Paper to be presented at
DRUID-Asia, Singapore, February 23-25, 2016
Co-organized by DRUID, NUS Business School and SMU - Lee Kong Chian School of Business.

Glocal: The Influence of Global Knowledge Networks and Local Cluster Characteristics on the Innovation Performance of Photovoltaics Clusters

Pieter Stek
Delft University of Technology
Faculty of Technology, Policy & Management
p.e.stek@tudelft.nl

Marina Van Geenhuizen
Delft University of Technology
Faculty of Technology, Policy & Management
m.s.vangeenhuizen@tudelft.nl

Abstract

The innovation output of the photovoltaics (PV) sector has rapidly expanded during recent decades thanks to the overall increase of investment in sustainable energy production and storage. This research provides insight into the development of the PV sector during the 2006-2010 period at the global and cluster level based on bibliometric data sources (patents and scientific publications). Using a novel cluster identification method, 56 photovoltaic clusters are identified around the world and bibliometric innovation and knowledge network indicators are extracted for each of the clusters. These indicators are based on, or adapted from, the existing literature. These indicators make it possible to quantitatively explore a number of current theories about agglomeration, spatial and relational proximity. The results and analysis suggest that agglomeration effects and the diversity of patent co-invention and co-assignment networks have a positive influence on cluster innovation performance. Small clusters appear to play a more active role in knowledge networks relative to their size; inter-cluster networking may therefore be a way to overcome the lack of agglomeration advantages of clusters. An in-depth analysis of six clusters, including their main innovation producers and the inter-cluster knowledge network, provide further insight into the diversity of individual cluster structures and networks.

Jelcodes:O31,F23

Glocal: The Influence of Global Knowledge Networks and Local Cluster Characteristics on the Innovation Performance of Photovoltaics Clusters

27 January 2016

Pieter E. Stek[1]
Doctoral Student

Prof.dr. Marina S. van Geenhuizen
Professor of Innovation and Innovation Policy in the Urban Economy
Faculty of Technology, Policy and Management, Delft University of Technology,
Jaffalaan 5, 2628 BX Delft, The Netherlands

Abstract
The innovation output of the photovoltaics (PV) sector has rapidly expanded during recent decades thanks to the overall increase of investment in sustainable energy production and storage. This research provides insight into the development of the PV sector during the 2006-2010 period at the global and cluster level based on bibliometric data sources (patents and scientific publications). Using a novel cluster identification method, 56 photovoltaic clusters are identified around the world and bibliometric innovation and knowledge network indicators are extracted for each of the clusters. These indicators are based on, or adapted from, the existing literature. These indicators make it possible to quantitatively explore a number of current theories about agglomeration, spatial and relational proximity. The results and analysis suggest that agglomeration effects and the diversity of patent co-invention and co-assignment networks have a positive influence on cluster innovation performance. Small clusters appear to play a more active role in knowledge networks relative to their size; inter-cluster networking may therefore be a way to overcome the lack of agglomeration advantages of clusters. An in-depth analysis of six clusters, including their main innovation producers and the inter-cluster knowledge network, provide further insight into the diversity of individual cluster structures and networks.

1 Introduction

This paper presents a quantitative assessment of the influence of global knowledge networks on the innovation performance of industry clusters in the photovoltaics (PV) sector. PV is a knowledge intensive and globally distributed sector that incorporates multiple technological domains, including electronics, materials science and optics. The PV sector has also received large inflows of public and private research investment during recent decades as part of a global push towards green growth; PV is expected to contribute to the reduction of greenhouse gas emissions and the diversification of energy supply (Breyer, Birkner, Meiss, Goldschmidt, & Riede, 2013). Along with R&D expenditure, renewable energy investment has also grown rapidly during the past decade, from US$61.82 billion in 2005 to US$328.93 billion in 2015, with solar energy accounting for just under half (US$161.5 billion) of all renewable energy investment in 2015; in fact, since 2015 global investment in renewable energy has exceeded investment in conventional energy sources (Bloomberg New Energy Finance, 2016).

It is therefore not surprising that the innovation performance of the PV sector has been the subject of

[1] E-mail: p.e.stek@tudelft.nl

significant attention, both in the academic community and among policy makers. Academic research has generally focused on the PV sector in one country or a small number of countries in the context of specific topics such as industrial policy, international technology transfer and international competition (de la Tour, Glachant, & Meniere, 2011; Grau, Huo, & Neuhoff, 2012; Kim & Kim, 2015; Klitkou & Godoe, 2013; Lo, Wang, & Huang, 2013; Vidican, McElvaney, Samulewicz, & Al-Saleh, 2012; Wu, 2014; Zheng & Kammen, 2014). There are also numerous studies of specific PV industry clusters which address similar aspects (Dewald & Fromhold-Eisebith, 2015; Klitkou & Coenen, 2013; Luo, Lovely, & Popp, 2013; West, 2014). And there is a small number of studies which explore the size, growth and geographic distribution of innovation output in the PV sector based on data sets that cover the entire world (Breyer et al., 2013; Leydesdorff, Alkemade, Heimeriks, & Hoekstra, 2014).

This study provides a "glocal" perspective on innovation, by identifying 56 regional PV clusters in 17 countries around the world, and by linking the influence of global knowledge networks and specific regional circumstances to the innovation performance of individual PV clusters. The positioning of the research lens at the intersection of global and local factors offers an original contribution to research on PV industry clusters. The global factors considered in this paper are global knowledge networks: networks of personal and institutional relationships that enable the transfer of knowledge on a global scale. Examples include the global research collaboration network (De Prato & Nepelski, 2014) and the global academic weblinks network (Barnett, Park, Jiang, Tang, & Aguillo, 2014). Other salient examples include the cross-border knowledge transfers that take place within multinational corporations (MNCs) which are enabled by MNCs private network of branch laboratories located in industry clusters worldwide (Castellani, Jimenez, & Zanfei, 2013). However the level of international participation in research, including the involvement of MNCs, appears to vary significantly between sectors (Alkemade, Heimeriks, Schoen, Villard, & Laurens, 2015).

While it is clear that international research activities are rapidly growing, especially in terms of their global distribution, which is aimed at exploiting local comparative advantages, and the growing importance of global knowledge networks (Audretsch, Lehmann, & Wright, 2014; Locke & Wellhausen, 2014), there are also indications that internationalization may weaken or limit the development of internal knowledge networks (Kwon, Park, So, & Leydesdorff, 2012; Van Geenhuizen & Nijkamp, 2012; Ye, Yu, & Leydesdorff, 2013), and lower the innovation performance of some knowledge intensive industries (Stek & van Geenhuizen, 2015). The influence of global knowledge networks on the innovation performance of industry clusters therefore remains unclear, and the PV sector is no exception.

To address this knowledge gap multiple regression analyses are carried out to assess the influence of internal cluster characteristics and global knowledge networks on the innovation performance of industry clusters. This multi-scalar approach is based on emerging theories about the role of spatial and relational proximity in innovation and a multidimensional understanding of knowledge networks as consisting of different types of relationships with different spatial reach. In this study three types of global knowledge network are investigated: personal collaboration networks between scientists and inventors, institutional collaboration networks (Dodgson, 1992), and MNC branch networks. Triple Helix (university-government-industry relations) networks (Etzkowitz & Leydesdorff, 2000) are only considered at the cluster level, mainly due to their sparsity at the global level. Other internal cluster characteristics being included in the model are agglomeration effects (Capello, 2009) and the various properties of an innovation system, including the presence of scientific and educational institutions, and entrepreneurship (Autio, Kenney, Mustar, Siegel, & Wright, 2014). Together these cluster characteristics and networks provide a broad and differentiated perspective on relational proximity and cluster innovation performance.

The paper makes a theoretical and methodological contribution to the understanding of innovation performance in industry clusters. The theoretical contribution consists of exploring how relational proximity (global networks) influences spatially concentrated knowledge creation in industry clusters, including the possibility of knowledge flows mainly in one direction, which might raise the innovation performance of the receiving industry cluster at the expense of the sending cluster. This latter option is considered to be a possibility for the MNC branch network, which are conceptualized as directed networks which map information flows *from* and *to* a particular network node (industry cluster) (Wasserman & Faust, 1994).

The paper makes a methodological contribution by using only bibliographic data to extract innovation indicators, combining spatial analysis and scientometric network analysis approaches. This process involves the identification of industry clusters, the extraction of cluster innovation input and output indicators and the construction of multiple inter-cluster knowledge networks from which network indicators are calculated. These innovation input, output and network indicators are then used in linear regression models of innovation performance. Because only bibliographic data are used, the method has the potential to be applied in many other knowledge intensive sectors, especially when other innovation data is not available, which is often the case for studies that encompass the regional and/or global scale.

The study is centered on the following research questions:

1. How has the PV sector developed globally in terms of growth in innovation performance, geographic distribution and the prevalence of knowledge networks during the 2006-2010 period?
2. Which forms of agglomeration influences PV cluster innovation performance?
3. Which internal cluster characteristics influence PV cluster innovation performance?
4. Do the local activities of MNCs influence PV cluster innovation performance?
5. Which global knowledge network characteristics influence PV cluster innovation performance?

These research questions are first discussed using the current literature, which leads to the formulation of hypotheses (section 2). Next, a research model is proposed (section 3). This is followed by a description of the data and methodology (section 4), preceding the results and analysis (section 5). The final section of the paper contains a discussion of the results and a conclusion (section 6).

2 Theory and Hypotheses

This section addresses a number of recent theoretical and empirical advances concerning the role of global knowledge networks in innovation performance as well as literature about agglomeration effects and innovation systems. Taken together a “glocal” perspective on cluster innovation performance emerges, which forms the basis of the research model and hypotheses.

