WIFO
ÖSTERREICHISCHES INSTITUT
FÜR WIRTSCHAFTSFORSCHUNG

WORKING PAPERS

Using PageRank in the Analysis of Technological Progress Through Patents

An Illustration for Biotechnological Inventions

Andreas Reinstaller, Peter Reschenhofer

543/2017

ÖSTERREICHISCHES INSTITUT FÜR WIRTSCHAFTSFORSCHUNG
AUSTRIAN INSTITUTE OF ECONOMIC RESEARCH

Using PageRank in the Analysis of Technological Progress Through Patents

An Illustration for Biotechnological Inventions

Andreas Reinstaller, Peter Reschenhofer

WIFO Working Papers, No. 543

October 2017

Abstract

This paper examines whether PageRank algorithms are a valid instrument for the analysis of technical progress in specific technological fields by means of patent citation data. It provides evidence for patent data in biotechnology. Recent literature has been critical with regard to the use of PageRank for the analysis of scientific citation networks. The results reported in this paper indicate, however, that with some minor adaptations and careful interpretation of the results the algorithm can be used to capture some important stylised facts of technical progress and the importance of single patents relatively well especially if compared to indicators based on direct inward citations only.

E-mail addresses: andreas.reinstaller@wifo.ac.at, peter.reschenhofer@wifo.ac.at

2017/260/W/0

Medieninhaber (Verleger), Hersteller: Österreichisches Institut für Wirtschaftsforschung • 1030 Wien, Arsenal, Objekt 20 • Tel. (43 1) 798 26 01-0 • Fax (43 1) 798 93 86 • http://www.wifo.ac.at/ • Verlags- und Herstellungsort: Wien

Using PageRank in the analysis of technological progress through patents. An illustration for biotechnological inventions

Andreas Reinstaller[*,†], Peter Reschenhofer[*]

October 10, 2017

Abstract

This paper examines whether PageRank algorithms are a valid instrument for the analysis of technical progress in specific technological fields by means of patent citation data. It provides evidence for patent data in biotechnology. Recent literature has been critical with regard to the use of PageRank for the analysis of scientific citation networks. The results reported in this paper indicate, however, that with some minor adaptations and careful interpretation of the results the algorithm can be used to capture some important stylised facts of technical progress and the importance of single patents relatively well especially if compared to indicators based on direct inward citations only.

Keywords: patent citations, technological progress, PageRank algorithm, biotechnology

JEL Codes: O31, O32, O34

[*]**Address:** Austrian Institute of Economic Research (WIFO), Arsenal Obj. 20, 1030 Vienna, Austria. E-mail: andreas.reinstaller@wifo.ac.at, peter.reschenhofer@wifo.ac.at

[†]author for correspondence

1 Introduction

Citations patents receive from other patents have widely been used as an indicator for their quality and commercial value in the economic literature. The underlying logic is that when a patent is cited this constitutes a reference to prior art and a knowledge spillovers from the cited to the citing patent. Hence, the more a patent is cited, the more impact it has on other inventions and this in turn goes along with increasing revenues from its commercial exploitation (cf. Trajtenberg, 1990).

In spite of several limitations of patent citations as indicator for knowledge spillovers and patent quality identified in recent years (see Jaffe and de Rassenfosse, 2016, for a comprehensive overview), the relationship between patent citations and patent value (cf. Harhoff et al., 1999), firm value (cf. Hall et al., 1991, 2005), innovation performance (cf. Katila, 2000; Hagedoorn and Cloodt, 2003) and strategic behaviour (cf. Lanjouw and Schankerman, 2001; Bessen, 2008; Abrams et al., 2013) has been thoroughly examined. Most contributions in these strands of research provide robust evidence that there is real economic value associated to patent citations.

One of the points of critique advanced against patent citations is that they suffer from the same problem as simple patent counts, which they are supposed to correct: while the number of citations to a patent is taken as a measure for quality, this count alone is not informative as to the quality of the citing patents. It should be assumed that a patent is of higher quality if it receives its citations from high quality rather than low quality patents, hence indirect citations are an important determinant of patent quality (cf. Atallah and Rodriguez, 2006). If this aspect is not taken into account the patent counts tend to be a relatively noisy indicator for quality.

From this perspective, the celebrated PageRank algorithm which constitutes the core of Google's internet search engine and variations thereof could potentially provide a valid indicator for the quality and the importance of a patented invention. However, recent contributions have questioned the validity of PageRank for scientific citation networks. This paper examines to what extent this critique applies also to patent citation networks when the aim is to characterise technical progress in a specific technological field. It presents an empirical analysis of patent applications in biotechnology as a case example.

Section 2 briefly characterises the data used in the empirical part of the paper. Section 3 provides a discussion of the critique advanced on the use of PageRank in the analysis of scientific citation networks. It presents a simple adaptation of the algorithm that takes on board important points of critique. Section 4 provides empirical evidence that PageRank scores for patents capture important stylised facts of technical progress and the importance of single patents in biotechnology relatively well especially if compared to indicators based on direct inward citations. Section 5 finally discusses the pros and cons of the use of PageRank based indicators for the importance of patented inventions.

2 Data

Modern biotechnology is rooted in genetics. As a technology field it has developed out of some key inventions, such as recombinant DNA, and has a relatively well identifiable starting point. For this reason, we analyse patents in the field of biotechnology. We rely on the PATSTAT database of the European Patent Office (EPO) (cf. de Rassenfosse et al., 2014, for an overview). To identify biotechnology patents, we have used a consolidated classification of IPC classes available from the OECD and EuroStat. The detailed list is reported in the appendix.

