapping map of science.

This research applied social network analysis to disclose the dynamic of publication network. Similar to other studies in this field, the result are highly depends on the data preparation procedure. In fact, neglecting the complexity of publication data may cause huge bias in network analysis. One approach is to consider other cleaning algorithms for publication records and compare them with current research. It is obvious that always there is a trade of between reliability of data and time consumption for preparing it and this is the research outline that determines the suitable procedure.

Current research discussed the evolution of network utilizing social network descriptive analysis. Therefore, it may be of interest to test the hypotheses within a larger geographical space (such as European countries) with sufficient observations utilizing efficient econometrics tools. My study has serious limitations in conducting a regional analysis for co-authorship network and so left it for future studies.

REFERENCES

Abbasi, A., L. Hossain and L. Leydesdorff (2012). "Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks." Journal of Informetrics**6**(3): 403-412.

Andrade., M. T. T., P. Braga., T. K. G. Carneiro., N. M. Ribeiro., M. A. Moret. and H. B. d. B. Pereira. (2014). "Contextualized Analysis of Social Networks: Collaboration in Scientific Communities." Social Networking**03**(02): 71-79.

Barabasi, A. L., H. Jeong, Z. Nda, E. Ravasz, A. Schubert and T. Vicsek (2002). "Evolution of the social network of scientific collaborations." Physica A**311**: 590–614.

Batagelj, V. and A. Mrvar (2004). Pajek—analysis and visualization of large networks. Graph drawing software, Springer**:** 77-103.

Bellanca, L. (2009). "Measuring interdisciplinary research: analysis of co-authorship for research staff at the University of York." Biocsience Horizons**2**(2).

Bojanowski, M. (2013). "intergraph: Coercion routines for network data objects in R."

Boyack, K. W. (2004). "Mapping knowledge domains: Characterizing PNAS." Proceedings of the National Academy of Sciences**101**(suppl 1): 5192-5199.

Butts, C. (2008). "network: A Package for Managing Relational Data in R." Journal of Statistical Software**24**(2).

Clauset, A., C. R. Shalizi and M. E. J. Newman (2009). "Power-Law Distributions in Empirical Data." Siam Review**51**(4): 661-703.

Cowan, R. and N. Jonard (2004). "Network structure and the diffusion of knowledge." Journal of Economic Dynamics and Control**28**(8): 1557-1575.

Csardi, G. and T. Nepusz (2006). "The igraph software package for complex network research." InterJournal Complex Systems**1695**: 1695.

Erfanmanesh., M., V. A. Rohani. and A. Abrizah. (2012). "Coauthorship Network of Scientometrics Research Collaboration." Malaysian Journal of Library & Information Science,**17**(3): 73-93.

Fleming, L., C. King and A. I. Juda (2007). "Small Worlds and Regional Innovation." Organization Science**18**(6): 938-954.

Gillespie, C. S. (2014). "Fitting heavy tailed distributions: the poweRlaw package." arXiv preprint arXiv:1407.3492.

González-Alcaide, G., R. Aleixandre-Benavent, C. Navarro-Molina and J. C. Valderrama-Zurián (2008). "Coauthorship networks and institutional collaboration patterns in reproductive biology." Fertility and Sterility**90**(4): 941-956.

Graf, H. (2011). "Gatekeepers in regional networks of innovators." Cambridge Journal of Economics**35**(1): 173-198.

Guo, Y., A. L. Porter and L. Huang (2009). Nanotechnology enhanced thin-film solar cells: analysis of global research activities with future prospects. Proceedings of 18th International Conference for the International Association of Management of Technology, Orlando, FL.

Kas, M., K. M. Carley and L. R. Carley (2012). "Trends in science networks: understanding structures and statistics of scientific networks." Social network analysis and mining**2**(2): 169-187.

Katz., J. S. and B. R. Martin (1997). "What is research collaboration?" Research Policy**26**: 1-18.

Krafft, J., F. Quatraro and P. P. Saviotti (2011). "The knowledge-base evolution in biotechnology: a social network analysis." Economics of Innovation and New Technology**20**(5): 445-475.
Leydesdorff, L. and C. S. Wagner (2008). "International collaboration in science and the formation of a core group." Journal of Informetrics**2**(4): 317-325.
Melin, G. and O. Persson (1996). "Studying research collaboration using co-authorships." Scientometrics**36**(3): 363-377.
Newman, M. E. J. (2001). "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality." Physical Review E**64**(1).
Newman, M. E. J. (2001). "Scientific collaboration networks.I. Network construction and fundamental results." Physical Review E**64**(1).
Persson, O., R. Danell and J. W. Schneider (2009). "How to use Bibexcel for various types of bibliometric analysis." Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday: 9-24.
Porter, A. L., J. D. Roessner, A. S. Cohen and M. Perreault (2006). "Interdisciplinary research: meaning, metrics and nurture." Research Evaluation**15**(3): 187-195.
Porter, A. L., J. Youtie, P. Shapira and D. J. Schoeneck (2007). "Refining search terms for nanotechnology." Journal of Nanoparticle Research**10**(5): 715-728.
Rafols, I., A. L. Porter and L. Leydesdorff (2010). "Science overlay maps: A new tool for research policy and library management." Journal of the American Society for Information Science and Technology**61**(9): 1871-1887.
Verspagen, B. and C. Werker (2004). "Keith Pavitt and the Invisible College of the Economics of Technology and Innovation." Research Policy**33**(9): 1419-1431.
Wagner, C. S. and L. Leydesdorff (2005). "Network structure, self-organization, and the growth of international collaboration in science." Research Policy**34**(10): 1608-1618.
Wagner, C. S., J. D. Roessner, K. Bobb, J. T. Klein, K. W. Boyack, J. Keyton, I. Rafols and K. Börner (2011). "Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature." Journal of Informetrics**5**(1): 14-26.
Wulf, T., P. Missner and F. F. v. Bernewitz (2010). "Future Scenarios for the German Photovoltaic Industry." HHL Working Paper**99**.
www.irena.org/Publications (2013). "Solar Photovoltaics- Technology Brief." IEA-ETSAP and IRENA- Technology Brief E11.
Ying, G., H. Lu and A. L. Porter (2009). Profiling research patterns for a New and Emerging Science and Technology: Dye-Sensitized Solar Cells. Science and Innovation Policy, 2009 Atlanta Conference.