Spatial and relational proximity

Regional approaches to international research interaction have been dominated for years by a rather one-dimensional approach in which spatial proximity enhances knowledge spillovers (Acs, Audretsch, & Feldman, 1994; Audretsch & Feldman, 1996). In the course of the past decades, an increased attention has grown for other types of proximity (Boschma, 2005; Breschi & Lissoni, 2001) and these include among others networks of cooperation in knowledge production (relational proximity) (Asheim & Isaksen, 2002; Bathelt, Malmberg, & Maskell, 2004). Relational proximity may be defined as the capability of regions or clusters (and their organizations and firms) to learn through cooperation with other regions or clusters (Camagni & Capello, 2002). Relational proximity is seen as being facilitated by socio/cultural proximity that enables the absorption of knowledge spillovers from other regional contexts, through a common set of values and beliefs (Fazio & Lavecchia, 2013). In addition to these two approaches, an approach which puts an emphasis on synergies between different proximities, like between spatial and relational proximity has also been developed (Ponds, Van Oort, & Frenken, 2007; Ponds, van Oort, & Frenken, 2010). Benefits from the two types of proximity will be discussed in more detail below, followed by an explanation of different types of knowledge and knowledge relations.

Agglomeration economies
The influence of spatial proximity on innovation performance under the label of agglomeration economies, has been studied theoretically and empirically for decades, inspired by scholars such as Marshall and Jacobs (Capello, 2009). Despite a controversy between ideas on specialization-based advantages (Marshall, 1920) compared to diversity-based advantages (Jacobs, 1970), agglomeration advantages can be discussed using three main dimensions. The first dimension of theses localized advantages relates to being spatially close in larger cities or clusters, including lower transport and transmission costs, proximity to final markets (for firms) or test/launching markets (for innovations), a larger chance for meeting of two agents, eventually leading to serendipity, and easier exchange of creative ideas (Morgan, 2004). The second dimension puts an emphasis on productivity increases due to cost reductions (scale effect) and localized accumulation of production skills. A third dimension draws attention to synergy and refers to the rise of a set of common values and beliefs which in fact act as the rationale for the reduction of transaction costs (Williamson, 1981). In so-far the advantages deal with knowledge, the appropriate concept is localized knowledge spillovers, and invention and innovative activity at universities, research institutes and companies in cities or clusters benefit from them (Acs et al., 1994; Anselin, Varga, & Acs, 1997; Audretsch & Feldman, 1996; Autant-Bernard, 2001; Jaffe, 1989). However, various doubt have been cast on the condition of spatial proximity in productive inter-organizational learning, summarized in the assumption that spatial proximity is neither a necessary nor a sufficient condition for creative learning and innovation. Rather, it merely facilitates (Boschma, 2005; D'Este, Guy, & Iammarino, 2012; Karlsson, 2010).

Clusters are often defined by spatial proximity, or to be more precise: by "geographic concentrations of industries related by knowledge, skills, inputs, demand, and/or other linkages" (Delgado, Porter, & Stern, 2012). However agglomeration effects are also related to the size of the cluster in the regional economy. Therefore agglomeration is often measured using the location quotient (LQ). Typically LQ is calculated based on the ratio between regional sectoral employment E_{SR} and total regional employment E_{TR}, and the ratio between national sectoral employment E_{SN} and total national employment E_{TN}.

$$LQ = (E_{SR} / E_{TR}) / (E_{SN} / E_{TN})$$

A high value for LQ indicates that a cluster is relatively large compared to the regional economy and may therefore exhibit Marshallian agglomeration advantages, while a low value suggests Jacobian agglomeration advantages. However both agglomeration advantages can co-exist: large clusters in large regional economies may benefit from both economies of scale and technological diversity. Hence the following hypotheses are proposed:

H1 PV cluster innovation performance correlates positively to the location quotient (LQ).
H2 PV cluster innovation performance correlates positively to total regional innovation activity.

Relational economies

The idea that ‘advantages of spatial proximity’ also work on a distance and in a similar manner compared to the ones that are localized, has emerged in the literature since the early 2000s and has been increasingly elaborated ever since (Ertur & Koch, 2011). Thus, Breschi and Lissone (2003) argue that collaborative networks are channels for knowledge spillovers that are not limited to local environments, instead, they can span long physical distances (Maggioni, Nosvelli, & Uberti, 2007; Maggioni & Uberti, 2009; Ponds et al., 2007). According to this line of thinking, the study of regional invention and innovation has shifted from a focus on close territorial relationships towards an emphasis on technological collaboration that increasingly occurs between cities or clusters as widely spread network-based systems through which knowledge circulates and is enriched by complementary needs and similar values and culture (Cohendet & Amin, 2006). And in the past years, it has been increasingly recognized that relational proximity between the organizations involved (whether local or global) is key in productive collaborative learning.

Often such a situation has been viewed as merely positive in enhancing innovation in clusters. Particularly in high-tech sectors, research collaboration through global networks has been regarded as crucial for corporate innovative performance, like in the biotechnology industry (Cooke, 2007; Gertler & Levitte, 2005), the automotive industry (Lorentzen & Gastrow, 2012), the electronics industry (Ernst, 2009) and the aerospace industry (Frenken, 2000). What however might occur, if local firms are strongly collaborating with Multinational Corporations (MNCs) from elsewhere or if they are established or acquired by such companies, is that these local firms develop knowledge strategies depending on their role in the production organization within the (parent) MNC. Particularly, the role of producing knowledge for the MNC means that MNCs elsewhere learn from their foreign subsidiaries, which is named ‘reverse’ knowledge transfer in some studies (Ambos, Ambos, & Schlegelmilch, 2006; Awate, Larsen, & Mudambi, 2014; Castellani et al., 2013; Dunning, 2000; Frost & Zhou, 2005; Frost, 2001). Under these conditions, foreign MNCs extract knowledge from the cluster and if this is based on exclusivity, the cluster might weaken instead of grow due to global research interaction. An 'excessive' level of MNC knowledge extraction could therefore lower PV cluster innovation performance.

H3 PV cluster innovation performance correlates positively to the net inflow of knowledge through the MNC branch network.

Firms within a cluster can leverage their relationships with partners inside and outside of the cluster. In smaller clusters which lack the local networking advantages of agglomeration economies, it is likely that knowledge relations with partners outside of the cluster play a more prominent role (Tödtling & Trippl, 2005). At the same time, the strong relational proximity of a cluster to other clusters may be strengthening the spatial agglomeration effect, as large clusters tend to be important nodes in national and global knowledge networks (Bathelt et al., 2004), for example in biotechnology (Huallacháin & Lee, 2014). The ability to establish “pipelines” (Bathelt et al., 2004) to access specific knowledge in clusters around the world and take advantage of local competitive advantages (Audretsch et al., 2014) suggests that network diversity benefits cluster innovation performance, while a high collaboration intensity with other clusters may signal 'reverse' knowledge transfers (Ambos et al., 2006) or a lack of local collaboration partners (Tödtling & Trippl, 2005). Smaller clusters may therefore have a higher intensity of exchanges with entities in other clusters relative to their size. The following hypotheses are proposed:

H4 *PV cluster innovation performance correlates positively to the diversity of inter-cluster knowledge relations.*
H5 *PV cluster innovation performance correlates negatively to the intensity of inter-cluster knowledge relations.*
H6 *PV cluster size correlates negatively to the intensity of inter-cluster knowledge relations per researcher.*

However, as the example of knowledge relations within MNCs shows, the type of knowledge and the type of knowledge relationship also influences a cluster's innovation performance.

Differentiation in knowledge and knowledge relations
Thus, in addition to considering the spatial dimension (local relationship inside the cluster and relationships between organizations in different clusters), there are significant differences in type of knowledge and knowledge relations. Most prominent is perhaps the distinction between tacit and codified knowledge (Gertler, 2003; Simmie, 2003), both of which play an important role in the innovation process. Because codified knowledge is easier to communicate, it is understandable that formal research collaboration between clusters has been identified as a factor that enhances regional innovation performance (Huallacháin & Lee, 2014), but this is not evident from formal within-cluster research collaboration (Fritsch, 2004). Within-cluster collaboration may involve primarily tacit exchanges of knowledge, which are not captured by formal research collaboration indicators such as co-invented patents.

In fact research collaboration, which has often been measured through co-authorship only accounts for a small share of the total knowledge transfer that takes place between institutions, including between university and industry (Bekkers & Bodas Freitas, 2008; Bukvova, 2010). In addition to this, research collaboration is sensitive to the power realities between the research partners (Hervas-Oliver & Albors-Garrigos, 2013; Van Geenhuizen & Nijkamp, 2012), with a stronger partner often exerting more influence over how the relationship is conducted and how potential benefits are appropriated. Even within MNCs, larger labs in more prominent clusters tend to have significant autonomy over how and what kind of research they conduct (Mudambi & Navarra, 2004).

Patterns of knowledge relations may also vary depending on the industry sector (Iammarino & McCann, 2006; Jensen, Johnson, Lorenz, & Lundvall, 2007), the knowledge resource base and the social capital of the region (Masciarelli, Laursen, & Prencipe, 2010; Tödtling & Trippl, 2005). Finally the type of actors involved in the collaboration and the benefits created by spanning mutual boundaries, such as between universities, government and industry as in the Triple Helix network, could make a difference in innovation performance (Etzkowitz & Leydesdorff, 2000; Ranga & Etzkowitz, 2011).

In the context of the spatial and relational proximity, the previous discussion suggests that a plurality of relations and networks co-exist which may influence innovation performance in different ways. It is therefore highly relevant to also explore the differences between global knowledge networks based on different network participants.

H7 *PV cluster innovation performance correlates positively to inter-cluster patent co-invention.*

H8 PV cluster innovation performance correlates positively to inter-cluster patent co-assignment.
H9 PV cluster innovation performance correlates positively to inter-cluster scientific co-authorship.