In section 4.1 we use the complete sample of biotechnology patents from PATSTAT that comprises all patent applications and granted patents. We refer to this as the global sample. We are interested in characterising the general development of the entire technological field using the PageRank algorithm. Therefore, each patent application or granted patent, independently on where it has been filed, has to be taken into account because it constitutes prior art for later patents. We have corrected for multiple filings referring to the same invention (identical priority), however, by pooling them into one observation and consolidating inward and outward citations across filings.

In sections 4.1 and 4.2 we analyse the relationship between PageRank scores and specific patent characteristics, and construct country rankings that capture the contribution of single countries to the development of this technological field over specific time horizons. For this part of the analysis, we rely on patent applications from EPO only to ensure that all patents in specific annual cohorts have been subject to identical legal and administrative procedures and that data are comparable. We refer to this sample as the EPO sample. Additional information on patent grants, patent family sizes, patent renewal, the number of claims and opposition used in section 4.2 were drawn from the related EPO INPADOC database. Citations data for the EPO data include citations of patents between 1978 and fall 2015 (with the the earliest cited patent dating 01-06-1978, and the earliest citing patent dating 25-04-1979).

As the goal of the present analysis is to characterise the overall historical development of a technological field, we do not restrict our analysis to a specific annual cohort of patents. This is usually done in studies relying on patent citations to make data comparable as older patents typically accumulate larger number of citations. Rather, we propose an approach to analyse annual cohorts and changes in rankings of annual cohorts over time in a consistent way.

Table 1: Descriptive Statistics

	Global sample	EPO sample
Patent Count	1,851,766	160,301
Patent Count with Inward Cit.	480,366	33,594
Max. Inward Cit.	6,117	1,049
Avg. Inward Cit.	2.12	1.54
Skewness Inward Cit.	101.45	33.51
Kurtosis Inward Cit.	27,923.74	2,349.97
Patent Count with Outward Cit.	595,940	75,818
Max. Outward Cit.	709	709
Avg. Outward Cit.	2.12	2.49

Table 1 presents some descriptive statistics for the global and EPO samples. The table shows that between twenty (global) and twenty five percent (EPO) of all patents do not receive any citations from later patents. The average number of inward citations is very low (2.12 and 1.54 respectively) and the maximum number of inward citations lies considerably above this average. The skewness and kurtosis statistics of the distribution of inward citations in biotechnology patents shows that it has fat tails on the right side. This distribution therefore produces more outliers, than if inward citations were normally distributed, while the mass of the observations is concentrated in the lower ranges of inward citations. In the calculation of PageRank scores, both inward and outward citations of a patent are used.

The summary statistics for outward citations show that the share of patents not citing earlier patents in both distributions is higher than for inward citations, i.e. many patents do not refer to prior art, but on average about two earlier sources get cited in a patent.

3 Indicators

As has been outlined in the introduction, simple citation weighted counts of patent applications or granted patents in a technological field are an incomplete measure for the quality of technological activity by a specific unit of observation such as a company or a country. The application of the famous PageRank algorithm (Brin and Page, 1998) to patent citation networks would allow taking recursively into account indirect citations when ranking patents. [1]

However, as Walker et al. (2007) and Chen et al. (2007) argue, the PageRank algorithm is blind to the age structure in citation networks. Its application to patents irrespective of their filing or priority dates will therefore lead to a historical ranking of the most important patents in a technology field, but it will not reflect the up-to-date relevance of patents. The resulting ranking will therefore be misleading for the simple reason that engineers and researchers when working on own inventions typically look at the most recent developments in a technology field and not at historical ones upon which their work is likely to build anyway (cf. Maslov and Redner, 2008). Walker et al. (2007) therefore propose an alternative measure called CiteRank which will be discussed in detail later.

[1] Shaffer (2011) has introduced a recursive algorithm to identify the technological significance and economic value of a patent he refers to as Patent Rank. It is mathematically equivalent to PageRank (cf. Langville and Meyer, 2003). For a discussion see Bruck et al. (2016).

More recently, Mariani et al. (2015) have argued that citation networks grow constantly over time. In this case the simple PageRank algorithm will be biased. If the decay of relevance of a patent is slow, then there will be only old patents in the top percentiles of the ranking produced by PageRank. The algorithm produces an "all-time" best list of inventions in a technology field. If the decay of relevance is fast, in contrast only recent Patents will occupy the top percentile which ignores the importance of older contributions. At fast decay rates a simply ranking according to inward citations will provide an unbiased measure of relevance given that they capture processes of preferential attachment and growth. This leads Mariani et al. (2015) to conclude that a static PageRank algorithm is an inappropriate measure to capture relevance in growing citation networks and that time-dependent algorithms based on temporal linking patterns should better be able to rank patents.

These concerns are valid if the principal aim of the ranking of scientific publications or patents is to identify emergent scientific or technological fields and the most important or relevant contributions in the recent past (possibly to assess the performance and importance of researchers or inventors in their current career stage). Key characteristics of emergent fields are novelty and growth. In this case the number of direct inward citations a patent acquires over some period of time is an appropriate measure (considering that patents enshrine some novelty by definition). However, when the goal of the ranking is to analyse the development of a technological field and the performance of some unit of observation in it over time, then identifying emergent fields and significant contributions in the recent past is an important but not the primary criterion of interest.[2]

[2] In the context of patent analysis scholars have used extreme value statistics to identify "superstar" patents, (cf. Silverberg and Verspagen, 2007; Castaldi et al., 2014).

Technical progress has a number of additional characteristics a ranking capturing the importance of an invention for the development of a specific technological field should ideally reflect. One important characteristics is its cumulativeness and thus the clustering of technical knowledge in specific subdomaims of knowledge in broader technological fields. In a recent paper Acemoglu et al. (2016), for instance, confirm earlier findings by means of a comprehensive analysis of US patent records, and show that technological progress is highly cumulative and the expansion in one technological field drives future work in linked fields. They also provide evidence that there will be more inventive activity in one field, if it can build on more past inventions in related fields. In this process different technological trajectories emerge that later bifurcate or merge (cf. Verspagen, 2007).