Despite the fact that co-authorship only measures a small part of total knowledge exchanges (Bekkers & Bodas Freitas, 2008; Bukvova, 2010), hypotheses H7, H8 and H9 cover inter-personal patent, inter-organizational patent and scientific collaboration, which differ significantly in terms of the knowledge process that give rise to such co-authorship or co-ownership. Co-assignment, which is co-ownership of intellectual property, usually signals R&D co-investment by two or more organizations or individuals, a form of "open innovation" (Chesbrough, 2006). If a single organization funds the research, they will typically receive full ownership of a patent even if other parties are also involved in carrying out the research (Gautam, Kodama, & Enomoto, 2014). This is in sharp contrast to scientific collaboration, where the listed authors usually contributed to the research, or were involved in the context of an other inter-personal relationship (Bukvova, 2010). Similarly, co-invention usually signals involvement of the listed individuals in the production of the patent, with patent inventors typically being associated with industry, and scientific publications originating predominantly from authors based at a university.

Local cluster characteristics
Several models for understanding innovation systems have been proposed at the national scale (Edquist, 1997; Freeman, 1995; Lundvall, 1992; Nelson, 1993), the regional scale (Cooke, 2001; Porter, 1998; Tödtling & Trippl, 2005) and at the sectoral level (Malerba & Orsenigo, 1997; Porter, 1998). All of these models incorporate university, industry and government actors in some form, and highlight the importance of interaction and co-evolution between the innovation system participants. The most explicit formulation of this concept is the presence of Triple Helix relations between academia, business and the public sector, which if well-balanced and well-developed enhance innovation performance (Etzkowitz & Leydesdorff, 2000). In addition to the Triple Helix there is a growing literature exploring the influence of factors such as entrepreneurship on innovation performance (Autio et al., 2014) and the contribution of universities to entrepreneurship and innovation (Looy et al., 2011; Perkmann et al., 2013). The local presence of strong universities not only contributes to the local scientific base upon which innovators can draw, it is also an indicator of the local availability of research talent: top universities attract top talent, and firms establish themselves in a cluster to gain access to that talent pool (Wuebker, Acs, & Florida, 2010). Therefore it is posited that:

H10 PV cluster innovation performance correlates positively to the size of the local scientific base of a cluster.
H11 PV cluster innovation performance correlates positively to the share of Triple Helix relations in total innovation output.

Innovation output in hypothesis H11 refers to patents and scientific publications. However it must be noted that the importance of the scientific base to innovation performance depends on the technological base of the industry, as well as on other factors (Asheim & Coenen, 2005; Iammarino & McCann, 2006; Mangematin & Errabi, 2012).

While it is difficult to measure entrepreneurship directly, it has been suggested that in the early stages of a cluster's life cycle, when it is most innovative, many new firms exist. Later on consolidation occurs and the total number of firms declines, along with innovation performance (Hervas-Oliver & Albors-Garrigos, 2013). Therefore the number of organizations in a cluster relative to its size could be seen as an indicator of entrepreneurship, which also influences innovation performance.

H12 PV cluster innovation performance correlates positively to the number of organizations per researcher in the cluster.

The above theoretical concepts and the eleven hypotheses are incorporated into the research model.

3 Research model

In this research the influence of local cluster characteristics and global knowledge networks on the innovation performance of PV industry clusters is evaluated. Therefore innovation performance is the dependent variable and the various characteristics and knowledge networks are the independent variables. Clusters are the unit of analysis.

The research model consists of five sub-models of innovation performance: an agglomeration model (incorporating hypotheses H1 and H2), a model of the presence of MNCs (hypothesis H3), a model of global knowledge networks (hypotheses H4, H5, H7, H8 and H9), a model of local cluster characteristics (hypotheses H10, H11 and H12) and a model of agglomeration-network effects (hypothesis H6). An illustration of the research model and the four sub-models is offered in figure 1.

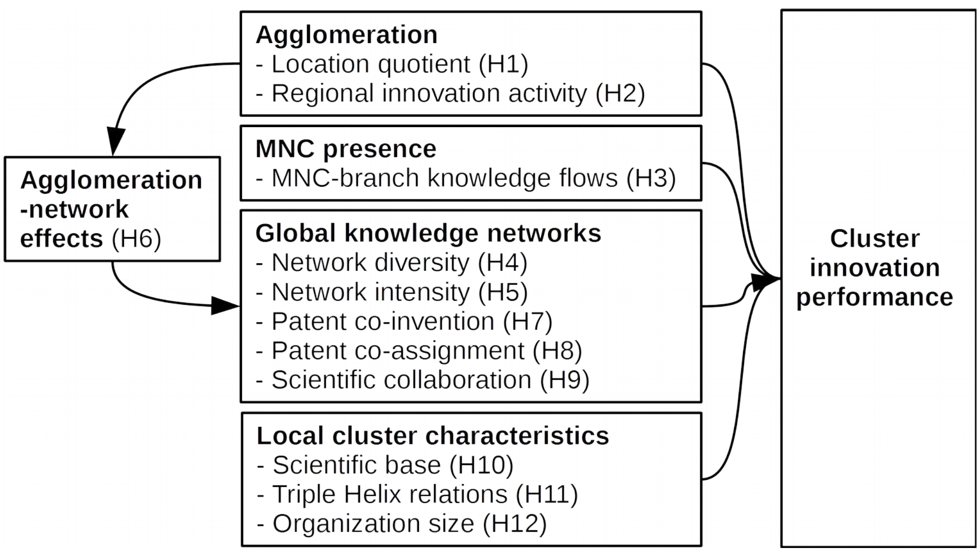

Figure 1: Research model

The use of sub-models has an important implication for the linear regression analysis. If all hypotheses and thus all indicators are incorporated into a single model, then statistically weaker effects are likely to be excluded from an indicator selection process. This exclusion is problematic given that the indicators being used are proxies for diverse and complex socioeconomic, geographic and technological phenomena. Different kinds of indicators may therefore carry different levels of statistical significance. All sub-models consist of similar kinds of indicators, except for agglomeration-network effects, where network and agglomeration indicators are combined. A description of the data and indicators follows below.

4 Data, methodology and indicators

This study considers the 2006-2010 period, which is the most recent 5-year period for which detailed bibliometric data is available. In this section the data, sector and cluster identification, indicator development and the regression model are described.

Data

This study is based exclusively on bibliographic sources (patents and scientific publications) which enable the observation of changes over longer time periods, while offering global coverage at a local scale by exploiting the address information contained in the bibliographic sources. These bibliographic sources form a “paper trail” of innovation activity (Jaffe, Trajtenberg, & Henderson, 1993), the data from which can be used to carry out regression analysis at the cluster-level.

On the one hand the use of bibliometric data as an innovation indicator has disadvantages, including variations in patenting propensity between sectors (Kleinknecht, Montfort, & Brouwer, 2002). On the other hand bibliometric indicators such as patent counts and citation counts seem like valid indicators in high-technology industries and tend to show close statistical overlap with other innovation indicators such as R&D inputs and new product announcements, which are also used to measure innovation performance (Hagedoorn & Cloodt, 2003). A summary of the data sources used in this study is provided in table 1.

Features	**Patent grants**	**Scientific publications**
Period (this study)	2006-2010	2006-2010
Home address	Available	None
Institutional address	Available	Available
Inbound citations data	Available	Available
Publication lag	up to 5 years, sometimes more	usually up to a year
Most active institutions	Industry	University
Data source (for this study)	Patent Statistical Database (PATSTAT) by the European Patent Office	Scopus® by Elsevier
Selection criterion/search term	Cooperative Patent Classification (CPC): Y02E/5... (“PV cells”) Y02E/6… (“Thermal PV hybrids”)	Advanced search: ABS(“solar cell” OR “photovoltaic cell”) AND SUBJAREA(ceng OR chem OR comp OR eart OR ener OR engi OR mate OR math OR phys)
Documents selected (this study)	11,631	9,241

Table 1: Patent grants and scientific publication data as sources for quantitative indicators

For this study, bibliometric data is obtained from two sources. Patent data is obtained from the Spring 2015 edition of the Patent Statistical Database (PATSTAT), which is published every 6 months by the European Patent Office (EPO) and contains data from all major patent offices. Scientific publications were downloaded from the *Scopus®* database, which is maintained by Elsevier, an academic publisher. All data processingis carried out using *R* (R Core Team, 2015). Further, *rvest* (Wickham, 2015) was used to communicate with the *Scopus* Application Programming Interface (API). *RMySQL* (Ooms, James, DebRoy, Wickham, & Horner, 2015) was used to communicate with and populate a MySQL database of bibliographic data.

Supplementing the bibliometric data are three further data sets, which aid in the calculation of specific indicators, they are:

1. Populated places data, which contains coordinates, place names and population information[2],
2. National patent applications and national population data, to calculate the average number of patent applications per resident[3],
3. Patent assignee classification data, to allocate patents to industry, university and government/non-profit groups, taken from the ECOOM-EUROSTAT-EPO PATSTAT Person Augmented Table (Du Plessis, Van Looy, Song, & Magerman, 2009; Magerman, Grouwels, Song, & Van Looy, 2009; Peeters, Song, Callaert, Grouwels, & Van Looy, 2009).

Sector Identification

For the delineation of the technological domain this study uses the new Y02E cooperative patent classification (CPC) from the United States Patent and Trademark Office (USPTO) and the European Patent Office (EPO), which contains renewable energy technologies, including a subcategory for photovoltaics (Y02E/5) and hybrid thermal and photovoltaic technologies (Y02E/6). For scientific publications the authors are not aware of a similar classification, and therefore rely on the advanced search option in *Scopus*®, where all documents (both conference proceedings and journal publications) which have abstracts that contain the phrases “photovoltaic cell” or “solar cell” and which fall within the *Scopus*®-designated subject areas of chemical engineering, chemistry, computers, earth science, energy, engineering, materials, maths or physics.

Cluster identification

As mentioned earlier, “clusters are geographic concentrations of industries related by knowledge, skills, inputs, demand, and/or other linkages” (Delgado, Porter, & Stern, 2014). Accordingly, this study uses two criteria to identify geographic clusters: concentration and connectedness. Both of these measures are extracted from patent data because the focus of this study is on innovation performance in industry, which is primarily reflected in patent output (Hagedoorn & Cloodt, 2003; Jaffe et al., 1993). Concentration refers to the intensity of patent output in geographic space based on the stated place of residence of inventors. Because patents can be assigned to organizations far away from where the actual innovation took place, inventor locations are the more reliable geographic indicator of innovation activity. On the other hand connectedness is based on the extent to which patenting organizations located within the cluster are connected to other organizations through co-assignment relationships. For a cluster to be recognized as such, the organizations located within it should be more connected than those located outside of a cluster. This connectedness criterion ensures that a single but large laboratory is not accidentally identified as a cluster.