In addition, one can observe an important clustering of inventive activities, where many minor and intermediate inventions are the consequences of major paradigmatic ones. On the one hand, this means that the 'size' or importance of an invention measured for instance through inward citations is drawn from a highly skewed and possibly fat-tailed distribution (cf. Silverberg, 2002; Valverde et al., 2007). On the other hand, it implies that technical progress in one domain is linked to earlier advances in related domains. Hence, earlier significant contributions are still important for current developments, even if they no longer get cited. New inventions build on the state of the art defined by past achievements. This is generally referred to as "standing on the shoulders of giants". All these aspects should be reflected in a time consistent ranking of inventive activity both at the level of single patents and in cross-country comparisons of inventive activity.

Finally, given the cumulativeness of technological development such a ranking should also capture the degree of novelty or the innovativeness of the patents considered in the ranking. While inward citations to a patent capture its importance for later patents, they do not tell us to what extent the invention it protects, draws on prior art. In scientific publications outward citations are often relatively selective and up to the discretion of the authors. In patents they are more meaningful in capturing prior art. Here outward citations get often inserted into a filed patent by patent examiners in order to delimit the exact scope and inventive step of an invention (or more precisely the claims in the patent) relative to prior art, and to correct in part for firm specific patenting behaviour.

As argued by Alcacer et al. (2009) this may introduce measurement bias, if the aim is to use outward citations to represent knowledge flows between inventions correctly. However, this critique is not valid if the aim is to measure the importance of a patent relative to prior art. In this case the goal is to judge the overall importance of a patent for the development in a specific technological field over time relative to other inventions. As a consequence, this information should be taken into account in the construction of indicators aiming at measuring the importance of an invention for the development of a technological field.

If one considers these aspects, then the PageRank approach remains an appropriate starting point for the construction of an indicator of patent quality despite the valid criticism on its age bias. The PageRank score x_i of a patent i is defined as:

$$\mathbf{x}^{\mathrm{PR}} = \alpha \mathbf{A} \mathbf{D}^{-1} \mathbf{x}^{\mathrm{PR}} + \beta \mathbf{1} \tag{1}$$

where $\mathbf{A}$ is the adjacency matrix with $A_{ij} = 1$ iff there is a link from patent j to patent i. The score $x_i \in \mathbf{x}$ of a patent i increases in the number of inward links it receives from patents j. As citations go from j to i but not the other way round, the graph defined by the adjacency matrix $\mathbf{A}$ is acyclic, furthermore if the columns and rows are ordered chronologically the $\mathbf{A}$ is an upper triangular matrix. The diagonal vector $\mathbf{D}$ defined as: $D_{ii} = \max\left(k_j^{\text{out}}, 1\right)$ and $D_{ij} = 0$ if $i \neq j$, introduces outward citations as a weighting factor. Any x_i in equation (1) is therefore defined through the weighted score x_j/k_j^{out} of all patents j citing it plus the constant β. Hence, for any given number of inward citations to patent j the weight it passes on to later patents i decreases proportionally to its outdegree.

Factor α establishes the importance of indirect citations to a patent i in line with the path length of indirectly connected patents. It is a dampening factor that geometrically discounts the impact of an indirectly citing patent on a cited patent over the length of the path between these nodes. With larger values of α indirect citations by patents that are more distant in the citation network get a higher weight, whereas for values close to zero only patent i's direct inward linkages influence the score. In the literature this parameter is typically set between 0.5 and 0.85.

In a recent paper, Bruck et al. (2016) show that for patent citation networks the optimal value of α is close to 0.5, whereas 0.85 does not lead to acceptable ranking results. Similar results are reported by Ding et al. (2009) for academic papers. The reason for this is that at higher values for α the peripheral parts of the citation network receive too much weight relative the more important central ones, which as a consequence become underrated (cf. Boldi et al., 2005; Avrachenko et al., 2008).

Table (2) shows for biotech patents filed from 2000 onward lying in the top 1 % inward citation percentile that the Spearman rank correlation for rankings produced with different values for α is on average highest for values

Table 2: Rank Correlation for rankings obtained from different values of α. The table shows results for patents from year 2000 onward in the top 1 % inward citation percentile

α	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95	Avg. Corr.
0.05	1.0000	0.9896	0.9603	0.9138	0.8530	0.7815	0.7019	0.6196	0.5403	0.4687	0.7829
0.15	0.9896	1.0000	0.9886	0.9565	0.9075	0.8455	0.7736	0.6970	0.6216	0.5524	0.8332
0.25	0.9603	0.9886	1.0000	0.9880	0.9553	0.9063	0.8447	0.7759	0.7060	0.6405	0.8766
0.35	0.9138	0.9565	0.9880	1.0000	0.9883	0.9563	0.9084	0.8501	0.7879	0.7278	0.9077
0.45	0.8530	0.9075	0.9553	0.9883	1.0000	0.9888	0.9579	0.9129	0.8607	0.8077	0.9232
0.55	0.7815	0.8455	0.9063	0.9563	0.9888	1.0000	0.9893	0.9606	0.9207	0.8769	0.9226
0.65	0.7019	0.7736	0.8447	0.9084	0.9579	0.9893	1.0000	0.9904	0.9656	0.9331	0.9065
0.75	0.6196	0.6970	0.7759	0.8501	0.9129	0.9606	0.9904	1.0000	0.9919	0.9725	0.8771
0.85	0.5403	0.6216	0.7060	0.7879	0.8607	0.9207	0.9656	0.9919	1.0000	0.9939	0.8389
0.95	0.4687	0.5524	0.6405	0.7276	0.8077	0.8769	0.9331	0.9725	0.9939	1.0000	0.7973

of α close to 0.5. This indicates that around this value the rank order varies least and ranks are most consistent and stable across specifications.The rank correlations are also slightly higher for parameter ranges between 0.55 and 0.95 indicating that the ranking obtained with the parameter set to 0.55 sufficiently takes into account also peripheral parts of the network, while giving more weight to the central ones. The results in Table (2) do not change when one observes a longer or shorter time horizon or broadens or narrows the sample with regard to the inward citation percentile. The parameter will therefore be set to 0.5 for all analyses in this paper.