To geographically locate patents (and scientific publications) the (partial) addresses of the affiliated institutions (including assignees) and addresses of inventors are used. These addresses can be assigned coordinates using commercial mapping applications such as the Bing Maps API and Google

---

[2] The Populated Places file was used (version 3.0.0), and is available at www.naturalearthdata.com. *Natural Earth* is an open-source project supported by universities, non-governmental organizations and the private sector.

[3] National patent application data and population data for the 2006-2010 period is available from the World Bank at http://data.worldbank.org/indicator/IP.PAT.RESD and http://data.worldbank.org/indicator/SP.POP.TOTL. Because the World Bank does not have data for Taiwan, the South Korea average is used for Taiwan.

Geolocation API. This approach has been used previously with a smaller data set (Leydesdorff et al., 2014; Leydesdorff & Persson, 2010). By using these two different mapping applications, and doing some manual searches, there is a high success rate of geolocating addresses which is between 94 and 98%.

The conversion of addresses into coordinates allows the locations of inventors to be plotted on a map. Based on a plot of these inventors, clusters are identified by using the standard "heat map" algorithm, formally known as kernel density estimation (Parzen, 1962; Rosenblatt, 1956). In this study a standard quartic (biweight) kernel shape is used with a radius of 30 km. Areas with the 95% highest concentrations are designated as cluster-cores, and a radius of again 30 km is drawn around these cores to designate the area of concentration, which is the first step in identifying a cluster. The 30 km radius is based on the 20 mile radius used in a previous study (Alcácer & Zhao, 2013) and coincides with a typical daily commuting distance, which is often used as a cluster radius (van Egeraat, Morgenroth, Kroes, Curran, & Gleeson, 2015). Although commuting distances vary widely, for instance the average daily commuting distance (going + returning) for cities in the United States varies from 15 km (Stockton-Lodi, CA) to 41 km (Atlanta-Sandy Springs-Roswell, GA) (Brookings Institution, 2015), 30 km seems like a reasonable average maximum distance at which intensive interaction can easily occur. The 95% concentration threshold yields the largest number of areas of concentration. Thresholds that are higher or lower yield fewer areas: lower thresholds tend to create a small number of very large areas, higher thresholds identify only a small number of small area (i.e. many false negatives).

The second step in the cluster identification process is based on connectedness. The connectedness of organizations located within the cluster is measured using the weighted degree centrality (Wasserman & Faust, 1994) of that organization in the patent co-assignment network. The average centrality of all organizations within an area of concentration can be compared to the average centrality of all organizations not located inside an area of concentration. Only those areas of concentration with an average centrality higher than the outside average are identified as clusters. In total 56 clusters are identified, see table 4 in the appendix. Clusters are named after the largest urban population center located inside the cluster. To avoid identifying very small "clusters" a minimum threshold of 5 patents and 5 scientific publications during the 2006-2010 period is imposed. A very small "cluster" could be a single laboratory, which produces a large number of patents in collaboration with headquarters, and would therefore be a clear outlier.

Indicators

The indicators used in this study are summarized in table 2 and organized based on the four sub-models outlined in figure 1. All indicators are calculated at the cluster-level.

Innovation performance is measured as the average number of patent citations received by patents with inventors from the cluster **(PATC)**. Patent citations are an indicator of patent value (Dernis, Guellec, & Van Pottelsberghe de la Potterie, 2001; Lanjouw & Schankerman, 2004; OECD, 2009; Squicciarini, Dernis, & Criscuolo, 2013; Webb, Dernis, Harhoff, & Hoisl, 2005). In case of inventors living in multiple clusters, the citations are divided based on the ratio of inventors. For example, if a patent has 2 inventors from cluster A and 3 inventors from cluster B, a total of 10 citations are assigned as 4 to cluster A and 6 to cluster B. The total number of researchers in a cluster **(RESR)** is measured by counting the number of unique inventor names registered to a cluster during a given year.

Besides the researcher population, agglomeration is also measured using regional patent output

(RPAT), which is calculated by adding the population of all population centers located inside of the cluster boundaries, and multiplying them by the average national patent output per resident (see also: *Data* section). To evaluate whether a region exhibits Jacobian (diversity) or Marshallian (specialization) agglomeration advantages the patenting location quotient **(PTLQ)** is calculated using the location quotient formula (see also: section 2) (McCann, 2013), where employment is replaced by patent output. Thus cluster patent output (*CPAT*) relative to regional patent output (RPAT) is compared to total cluster patent output ($CPAT_T$, for all 56 clusters) relative to total regional patent output ($RPAT_T$).

$$PTLQ = (CPAT / RPAT) / (CPAT_T / RPAT_T)$$

The indicator for MNC presence is based on the net inflow to a cluster from the inventor to assignee network. The inventor to assignee network is a directed network, with knowledge flowing from a branch to headquarters. The network is derived from patents where the inventor and assignee are from different clusters, the inventor's cluster is presumed to be the “branch” location, the assignee's cluster is presumed to be the “headquarters”. This approach is not perfect, as firms sometimes file patents under subsidiaries (i.e. not the headquarters) and sometimes write the headquarters' address as the inventor's address. However for most patents which fall within this category, perhaps 90% or more, the inventor and headquarter addresses appear to be correct. The inflow to a cluster is calculated by subtracting the outbound degree centrality from the inbound degree centrality. The degree centrality is a measure of the number of connections a cluster has to other clusters; simple degree centrality counts each connection to a separate cluster as 1, weighted degree centrality counts the total number of connections between different clusters (Wasserman & Faust, 1994). Degree centrality is calculated using the *igraph* package in R (Csardi & Nepusz, 2006). The presence of MNCs in a cluster is therefore measured based on the simple net inflow **(FLWS)** and the weighted net inflow per researcher **(FLWW)**. The weighted degree centrality is calculated per researcher because the total number of connections is likely to be influenced by cluster size.

Unlike the inventor to assignee network, the other global knowledge networks are undirected: the relationship between two parties is presumed to be equal. The patent co-invention network is derived from patents with two or more inventors located in different clusters. The patent co-assignment network is derived from patents with two or more assignees located in different clusters. And the scientific co-authorship network is extracted from scientific publications with authors affiliated to institutions located in two or more clusters. For all three networks the simple degree centrality **(INNS, ASNS, SCNS)** and weighted degree centrality per researcher **(INNW, ASNW, SCNW)** are calculated.

Then concerning local cluster characteristics: the scientific base **(SCIB)** is measured by calculating the total number of citations received by scientific publications with authors affiliated to an institution inside the cluster. The scientific base is measured relative to the size of patent output, so total scientific publications are divided by total patent output. Scientific publications are an indicator of scientific quality (Waltman et al., 2012). Adjustments are made for scientific publications from multiple clusters in the same way as for patent citations.

Theoretical concept	**Indicator(s)**	**Transfor -mation**	**Statistical summary** (n = 56)

Innovation performance	*PATC*	Patent citations per researcher	none	min.: 0.037 max.: 0.263	mean: 0.106
Agglomeration	*RESR*	Researcher population	$\log_{10}$	min.: 1.663 max.: 3.858	mean: 2.540
	RPAT	Regional patent output	$\log_{10}$	min.: 2.291 max.: 5.739	mean: 3.856
	PTLQ	Patenting location quotient	$\log_{10}$	min.: -1.109 max.: 1.515	mean: 0.287
MNC presence	*FLWS*	Net inventor to assignee knowledge flow, simple	none	min.: max.:	mean:
	FLWW	Net inventor to assignee knowledge flow, weighted per researcher	none	min.: max.:	mean:
Global knowledge networks	*INNS*	Patent co-invention network, simple	none	min.: 3 max.: 62	mean: 18.7
	INNW	Patent co-invention network, weighted per researcher	$\log_{10}$	min.: -1.619 max.: -0.297	mean: -0.748
	ASNS	Patent co-assignment network, simple	none	min.: 1 max.: 22	mean: 5.2
	ASNW	Patent co-assignment network, weighted per researcher	$\log_{10}$	min.: -2.708 max.: -0.327	mean: -1.675
	SCNS	Scientific co-authorship network, simple	none	min.: 1 max.: 19	mean: 4.2
	SCNW	Scientific co-authorship network, weighted per researcher	$\log_{10}$	min.: -2.918 max.: -0.155	mean: -1.786
Local cluster characteristics	*SCIB*	Relative scientific base	$\log_{10}$	min.: -0.948 max.: 1.641	mean: 0.048
	THSS	Triple Helix share, scientific publications	none	min.: 0.01 max.: 0.353	mean: 0.108
	THSP	Triple Helix share, patents	$\log_{10}$	min.: -2.699 max.: -0.540	mean: -1.530
	ORGS	Average organization size	$\log_{10}$	min.: -1.575 max.: 0.263	mean: -1.025

Table 2: Description of model indicators and statistical summaries.

The Triple Helix indicators are calculated as the share of all scientific publications **(THSS)** or patents **(THSP)** which have two or more affiliations or assignees from the following groups: universities, government and industry. For patents, institutions are assigned to groups based on the aforementioned ECOOM-EUROSTAT-EPO PATSTAT Person Augmented Table. For scientific publications a simpler categorization is used, which uses key words such as "company", "Ltd." "university", "college", "national institute", etc. in various languages to categorize institutions into one of the three groups (see table 5 in the appendix). An inter-cluster Triple Helix network can also be constructed in this way, but there are too few data points in the present study to conduct such an analysis. This suggests that inter-cluster Triple Helix relations are rare.

Average organization size **(ORGS)**, a measure of entrepreneurship, is calculated by dividing the number of unique organizations with an address in the cluster that patent during a particular year, by the number of researchers.

For the multiple regression analysis to be predictive, all indicators are modified so that they are as close as possible to a normal distribution, hence the transformations in table 2. To avoid losing data, 0-values are replaced with 1 for centrality indicators and with 0.01 for other indicators.

Multiple linear regression
The multiple regression analysis and related statistical diagnostic tests are carried out in Stata (StataCorp, 2015). An overview of the statistical diagnostics tests is given in table 6 (appendix). A description of the regression models and results follows below.

5 Results and analysis

The results and analysis start with a global overview, which leads into the regression analysis. Based on the insights from the regression analysis, six individual clusters are described in more detail.

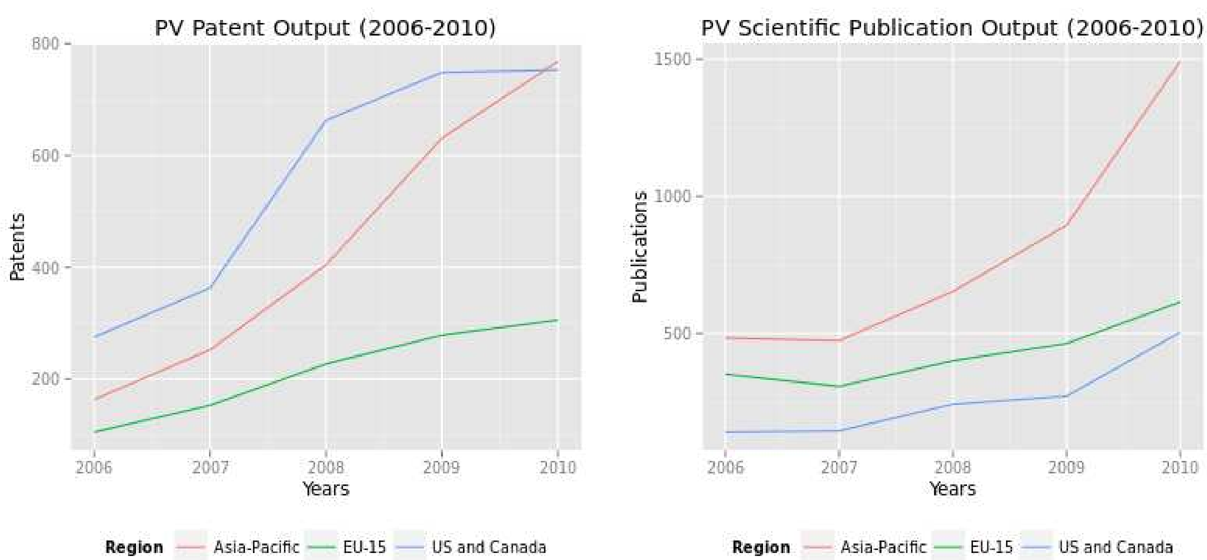

Figure 2: Global PV innovation output, 2006-2010

Global Overview
The innovation output of the PV sector has rapidly grown during the 2006-2010 period, see figure 2. Both in terms of patent output and scientific publications, growth in the Asia-Pacific region[4] has been

[4] The Asia-Pacific region as defined in this paper consists of: Australia, China, Hong Kong, Japan, Korea, Macao, New

most rapid, with 2010 marking the first year in which knowledge production in Asia-Pacific overtook the US & Canada. The EU-15[5] has been lagging in terms of patent output, but has held the second-place globally in terms of scientific publication output, while the US & Canada have been lagging in scientific publication output, but before 2010 were the largest in patent output.

These aggregate global statistics on PV innovation paint a picture of a sector that is in flux, with the expansion of Asia-Pacific R&D activity, and large imbalances between the three major regions in terms of patent output and scientific publications.

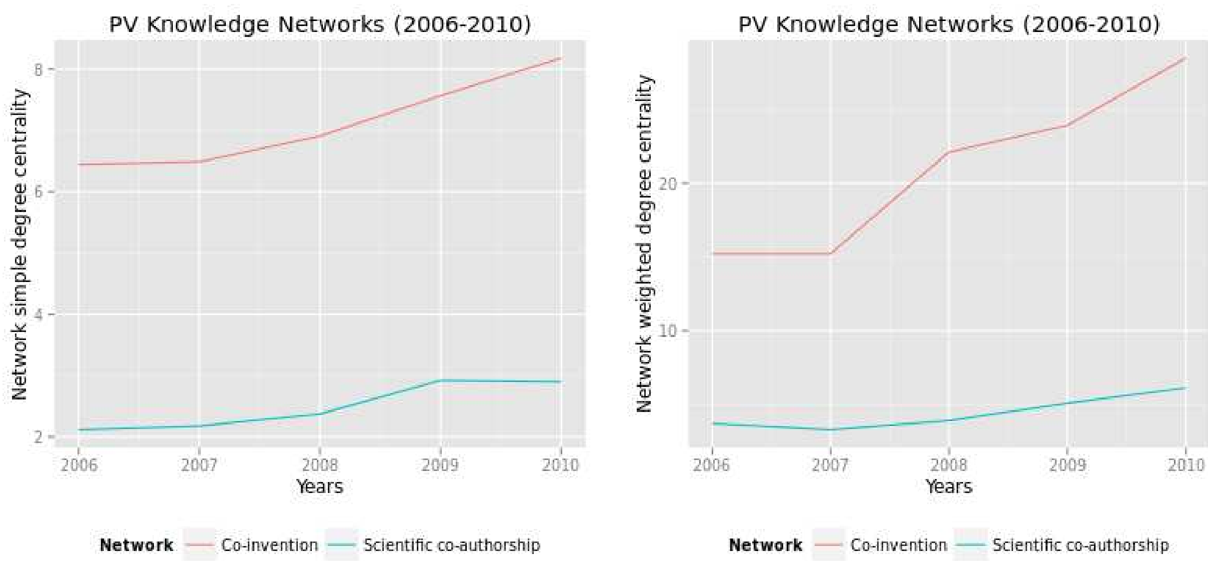

Figure 3: PV inter-cluster knowledge networks, 2006-2010

While innovation output has been increasing rapidly, see figure 3, the growth of knowledge networks appears to have been slower. Although both the co-invention and scientific co-authorship networks are becoming increasingly inter-connected, the total number of connections, as shown by the average network weighted degree centrality, appears to be slower. Furthermore, the patent co-invention network is significantly better in terms of diversity and intensity than the scientific co-authorship network. Since industry accounts mainly for patent output (see also table 7, appendix), and universities and public research institutions produce most scientific publications, this suggests that industry may be the primary driver of inter-cluster and global knowledge exchanges. The regression analysis provides further insight into the role of knowledge networks on innovation performance.

Regression Analysis
As explained in section 3, the linear regression is carried out based on five sub-models: agglomeration, MNC presence, global knowledge network, local cluster characteristics and agglomeration-network effects. An overview of the regression analysis results is given in table 3.

The agglomeration sub-model ("Model A") regression results suggest that cluster size and regional

Zealand, Singapore and Taiwan.

5 The EU-15 consists of the fifteen "core" countries that joined the European Union in the year 1995 or earlier. They include France, Germany and the United Kingdom.

innovation output may have a positive influence on innovation performance, and therefore hypotheses H1 and H2 are supported. However the explanatory power of the model is relatively low (R^2: 0.10).

The MNC presence sub-model, which is based on inventor to assignee knowledge flows, does not yield any statistically significant results, for both the simple and weighted indicators (*FLWS*, *FLWW*, R^2: 0.00). Therefore hypothesis H3 is not supported.

The global knowledge network sub-model is divided into a regression of simple network indicators ("Model GS") and weighted network indicators ("Model GW"). The simple indicators are stronger predictors of innovation performance (R^2: 0.24) than the weighted indicators (R^2: 0.15).[6] Although diversity of the co-inventor and co-assignment networks correlate positively to innovation performance, diversity in the scientific co-authorship network has a negative correlation. Therefore hypothesis H4 is only partially supported. Hypothesis H5 too, is only partially supported, with a negative correlation the intensity of the scientific co-authorship network and innovation performance, but no statistically significant results for the other weighted network indicators. Thus hypotheses H7 and H8 are partially supported, while hypothesis H9 is rejected.

The rejection of hypothesis H9 is also connected to the rejection of hypotheses H10 and H11. There appears to be a negative correlation between "scientific" indicators and innovation performance as the scientific co-authorship network indicators (*SCNS*, *SCNW*), the relative scientific base (*SCIB*) and the Triple Helix share for patents (*THSP*) all correlate negatively to innovation performance, both in the global knowledge networks sub-model ("Model GS" and "Model GW") and in the local cluster characteristics sub-model ("Model L"). This is surprising because access to scientific knowledge should benefit local knowledge-intensive industry, and not lower its innovation performance.

This "anomaly" can be understood if one considers the fact that the relative scientific base (SCIB) declines with the researcher population of a cluster (*RESR*); the two indicators have a negative correlation (Pearson's r: -0.49). This is also true when comparing the weighted scientific co-authorship network (SCNW, r: -0.57), the Triple Helix share for patents (THSP, r: -0.20) and the simple scientific co-authorship network (*SCNS*) divided by the researcher population (r: -0.56) to the researcher population. So the four "scientific" indicators are essentially negative proxies for cluster size, and are also closely correlated to each other.

6 *ASNS* is excluded from Model GS because of multicollinearity. If *INNS* is excluded instead, similar results are obtained, with a statistically significant positive correlation of *ASNS*.

Indicators	Model A	Model GS	Model GW	Model L	Model AN
RESR					-0.337*** (-5.76)
RPAT	0.031** (2.12)				
PTLQ	0.026** (2.40)				0.104* (1.94)
INNS		0.002*** (3.84)			
ASNS					
SCNS		-0.005*** (-2.89)			
INNW			0.021 (0.94)		
ASNW			0.001 (0.08)		
SCNW			-0.035*** (-2.97)		
SCIB				-0.052*** (-4.59)	
THSS				0.035 (0.57)	
THSP				-0.022* (-1.76)	
ORGS				0.039 (1.61)	
Constant	-0.003	0.092	0.060	0.112	0.077
R^2	0.10	0.24	0.15	0.38	0.40
DV	*PATC*	*PATC*	*PATC*	*PATC*	*INNW*

Table 3: Multiple linear regression analysis results with coefficients, *, ** and *** denote significance at the 10%, 5% and 1% level, respectively. DV is the dependent variable.