Finally, β is an additional, network independent observation based weight that in the standard PageRank approach is normally set to one, if no suitable network independent information is available. Solving for $\mathbf{x}^{\mathrm{PR}}$ equation (1) can then be reformulated as:

$$\mathbf{x}^{\mathrm{PR}} = \left(\mathbf{I} - \alpha\mathbf{A}\mathbf{D}^{-1}\right)^{-1}\mathbf{1} = \mathbf{D}\left(\mathbf{D} - \alpha\mathbf{A}\right)^{-1}\mathbf{1} \tag{2}$$

CiteRank (Walker et al., 2007) is now an alternative measure of centrality in a citation network to overcome the inherent age bias of PageRank. To this end it introduces a weight $p_i^{\mathrm{CR}} = e^{\mathrm{age}_i/\tau_i}$, which is drawn from a traffic model capturing the probability that an inventor of later patents "stumbles across" patent i:[3]

$$\mathbf{x}^{\mathrm{CR}} = \sum_{i=0}^{\infty} \left(\alpha \mathbf{A}\mathbf{D}^{-1}\right)^{i} \mathbf{p}^{\mathrm{CR}} = \left(\mathbf{I} - \alpha \mathbf{A}\mathbf{D}^{-1}\right)^{-1} \mathbf{p}^{\mathrm{CR}}$$
$$= \mathbf{D}\left(\mathbf{D} - \alpha \mathbf{A}\right)^{-1} \mathbf{p}^{\mathrm{CR}} \tag{3}$$

CiteRank thus ranks down historically important contributions, as the weight p_i decreases exponentially with patent age proportional to a decay parameter τ_i. Hence, if the unit of observation is an inventor or a country and the goal is to establish the direct influence of their older inventions on current inventive activities in a technological field, CiteRank provides a consistent ranking on the most recent time margin.

If the goal of the ranking is to identify the units of observations responsible for the most important inventions in a technological field and how their contribution to the development of the technological field has changed over time (building cumulatively on earlier inventions) across units of observation, CiteRank will not provide the best suited indicator, as earlier paradigmatic patents are unlikely to rank high in CiteRank for a specific period t despite their importance for later patents. In addition, parameter τ_i is generally difficult to determine for each observation. It may be specific to technological fields and also change over time. This introduces an inaccuracy and discretion in CiteRank.

To obtain a time consistent ranking of patents, the vector of weights can be used to construct a moving or sliding time window over annual cohorts. Instead of an exponential weighting, we use binary weights with ones for patents newer than a threshold date d, and zeros for older patents:

$$\mathbf{x}^{\mathrm{SW}(d)} = \mathbf{D}\left(\mathbf{D} - \alpha \mathbf{A}\right)^{-1} \mathbf{p}^{\mathrm{SW}(d)} \tag{4}$$

[3]Walker et al. (2007) define CiteRank as $\mathbf{x}^{\mathrm{CR}} = \mathbf{I}\mathbf{p}^{\mathrm{CR}} + \alpha \mathbf{W}\mathbf{p}^{\mathrm{CR}} + \alpha^2 \mathbf{W}^2 \mathbf{p}^{\mathrm{CR}} + \cdots = \sum_{i=0}^{\infty} \left(\alpha \mathbf{W}\right)^{i} \mathbf{p}^{\mathrm{CR}}$ with $\mathbf{W} = \mathbf{A}\mathbf{D}^{-1}$ in the notation of this paper.

with

$$p_i^{\mathrm{SW}(d)} = \begin{cases} 1 & \text{if } p_i \text{ is newer than } d \\ 0 & \text{otherwise} \end{cases}$$

Note that if the patents are ordered chronologically, the matrix $\mathbf{D}\,(\mathbf{D} - \alpha\mathbf{A})$ is also an upper triangular matrix and the vector $\mathbf{p}^{\mathrm{SW}}$ consists of two blocks, zeros until the row $i-1$ where i is first patent that is newer than the threshold date d. All other rows $\{i, i+1, \ldots, n\}$ are ones with n denoting the number of patents, and therefore $\mathbf{x}_j^{\mathrm{SW}} = \mathbf{x}_j^{\mathrm{PR}}$ for $i \le j \le n$ holds.[4] This implies that time consistent rankings can now be obtained by calculating the PageRank algorithm over the whole patent sample using equation (2) once, and evaluate only patent scores newer than d to construct rankings for the units of observations.[5]

In the next section, we assess the sensitivity of the moving time window PageRank approach to the weighting with the patent's outdegree both for patent as well as the initial weight the patent has calculation of the scores.[6] For this purpose, we calculate alternative indicators in moving time windows. Centrality measures that are frequently used for the analysis of directed acyclic graphs are Alpha centrality (Bonacich and Lloyd, 2001) and Katz centrality (Katz, 1953). Alpha centrality is essentially a PageRank

[4]The inverse matrix of the upper triangular matrix

$$\mathbf{I} - \alpha\mathbf{A}\mathbf{D}^{-1} = \begin{pmatrix} 1 & -\frac{\alpha A_{1,2}}{D_{2,2}} & \cdots & -\frac{\alpha A_{1,n}}{D_{n,n}} \\ 0 & \ddots & & \vdots \\ 0 & \cdots & 1 & -\frac{\alpha A_{n-1,n}}{D_{n,n}} \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

can be simple calculated from bottom up, so that for the calculation of row i of the inverse only the rows $j \ge i$ are needed.