The negative correlation can perhaps be interpreted as a cluster relying *too much* on universities and public research institutions as sources of innovation. Universities and public research institutions are the main actors in the above-mentioned “scientific” indicators, but in terms of innovation performance they under-deliver: in the PV sector patent citation rates of university and public research institution (government or non-profit) assigned patents are lower than for industry-held patents (see table 7, appendix). This does not necessarily mean that university or government-produced patents are inferior, because firms will often appropriate the best patents if they fund university research (Gautam et al., 2014). Rather, it suggests that industry R&D is relatively small compared to the scientific resources available in the cluster, and that small clusters size is due to constraints other than these.

Although not statistically significant (but almost, p-value: 0.88), the average organization size (*ORGS*) appears to have a positive influence on innovation performance, thus providing some support for hypothesis H12.

Finally, the agglomeration-network interaction model (“Model AN”) regression results suggest that cluster size (*RESR*) correlates negatively to the weighted inventor network degree centrality (*INNW*), confirming hypothesis H6. Furthermore, clusters that account for a relatively large share of regional patent output, thus having a high patent location quotient (PTLQ), tend to correlate positively to the network indicator, as presumably the region provides fewer opportunities for collaboration, necessitating inter-cluster relations.

Summarizing the results: agglomeration effects and network diversity both correlate positively to innovation performance, with smaller clusters that lack the benefits of agglomeration having more network connections than larger clusters. Unexpectedly, the relative size of the scientific base and related indicators are proxies for the smallness of a cluster.

Cluster Analysis

The relationship between cluster size and connectedness to global knowledge networks, and their positive influence on innovation performance has important implications for cluster development. Figure 4 shows the positioning of clusters in terms of size and connectedness, with clusters above the fitted value line being *more* connected than their size would suggest, and clusters below the fitted value line being *less* connected. Note that the cluster naming is based on the largest population center in the cluster, which may not always coincide with the center of gravity of innovation activity.

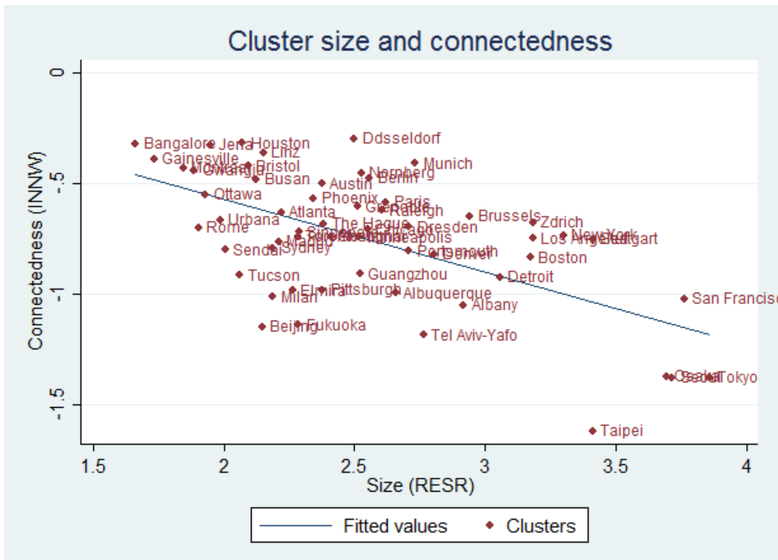

Figure 4: Cluster size and connectedness; for indicator details see table 2.

To further understand the effects of size and under- and highly-connectedness, six extreme clusters of different sizes are chosen. Highly-connected clusters San Francisco, Brussels and Busan are compared to under-connected clusters Taipei, Milan and Tucson. A summary of these six clusters, describing their size and global share of PV innovation output, as well as major organizations and the inter-cluster network is provided in table 8 (appendix).

Among the six clusters, the strongest innovation performance is from San Francisco (0.19 citations per researchers), followed by Tucson (0.13 citations per researcher). The weakest innovation performance was in Milan (0.049 citations per researcher). The other three clusters have a similar innovation performance (between 0.082 and 0.089 citations per researcher).

Milan is also unique in the sense that no organization is listed among the top five patenting *and* the top five scientific publishing organizations. Most other clusters have an organization that appears in both lists, including Brussels (Interuniversity Micro-Electronics Center and Katholieke Universiteit Leuven), San Francsico (Palo Alto Research Center, PARC), Taipei (Taiwan National University) and Tucson (University of Arizona). Busan is similar to Milan, however public research institutions appear in both top-five lists. In the case of Milan, no universities appear to be patenting except the Federal Institute of Technology Lausanne, which is Swiss and thus a partner-institution of Milan-based organizations. The presence of universities as significant patenting and scientific publishing organizations may signal their role in commercializing their research. In the case of San Francisco

PARC is not a university, but it may be fulfilling a similar role as it carries out contract research for industry and government[7].

Brussels, Busan and Taipei all have two or more universities or public research institutions among their top five patenting organization, in contrast to San Francisco, which has none. This suggests that San Francisco, Milan and Tucson have more industry-driven R&D activity.

Among the highly-connected clusters, San Francisco and Busan are primarily connected to other clusters within their respective countries, while Brussels is mainly connected to clusters within Europe. This suggests that these clusters may be more closely integrated into national/European networks. The under-connected clusters of Taipei and Tucson have relatively diverse networks with connections to clusters in Europe, Asia and the USA, although Milan's few connections are only in Europe. Taipei and Tucson may therefore be active in more niche areas, or in the case of Taipei, lack a national network.

6 Discussion and conclusion

Returning to the research questions, the PV sector has undergone rapid growth during the 2006-2010 period, especially in the Asia-Pacific region. Most significant is the expansion of R&D activity in the Asia-Pacific region and the imbalances in terms of patenting and scientific output between regions, and the relative strength of the patent co-invention network compared to the scientific co-authorship network.

Within the context of these macro-trends, agglomeration effects appear to enhance a cluster's innovation performance, with larger clusters in regions with high innovation activity performing best – thus Marshallian (scale) and Jacobian (diversity) agglomeration benefits are both noticeable. Smaller clusters tend to have more connections outside the cluster, which may be a strategy of overcoming agglomeration disadvantages.

A secondary agglomeration effect is also visible in the results, which is the negative influence of local scientific output and related Triple Helix indicators on innovation performance. This result is likely due to the fact that a relatively large scientific base suggests the presence of a small industry cluster, which lacks the benefits of agglomeration, and therefore has lower innovation performance.

While the local presence of MNCs appears to have no influence on cluster innovation performance, the diversity of global knowledge networks, specifically patent co-invention and patent co-assignment networks, correlate positively to innovation performance.

The six cluster cast studies suggest that several cluster typologies may exist in terms of their network and cluster institutions. Some clusters appear to be strongly connected in national or regional networks, while others have a small but diverse number of global connections beyond the nation and region. Although some of these typologies may be due to the cluster's institutions, they could also be related to the specific technologies being developed and the role that the clusters play in the global value chain.

Nevertheless, the results demonstrate the viability of using bibliometric indicators to analyze innovation performance at the global and cluster levels, and extracting a variety of bibliometric and network indicators. However there are naturally also a number of limitations.

[7] See also: http://www.parc.com.

Limitations of the study

Although this study shows that a bibliometric model of industry clusters and knowledge networks is feasible, and holds great promise for use in other studies in which bibliometrics is one or the only consistent source of data, the study also shows the limitations of this methodology. First of all some knowledge networks, such as the inter-organizational and Triple Helix networks, only appear in very large clusters and so there is a limitation in the study of smaller clusters.

Secondly, this study has only addressed the PV sector. Although a simple and transparent method for cluster identification, the use of bilbiometrics means that only one dimension of the sector, and one dimension of innovation, is being measured. It is therefore likely that the method is blind to clusters which may have significance in terms of production or production-related R&D but not in terms of more general R&D that is eventually published as patents or peer-reviewed scientific publications. Since only the PV sector was explored, different sectors may display significantly different agglomeration and network effects.

Thirdly, many of the indicators, and the way in which they are used, are novel and may therefore be further improved and validated. For example, innovation performance could be measured by including scientific publications, or more narrowly by using only industry patents. The network analysis is also quite simple, and a more detailed structural analysis of the networks that are part of the present data set could yield additional insights, such as the correlation between different networks and the changing position of clusters within them.

Appendix

Due to length constraints, the paper containing the full appendix is available at:
http://stek.in/research/photovoltaics_druid.pdf
Alternatively, please send an e-mail p.e.stek@tudelft.nl to receive a copy.

Table 4: Cluster indicators, raw (not transformed)
Table 5: Triple Helix group identification keywords
Table 6: Multiple linear regression diagnostics test outcomes (n = 56)
Table 7: Citation rates per assignee group, includes co-assigned patents
Table 8: Selected cluster profiles

References

Acs, Z. J., Audretsch, D. B., & Feldman, M. P. (1994). R&D spillovers and innovative activity. *Managerial and Decision Economics*, *15*(2), 131–138.

Alcácer, J., & Zhao, M. (2013). Zooming In: A Practical Manual for Identifying Geographic Clusters. *Harvard Business School Strategy Unit Working Paper.* http://doi.org/dx.doi.org/10.2139/ssrn.2359611

Alkemade, F., Heimeriks, G., Schoen, A., Villard, L., & Laurens, P. (2015). Tracking the internationalization of multinational corporate inventive activity: national and sectoral characteristics. *Research Policy*, *44*(9), 1763–1772. http://doi.org/10.1016/j.respol.2015.01.007

Ambos, T. C., Ambos, B., & Schlegelmilch, B. B. (2006). Learning from foreign subsidiaries: An empirical investigation of headquarters' benefits from reverse knowledge transfers. *International Business Review*, *15*(3), 294–312.