[5]Another approach to get a time consistent PageRank would be the calculation of the PageRank on a subsample of the patents from patents newer than the date d. If in that calculation the matrix $\mathbf{D}$ is taken from the original graph (meaning that the count of outgoing citations to patents older than the threshold date d are still considered) and sliced to the dimensions of the subsample adjacency matrix $\mathbf{A}$ we obtain an identical ranking.

[6]Assigning a weight to each patent is necessary to obtain meaningful centrality measures in acyclic graphs.

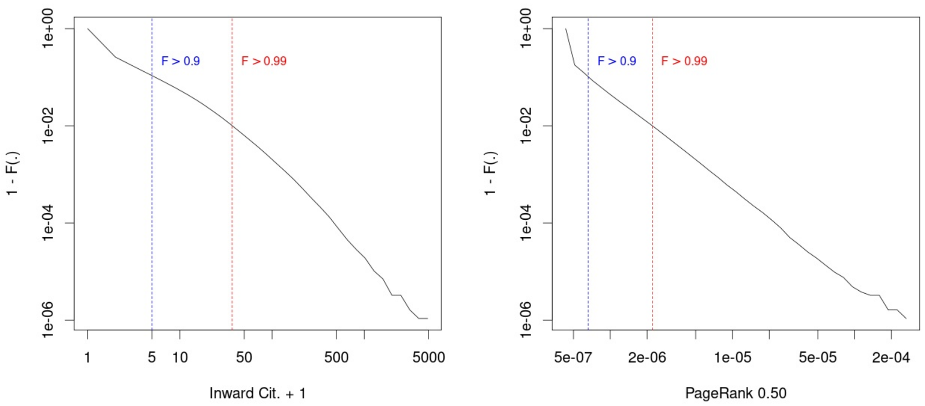
centrality without taking into account the outdegree of patents in the calculation of the score. Katz centrality is largely identical to Alpha centrality, with the exception that the initial weight attached to each patent in the network is subtracted from the final score. This scales down countries in a ranking, when they have a large number of patents with zero inward citations. Our examinations show that Katz and Alpha centrality do not yield qualitatively different results. For this reason, in what follows, we compare only PageRank with inward citations and Alpha centrality. We provide a technical description of Alpha centrality in the appendix.

4 Analysis

4.1 PageRank scores for patent citations: The importance of inventions and clustering in time

We first examine Pareto plots for the EPO citation data. These plots represent right cumulative distributions of the number of observations plotted on the ordinate with a value greater than or equal to a specific amount (count or score) plotted on the abscissa on a log-log scale. A Pareto or power law distribution would appear as a linear downward sloping line in such a plot. Figure 1a shows the Pareto plot for the number of citations for the global sample of patents in biotechnology.[7] The figure is in line with earlier findings (e.g. Scherer and Harhoff, 2000) that have shown, that the cumulative distribution of citations shows some slight curvature with the tail of the distribution being rather linear and extensive. While this is evidence that the

[7]In the appendix we present Pareto plots for the EPO sample. Qualitatively the differences between the distribution of inward citations and PageRank scores are largely identical to the ones reported here for the global sample. However, the Pareto plot for PageRank scores for the EPO sample shows a slight downward sloping curvature for extreme values indicating that extreme outliers are less frequent in the EPO sample than in the global sample.



(a) Distribution of inward citations (b) Distribution PageRank scores

Figure 1: Authors' calculations based on the global sample of patents in biotechnology extracted from the EPO PATSTAT database. Vertical lines indicate 90% and 99% of total number of patents in the sample.

data are highly skewed, the literature has argued that it is not clear whether they are better represented by a fat-tailed distribution such as a Pareto or by medium-tailed distributions such as the log-log.

Figure 1b shows the Pareto plot for the PageRank scores for the same sample of patents. The cumulative distribution of citations is clearly linear with the tail of the distribution again being rather linear and extensive. If indirect citations play indeed an important role in the determination of patent quality, this evidence supports the findings of Silverberg and Verspagen (2007) who have argued that the evidence of heavy tails in the distribution of patent citations has significant implications for technology policy insofar, as it hints at interdependent realisation of invention outcomes and thus clustering processes in inventive activity in time where many of minor and intermediate inventions are the consequences of major paradigmatic ones.

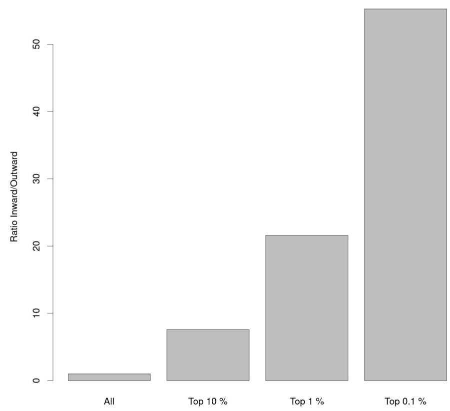


Figure 2: Ratio of inward to outward citations by inward citation percentiles. Authors' calculations based on the global sample of patents in biotechnology extracted from the EPO PATSTAT database.

Figure 2 provides further evidence for this clustering process in time. It shows the ratio of inward to outward citations of the patents in our sample for different sections of the distribution of PageRank scores. High ranking patents are on average cited by considerably more patents than they cite themselves, or in other words, they constitute prior art to considerably more subsequent patents than they themselves draw on prior art from earlier patents. This ratio steadily increases as one moves through the sections of the PageRank distribution.

If one looks at the top 20 patents identified in the distribution of PageRank scores (executed on all patents without moving time window) in Table 3, one can identify, for example, the paradigmatic patents of Cohen/Boyer (1974), Mullis(1985) or Ehrlich et al. (1986) that were key inventions enabling the development of modern biotechnology and the emergence of the biotechnology industry with the former covering the principles of recombinant DNA and the latter ones PCR technology.[8] PageRank is thus able to

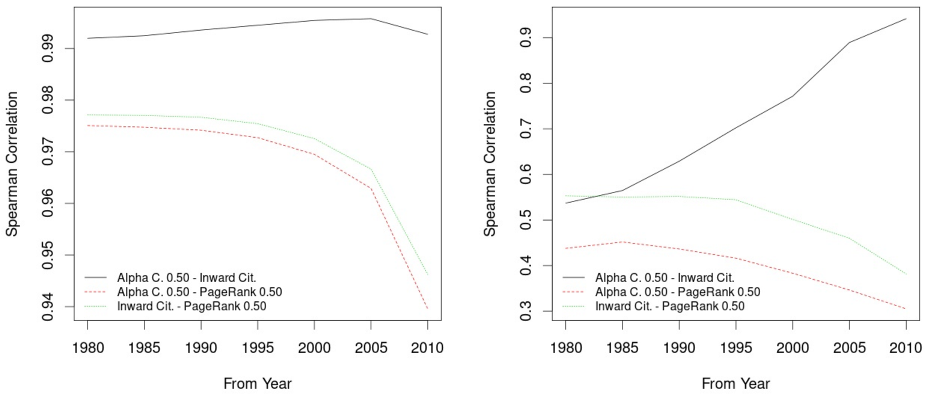
[8] In the appendix we list the top 20 patents in biotechnology for EPO applications only.

Table 3: Top 20 patents in biotechnology in the global sample from the EPO PATSTAT database

Patent Nr.	Prio. Date	Title	Inventors	R. PR50	R. Inward
US19940265483	1994-06-24	Mutant dwarfism gene of petunia	Holtrop, Mathilde H. (US) / Hunsperger, Mary H. (US)	1	4
US19860828144	1986-01-30	Process for amplifying, detecting, and/or-cloning nucleic acid sequences	Erlich, Henry A. (US) / Saiki, Randall K. (US) / Mullis, Kary B. (US) / Horn, Glenn T. (US) / Arnheim, Norman (US) / Scharf, Stephen J. (US)	2	2
US19850791308	1985-03-28	Process for amplifying nucleic acid sequences	Mullis, Kary B. (US)	3	1
US19810294227	1981-08-19	Method for the direct analysis of sickle cell anemia	Geever, Robert F. (US) / Wilson, John T. (US) / Wilson, Lois B. (US)	4	4301
US19940328962	1993-04-07	Method of screening for cDNA encoding novel secreted mammalian proteins in yeast	Jacobs, Kenneth (US)	5	23
US19790001021	1974-11-04	Process for producing biologically functional molecular chimeras	Boyer, Herbert W. (US) / Cohen, Stanley N. (US)	6	294
US19800213803	1980-12-08	Specific DNA probes in diagnostic microbiology	Falkow, Stanley (US) / Moseley, Stephen L. (US)	7	104
US19790035619	1978-05-04	Diagnostic reagent	Parikh, Indu (US) / Cuatrecasas, Pedro (US)	8	3103
US19970825469	1997-03-28	Soybean transformation and regeneration methods	Reichert, Nancy A. (US) / Dan, Yinghui (US)	9	51
US19900492462	1989-06-07	LARGE SCALE PHOTOLITHOGRAPHIC SOLID PHASE SYNTHESIS OF POLYPEPTIDES AND RECEPTOR BINDING SCREENING THEREOF	Fodor, Stephen P. A. (US) / Pirrung, Michael C. (US) / Read, J. Leighton (US) / Stryer, Lubert (US)	10	7
US19830483457	1983-04-08	Recombinant immunoglobin preparations	Heyneker, Herbert L. (US) / Riggs, Arthur D. (US) / Holmes, William E. (US) / Cabilly, Shmuel (US) / Wetzel, Ronald B. (US)	11	3
US19720260488	1972-06-07	MICROORGANISMS HAVING MULTIPLE COMPATIBLE DEGRADATIVE ENERGY-GENERATING PLASMIDS AND PREPARATION THEREOF		12	21608
US19680762120	1968-09-24	METHOD FOR THE DETERMINATION OF ANTIGENS AND ANTIBODIES	ANTONIUS HERMANUS WILHELMUS MARIA SCHUURS (NL) / BAUKE KLAAS VAN WEEMEN (NL)	13	509
US19870045890	1986-05-02	Secretory signal selection vectors for extracellular protein synthesis bacilli	Bron, Sierd (NL) / Peeters, Ben P. H. (NL) / Smith, Hilde E. (NL) / Van Ee, Jan H. (NL) / Venema, Gerard (NL)	14	39757
US19970835770	1993-10-18	Antisense oligonucleotide modulation of multidrug resistance-associated protein	Bennett, C. Frank (US) / Dean, Nicholas M. (US) / Baracchini, Edgardo (US)	15	43
US19850727503	1985-04-26	Protogyny in Zea mays	Orr, Peter M. (US) / Armond, Paul A. (US)	16	115
EP19840302368	1983-04-08	Recombinant immunoglobulin preparations	Cabilly, Shmuel (US) / Holmes, William Evans (US) / Wetzel, Ronald Burnell (US) / Heyneker, Herbert Louis (US) / Riggs, Arthur Dale (US)	17	87
US19890299617	1987-09-21	Single polypeptide chain binding molecules	Ladner, Robert C. (US) / Bird, Robert E. (US) / Hardman, Karl (US)	18	6
US19870063647	1986-08-22	Process for amplifying, detecting, and/or cloning nucleic acid sequences using a thermostable enzyme	Mullis, Kary B. (US) / Erlich, Henry A. (US) / Gelfand, David H. (US) / Saiki, Randall K. (US) / Horn, Glenn (US)	19	8
US19930068540	1991-09-04	Method and apparatus for conducting an array of chemical reactions on a support surface	Brennan, Thomas M. (US)	20	39

identify these major paradigmatic inventions in a technology field. From the table it is also evident, that the ranking of patents in a technology field may differ sometimes considerably depending on whether one uses inward citations or PageRank scores. Patents with a high number of direct citations cited by relatively unimportant patents would rank considerably higher in an inward citation based ranking, whereas patents with fewer but more important citations (in terms of the number of citations of the citing patents) would rank lower. PageRank corrects for this problem. On the other hand, the age bias for which the PageRank algorithm has been criticised, clearly emerges in Table 3. It ranks the historically paradigmatic patents high, whereas recent important developments will typically rank low as has been argued by Walker et al. (2007).