Anselin, L., Varga, A., & Acs, Z. (1997). Local geographic spillovers between university research and high technology innovations. *Journal of Urban Economics*, *42*(3), 422–448.

Asheim, B. T., & Coenen, L. (2005). Knowledge bases and regional innovation systems: Comparing Nordic clusters. *Research Policy*, *34*(8), 1173–1190. http://doi.org/10.1016/j.respol.2005.03.013

Asheim, B. T., & Isaksen, A. (2002). Regional innovation systems: the integration of local "sticky"and global "ubiquitous" knowledge. *The Journal of Technology Transfer*, 1–14. Retrieved from http://link.springer.com/article/10.1023/A:1013100704794

Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American Economic Review*, 630–640.

Audretsch, D. B., Lehmann, E. E., & Wright, M. (2014). Technology transfer in a global economy. *Journal of Technology Transfer*, *39*(3), 301–312. http://doi.org/10.1007/s10961-012-9283-6

Autant-Bernard, C. (2001). Science and knowledge flows: evidence from the French case. *Research Policy*, *30*(7), 1069–1078.

Autio, E., Kenney, M., Mustar, P., Siegel, D., & Wright, M. (2014). Entrepreneurial innovation: The importance of context. *Research Policy*, *43*(7), 1097–1108. http://doi.org/10.1016/j.respol.2014.01.015

Awate, S., Larsen, M. M., & Mudambi, R. (2014). Accessing vs sourcing knowledge: A comparative study of R&D internationalization between emerging and advanced economy firms. *Journal of International Business Studies*.

Barnett, G. A., Park, H. W., Jiang, K., Tang, C., & Aguillo, I. F. (2014). A multi-level network analysis of web-citations among the world's universities. *Scientometrics*, *99*(1), 5–26.

Bathelt, H., Malmberg, A., & Maskell, P. (2004). Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, *28*(1), 31–56.

Bekkers, R., & Bodas Freitas, I. M. (2008). Analysing knowledge transfer channels between universities and industry: To what degree do sectors also matter? *Research Policy*, *37*(10), 1837–1853.

Bloomberg New Energy Finance. (2016). *Clean Energy Investment: Q4 2015 Factpack*. London. Retrieved from http://www.bnef.com

Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies*, *39*(April 2010), 61–74. http://doi.org/10.1080/0034340052000320887

Breschi, S., & Lissoni, F. (2001). Knowledge Spillovers and Local Innovation Systems: A Critical Survey. *Industrial and Corporate Change*, *10*(4), 975–1005.

Breyer, C., Birkner, C., Meiss, J., Goldschmidt, J. C., & Riede, M. (2013). A top-down analysis: Determining photovoltaics R&D investments from patent analysis and R&D headcount. *Energy Policy*, *62*, 1570–1580. http://doi.org/10.1016/j.enpol.2013.07.003

Brookings Institution. (2015). *The growing distance between people and jobs in metropolitan America*. Washington, DC. Retrieved from http://www.brookings.edu/~/media/research/files/reports/2015/03/24-job-proximity/srvy_jobsproximity.pdf

Bukvova, H. (2010). Studying Research Collaboration: A Literature Review, *10*.

Camagni, R., & Capello, R. (2002). Milieux Innovateurs and Collective Learning: From Concepts to Measurement. In Z. Acs, H. F. de Groot, & P. Nijkamp (Eds.), *The Emergence of the Knowledge Economy SE - 2* (pp. 15–45). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-24823-1_2

Capello, R. (2009). Spatial spillovers and regional growth: a cognitive approach. *European Planning Studies*, *17*(5), 639–658.

Castellani, D., Jimenez, A., & Zanfei, A. (2013). How remote are R&D labs[quest] Distance factors and international innovative activities. *J Int Bus Stud*, *44*(7), 649–675. Retrieved from http://dx.doi.org/10.1057/jibs.2013.30

Chesbrough, H. (2006). Open Innovation : A New Paradigm for Understanding Industrial Innovation.

Cohendet, P., & Amin, A. (2006). Epistemic communities and communities of practice in the knowledge-based firm1. In *New Frontiers in the Economics of Innovation and New Technology: Essays in Honour of Paul A. David* (p. 296). Edward Elgar Publishing.

Cooke, P. (2001). Regional innovation systems, clusters, and the knowledge economy. *Industrial and Corporate Change*, *10*(4), 945–974.

Cooke, P. (2007). *Regional knowledge economies: markets, clusters and innovation*. Edward Elgar Publishing.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Sy*, 1695. Retrieved from http://igraph.org

D'Este, P., Guy, F., & Iammarino, S. (2012). Shaping the formation of university–industry research collaborations: what type of proximity does really matter? *Journal of Economic Geography*, lbs010.

de la Tour, A., Glachant, M., & Meniere, Y. (2011). Innovation and international technology transfer: The case of the Chinese photovoltaic industry. *Energy Policy*, *39*(2), 761–770. http://doi.org/10.1016/j.enpol.2010.10.050

De Prato, G., & Nepelski, D. (2014). Global technological collaboration network: natwork analysis of international co-inventions. *Journal of Technology Transfer.*

Delgado, M., Porter, M. E., & Stern, S. (2012). Defining Clusters of Related Industries. *NBER Working Papers*.

Delgado, M., Porter, M. E., & Stern, S. (2014). *Defining clusters of related industries*. National Bureau of Economic Research.

Dernis, H., Guellec, D., & Van Pottelsberghe de la Potterie, B. (2001). *USing Patent Counts for Cross-Country Comparisons of Technology Output* (No. 27). Paris. Retrieved from http://www.oecd.org/sti/inno/21682515.pdf

Dewald, U., & Fromhold-Eisebith, M. (2015). Trajectories of sustainability transitions in scale-transcending innovation systems: The case of photovoltaics. *Environmental Innovation and Societal Transitions*.

Dodgson, M. (1992). The strategic management of R&D collaboration. *Technology Analysis & Strategic Management*, *4*(3), 227–244. http://doi.org/10.1080/09537329208524096

Du Plessis, M., Van Looy, B., Song, X., & Magerman, T. (2009). *Data Production Methods for Harmonized Patent Indicators: Assignee sector allocation*. Luxembourg.

Dunning, J. H. (2000). The eclectic paradigm as an envelope for economic and business theories of MNE activity. *International Business Review*, *9*(2), 163–190.

Edquist, C. (1997). *Systems of innovation: technologies, institutions, and organizations*. Psychology Press.

Ernst, D. (2009). A new geography of knowledge in the electronics industry?: Asia's role in global innovation networks.

Ertur, C., & Koch, W. (2011). A contribution to the theory and empirics of Schumpeterian growth with worldwide interactions. *Journal of Economic Growth*, *16*(3), 215–255.

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. *Research Policy*, *29*(2), 109–123. http://doi.org/10.1016/S0048-7333(99)00055-4

Fazio, G., & Lavecchia, L. (2013). Social capital formatio across space: proximity and trust in European regions. *International Regional Science Review, 26*(3), 296–321.

Freeman, C. (1995). The "National System of Innovation"in historical perspective. *Cambridge Journal of Economics*, *19*(1), 5–24.

Frenken, K. (2000). A complexity approach to innovation networks. The case of the aircraft industry (1909–1997). *Research Policy*, *29*(2), 257–272.

Fritsch, M. (2004). Cooperation and the efficiency of regional R&D activities. *Cambridge Journal of Economics*, *28*(6), 829–846.

Frost, T. S. (2001). The geographic sources of foreign subsidiaries' innovations. *Strategic Management Journal*, *22*(2), 101–123.

Frost, T. S., & Zhou, C. (2005). R&D co-practice and "reverse"knowledge integration in multinational firms. *Journal of International Business Studies*, *36*(6), 676–687.

Gautam, P., Kodama, K., & Enomoto, K. (2014). Joint bibliometric analysis of patents and scholarly publications from cross-disciplinary projects: implications for development of evaluative metrics. *Journal of Contemporary Eastern Asia*, *13*(1), 19–37.

Gertler, M. S. (2003). Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography*, *3*(1), 75–99. Retrieved from http://joeg.oxfordjournals.org/cgi/doi/10.1093/jeg/3.1.75

Gertler, M. S., & Levitte, Y. M. (2005). Local nodes in global networks: the geography of knowledge flows in biotechnology innovation. *Industry and Innovation*, *12*(4), 487–507.

Grau, T., Huo, M., & Neuhoff, K. (2012). Survey of photovoltaic industry and policy in Germany and China. *Energy Policy*, *51*, 20–37. http://doi.org/10.1016/j.enpol.2012.03.082

Hagedoorn, J., & Cloodt, M. (2003). Measuring innovative performance: Is there an advantage in using multiple indicators? *Research Policy*, *32*(8), 1365–1379. http://doi.org/10.1016/S0048-7333(02)00137-3

Hervas-Oliver, J.-L., & Albors-Garrigos, J. (2013). *Are Technological Gatekeepers Constraining my Cluster ? Unfolding the paradox of gatekeepers resilience across cluster life cycle stages* (No. 13-04).

Huallacháin, B. Ó., & Lee, D.-S. (2014). Urban centers and networks of co-invention in American biotechnology. *The Annals of Regional Science*, *52*(3), 799–823. http://doi.org/10.1007/s00168-014-0610-8

Iammarino, S., & McCann, P. (2006). The structure and evolution of industrial clusters: Transactions, technology and knowledge spillovers. *Research Policy*, *35*(7), 1018–1036. http://doi.org/10.1016/j.respol.2006.05.004

Jacobs, J. (1970). The economy of cities. *The Economy of Cities*.

Jaffe, A. B. (1989). Real effects of academic research. *The American Economic Review*, 957–970.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics*, *108*(3), 577–598. http://doi.org/10.2307/2118401

Jensen, M. B., Johnson, B., Lorenz, E., & Lundvall, B. Å. (2007). Forms of knowledge and modes of innovation. *Research Policy*, *36*(5), 680–693.