Figure 3 presents Spearman rank correlations between (global) patent rankings for bio tech patents over time obtained from inward citations, the PageRank and Alpha centrality. In this way, we assess to what extent these rankings are sensitive to indirect citations and weighting through outward citations. As explained earlier, Alpha centrality is identical to PageRank without taking into account the outdegree k_j^{out} of patents in the calculation of the score. Hence, the comparison between the ranking resulting from inward citations and the rankings obtained from PageRank (Inward Cit. - PageRank 0.50) and Alpha centrality (Alpha C. 0.50 - Inward Cit.) shows to what extent the ranking is influenced by indirect citations, whereas the Spearman rank correlation between the rankings obtained from PageRank and Alpha centrality (Alpha C. 0.50 - PageRank 0.50) indicate how outward citations influence the resulting ranking. The analysis has been carried out for the entire sample (Figure 3a) and in the decile of the sample with the highest scores of each indicator (Figure 3b).



(a) All patents

(b) Top decile

Figure 3: Spearman rank correlation between patent rankings constructed from different network centrality measures. Authors' calculations based on EPO PATSTAT data.

Figure 3a shows that the rankings obtained from inward citations and Alpha centrality almost perfectly correlate. Alpha centrality and inward citation based rankings instead correlate strongly with PageRank based ones, but the closer time windows moves to more recent periods the weaker it becomes even though it remains at very high levels (> 0.94). However, the high correlation coefficients are determined by the large number of patents with no or very few inward citations, which all rank the same. The correlations in Figure 3a are therefore not informative.

Figure 3b therefore concentrates on the right tail of the distribution where essentially all patents receive more than ten inward citations (see Figure 1a). Now, the influence of the difference between the indicators becomes clearer. The correlation between Alpha centrality and inward citation becomes stronger, the closer the time windows moves to recent periods and approaches values in the order of 0.9. As outward citations are not used in the construction of Alpha centrality and as fewer and fewer patents by definition can get cited by later patents, Alpha centrality and inward citations should converge. This is what we observe.

On the other hand, Alpha centrality and inward citation based rankings correlate weakly with PageRank based ones, and the correlation deteriorates the closer the time window moves to recent periods. Taking into account outward citations significantly affects the ranking. As in our interpretation outward citations are a (rough) proxy for the cumulativeness of an invention. Given that they reflect references to prior art, PageRank based rankings therefore penalises patents citing many earlier sources. PageRank therefore makes a first guess on the radicality of a patent on the basis of its supposed cumulativeness inferred from the number of outward citations. Whether this guess will turn out to be correct, can only be inferred over longer periods of observation, when inward citations and indirect citations can be factored in, which PageRank does. Inward citation and Alpha centrality based rankings ignore this aspect, but are less charged with a-priori interpretations on the meaning of outward citations.

Overall, this first descriptive evidence indicates that PageRank based patent rankings are able to capture clustering processes in inventive activity in time and the importance of inventions better than simple inward citation based ones.

4.2 PageRank and the clustering in knowledge space

As has been argued earlier, technical change is cumulative and new technologies build on prior advances in other technological domains. Specific technological trajectories branch out into new ones, merging earlier knowledge with new knowledge domains. Hence, technical progress goes along with a process of diversification in which new knowledge evolves out of earlier knowledge.

Recent research (cf. Kogler et al., 2013; Acemoglu et al., 2016) has tried to analyse this process by using information on technological classes contained in patents, typically IPC classes, to construct networks of related knowledge domains in a technological field and analyse how these change over time. Kogler et al. (2013) refer to this network as “knowledge space”. In this section we examine the relationship between citation networks and the evolution of the knowledge space. The former is represented by PageRank as well as inward citation scores, the latter by the network of related technological fields in biotechnological inventions.

One should expect that over at the beginning of the development of a technological field, a dense network of interrelated knowledge domains in knowledge space emerges in which paradigmatic patents at first accumulate high direct and indirect citations scores and therefore high PageRank scores. As time goes by, new knowledge domains closely related to the earlier network emerge, i.e. new developments take place in the neighbourhood of more central parts of the knowledge space. Again, some key patents in these new domains should start attracting citations and thus develop high PageRank scores.

To examine this conjecture, we construct first the knowledge space using the global sample of biotechnology patents following the approach by Kogler et al. (2013) It exploits information on the co-appearance of specific technological (IPC) classes across patent applications in a particular year to construct a network of technological proximity or knowledge relatedness across bio tech patents. We construct this knowledge space at the most disaggregated level of IPC classes consisting of 514 subclasses. If now, $I_{ki} = 1$ when patent i lists IPC code k and zero otherwise, then the total number P_k of patents listing technology class k in a particular year is $P_k = \sum_i I_{ik}$, where index i runs over all patent applications in a particular year. Similarly, the co-appearance of two IPC classes k and l is defined as $P_{k,l} = \sum_i I_{ik} \times I_{il}$. The proximity between any two technology classes k and l is then given by

$$\text{proximity}_{k,l} = \frac{P_{k,l}}{\sqrt{P_k^2 \times P_l^2}},$$

which is a standardised co-occurence measure of technology classes. It reflects the likelihood of recombination of specific knowledge domains in a technology field, and also represents the edges in the network of knowledge domains, or the knowledge space.