Karlsson, C. (2010). *Handbook of research on cluster theory* (Vol. 1). Edward Elgar Publishing.

Kim, K., & Kim, Y. (2015). Role of policy in innovation and international trade of renewable energy technology: Empirical study of solar PV and wind power technology. *Renewable and Sustainable Energy Reviews*, *44*, 717–727.

Kleinknecht, A., Montfort, K. V. A. N., & Brouwer, E. (2002). THE NON-TRIVIAL CHOICE BETWEEN INNOVATION INDICATORS, *11*(2), 109–121. http://doi.org/10.1080/10438590290013906

Klitkou, A., & Coenen, L. (2013). The emergence of the Norwegian solar photovoltaic industry in a regional perspective. *European Planning Studies*, *21*(11), 1796–1819.

Klitkou, A., & Godoe, H. (2013). The Norwegian PV manufacturing industry in a Triple Helix perspective. *Energy Policy*, *61*(1), 1586–1594. http://doi.org/10.1016/j.enpol.2013.06.032

Kwon, K.-S., Park, H. W., So, M., & Leydesdorff, L. (2012). Has globalization strengthened South

Korea’s national research system? National and international dynamics of the Triple Helix of scientific co-authorship relationships in South Korea. *Scientometrics*, *90*(1), 163–176.

Lanjouw, J. O., & Schankerman, M. (2004). Patent quality and research productivity: Measuring innovation with multiple indicators*. *The Economic Journal*, *114*(495), 441–465.

Leydesdorff, L., Alkemade, F., Heimeriks, G., & Hoekstra, R. (2014). Geographic and Technological Perspectives on “Photovoltaic Cells:” Patents as Instruments for Exploring Innovation Dynamics. *arXiv Preprint arXiv:* …, 1–44. Retrieved from http://arxiv.org/abs/1401.2778

Leydesdorff, L., & Persson, O. (2010). Mapping the geography of science: Distribution patterns and networks of relations among cities and institutes. *Journal of the American Society for Information Science and Technology*, *61*(8), 1622–1634. http://doi.org/10.1002/asi.21347

Lo, C. C., Wang, C. H., & Huang, C. C. (2013). The national innovation system in the Taiwanese photovoltaic industry: A multiple stakeholder perspective. *Technological Forecasting and Social Change*, *80*(5), 893–906. http://doi.org/10.1016/j.techfore.2012.08.016

Locke, R. M., & Wellhausen, R. L. (2014). *Production in the innovation economy*. MIT Press.

Looy, B. Van, Landoni, P., Callaert, J., Pottelsberghe, B. Van, Sapsalis, E., & Debackere, K. (2011). Entrepreneurial effectiveness of European universities : An empirical assessment of antecedents and trade-offs. *Research Policy*, *40*(4), 553–564. http://doi.org/10.1016/j.respol.2011.02.001

Lorentzen, J., & Gastrow, M. (2012). Multinational strategies, local human capital, and global innovation networks in the automotive industry: case studies from Germany and South Africa. *Innovation and Development*, *2*(2), 265–284.

Lundvall, B.-A. (1992). National systems of innovation: An analytical framework. *London: Pinter*.

Luo, S., Lovely, M. E., & Popp, D. (2013). Intellectual Returnees As Drivers of Indigenous Innovation: Evidence From the Chinese Photovoltaic Industry, 40.

Magerman, T., Grouwels, J., Song, X., & Van Looy, B. (2009). *Data Production Methods for Harmonized Patent Indicators: Patentee Name Harmonization*. Luxembourg.

Maggioni, M. A., Nosvelli, M., & Uberti, T. E. (2007). Space versus networks in the geography of innovation: A European analysis*. *Papers in Regional Science*, *86*(3), 471–493.

Maggioni, M. A., & Uberti, T. E. (2009). Knowledge networks across Europe: which distance matters? *The Annals of Regional Science*, *43*(3), 691–720.

Malerba, F., & Orsenigo, L. (1997). Technological regimes and sectoral patterns of innovative activities. *Industrial and Corporate Change*, *6*(1), 83–118.

Mangematin, V., & Errabi, K. (2012). THE DETERMINANTS OF THE SCIENCE-BASED CLUSTER GROWTH : THE CASE OF NANOTECHNOLOGIES THE DETERMINANTS OF

THE SCIENCE - BASED CLUSTER GROWTH : THE CASE OF. *Environment and Planning C: Government and Policy*, *30*(1).

Marshall, A. (1920). *Principles of Economics*. London: MacMillan.

Masciarelli, F., Laursen, K., & Prencipe, A. (2010). Trapped by over-embeddedness: the effects of regional social capital on internationalization. *DRUID Working Papers*.

McCann, P. (2013). *Modern urban and regional economics*. Oxford University Press.

Morgan, K. J. (2004). The Exaggerated Death of Geography: Learning, Proximity and Territorial Innovation Systems, *4*(1), 3–21. http://doi.org/10.1093/jeg/4.1.3

Mudambi, R., & Navarra, P. (2004). Is knowledge power ? Knowledge flows , subsidiary power and rent-seeking within MNCs, 385–406. http://doi.org/10.1057/palgrave.jibs.8400093

Nelson, R. R. (1993). *National Innovation Systems: a Comparative Study*. New York: Oxford University Press.

OECD. (2009). *OECD Patent Statistics Manual*. Paris.

Ooms, J., James, D., DebRoy, S., Wickham, H., & Horner, J. (2015). RMySQL: Database Interface and MySQL Driver for R. Retrieved from http://cran.r-project.org/package=RMySQL

Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *Annals of Mathematical Statistics*, *33*(3), 1065–1076.

Peeters, B., Song, X., Callaert, J., Grouwels, J., & Van Looy, B. (2009). *Harmonizing harmonized patentee names: an exploratory assessment of top patentees*. Luxembourg.

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'Este, P., … Hughes, A. (2013). Academic engagement and commercialisation: A review of the literature on university–industry relations. *Research Policy*, *42*(2), 423–442.

Ponds, R., Van Oort, F., & Frenken, K. (2007). The geographical and institutional proximity of research collaboration*. *Papers in Regional Science*, *86*(3), 423–443.

Ponds, R., van Oort, F., & Frenken, K. (2010). Innovation, spillovers and university-industry collaboration: An extended knowledge production function approach. *Journal of Economic Geography*, *10*(2), 231–255. http://doi.org/10.1093/jeg/lbp036

Porter, M. E. (1998). *Clusters and the new economics of competition* (Vol. 76). Harvard Business Review Boston.

R Core Team. (2015). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from http://www.r-project.org/

Ranga, M., & Etzkowitz, H. (2011). *Triple Helix Systems : An Analytical Framework for Innovation*

Policy and Practice in the Knowledge Society.

Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, *27*(3), 832–837.

Simmie, J. (2003). Innovation and Urban Regions as National and International Nodes for the Transfer and Sharing of Knowledge. *Regional Studies*, *37*(October), 607–620. http://doi.org/10.1080/0034340032000108714

Squicciarini, M., Dernis, H., & Criscuolo, C. (2013). *Measuring Patent Quality Indicators of Technological and Economic Value* (No. 3). Paris. Retrieved from http://www.oecd-ilibrary.org/science-and-technology/measuring-patent-quality_5k4522wkw1r8-en

StataCorp. (2015). *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.

Stek, P. E., & van Geenhuizen, M. S. (2015). The influence of international research interaction on national innovation performance: A bibliometric approach. *Technological Forecasting and Social Change*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0040162515002796

Tödtling, F., & Trippl, M. (2005). One size fits all? *Research Policy*, *34*(8), 1203–1219. http://doi.org/10.1016/j.respol.2005.01.018

van Egeraat, C., Morgenroth, E., Kroes, R., Curran, D., & Gleeson, J. (2015). A Measure for identifying substantial geographic Concentrations. *MRPA Archive*, (65954).

Van Geenhuizen, M., & Nijkamp, P. (2012). Knowledge virtualization and local connectedness among young globalized high-tech companies. *Technological Forecasting and Social Change*, *79*(7), 1179–1191.

Vidican, G., McElvaney, L., Samulewicz, D., & Al-Saleh, Y. (2012). An empirical examination of the development of a solar innovation system in the United Arab Emirates. *Energy for Sustainable Development*, *16*(2), 179–188.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E., Tijssen, R. J. W., Eck, N. J., … Wouters, P. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, *63*(12), 2419–2432.

Wasserman, S., & Faust, C. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.

Webb, C., Dernis, H., Harhoff, D., & Hoisl, K. (2005). *Analysing European and International Patent Citations: A Set of EPO Patent Database Building Blocks* (No. 2005/9). Paris. Retrieved from http://www.oecd.org/sti/35520805.pdf

West, J. (2014). Too little, too early: California's transient advantage in the photovoltaic solar industry. *The Journal of Technology Transfer*, *39*(3), 487–501.

Wickham, H. (2015). rvest: Easily Harvest (Scrape) Web Pages. Retrieved from http://cran.r-project.org/package=rvest

Williamson, O. E. (1981). The Economics of Organization : The Transaction Cost Approach. *The American Journal of Sociology*, *87*(3), 548–577.

Wu, C.-Y. (2014). Comparisons of technological innovation capabilities in the solar photovoltaic industries of Taiwan, China, and Korea. *Scientometrics*, *98*(1), 429–446.

Wuebker, R., Acs, Z. J., & Florida, R. (2010). The Globalization of Innovation and Entrepreneurial Talent: The Changing Context of Venture Capital Investment. *George Mason University School of Public Policy Research Paper*, *6*.

Ye, F. Y., Yu, S. S., & Leydesdorff, L. (2013). The Triple Helix of University-Industry-Government Relations at the Country Level, and its Dynamic Evolution under the Pressures of Globalization. *Journal of the American Society for Information Science and Technology*, 1–21.

Zheng, C., & Kammen, D. M. (2014). An innovation-focused roadmap for a sustainable global photovoltaic industry. *Energy Policy*, *67*, 159–169. http://doi.org/10.1016/j.enpol.2013.12.006