The vertices in this network, in turn, consist of PageRank weighted fractional counts of patents falling in a specific knowledge domain. This is shown in Figure (4) for the years 1980, 1995 and 2010. The vertices are the bigger, the higher the PageRank score of the patents allocated to it. The relative position of the vertices has been kept constant over time. The colour of the vertices corresponds to specific aggregate technology classes shown in the legend of the figures. For better readability of the figures, we show the network at the level of aggregated (four digit) IPC classes.

For each patent i, we calculate a centrality measure, centrality_i, which is defined as the maximum degree centrality of the corresponding IPC classes:

$$\text{centrality}_i = \max_{k \in C_{IPC}: I_{ki}=1} \sum_{l \in C_{IPC}} \text{proximity}_{k,l}.$$

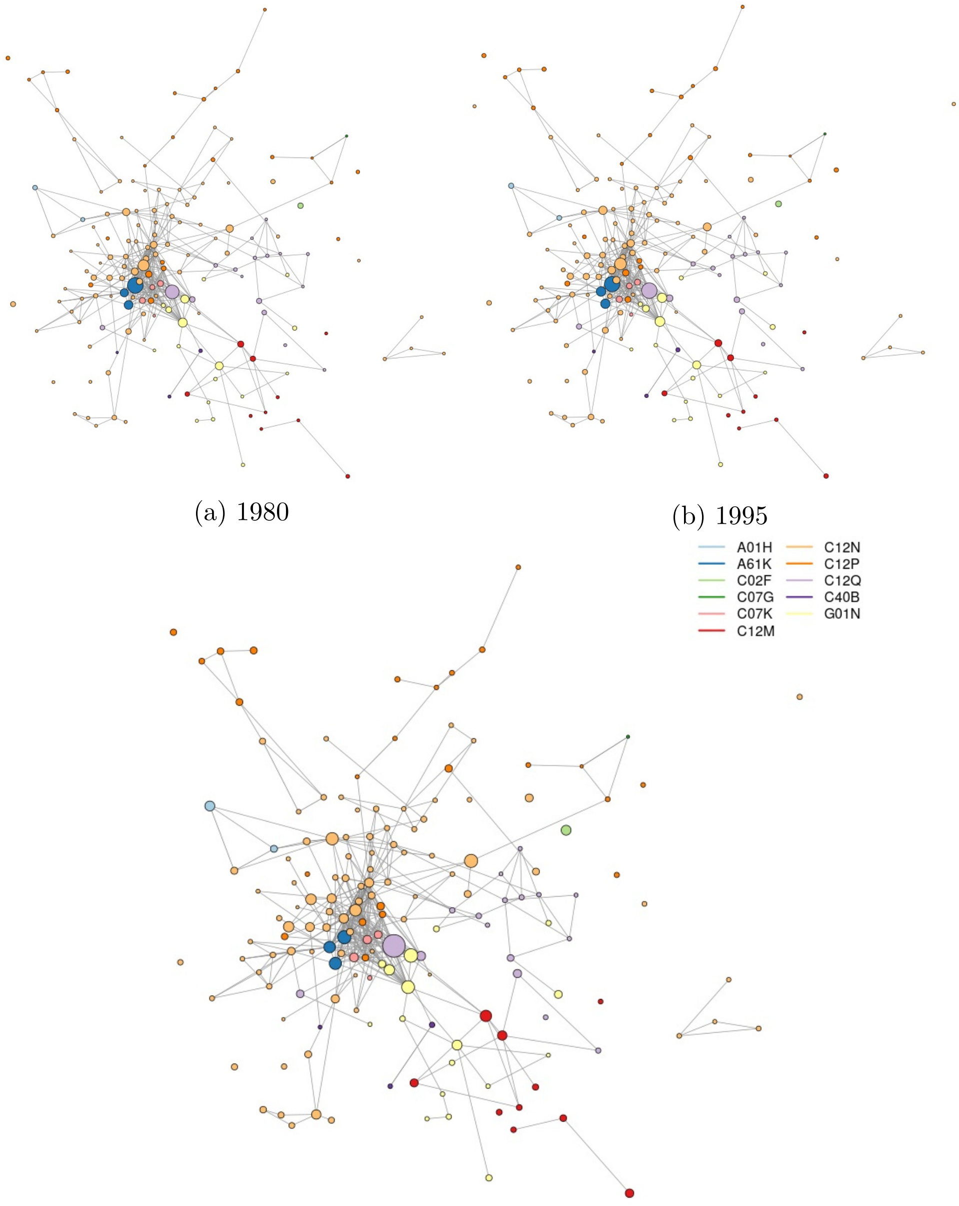
We use this indicator later in the regression analysis, to examine how the position of patent in the knowledge space affects its PageRank and Alpha centrality scores as well as the number of inward citations it receives.[9]

A first look at Figure (4) shows that there is a dense part in the knowledge space for biotechnology patents. At the centre of this dense part in 1980, we observe (Figure 4a) two big vertices at the centre and a two slightly smaller ones in the periphery of the network. The two central vertices are related to the high level IPC classes "preparations for medical purposes" (A61K) and "micro-organisms and enzymes" (C12N), whereas the two peripherical ones are related to the high level IPC classes "Measuring and testing processes involving enzymes" (C12Q) and again the class "micro-organisms and enzymes" (C12N). As we move through time, these central nodes loose importance and the node related to "Measuring and testing processes involving enzymes" (C12Q) positioned slightly at the periphery of the core gets increasingly important, whereas the other knowledge domains lose importance in terms of patenting activity and PageRank scores. This may be taken as a first evidence, for the process of technical progress outlined above.

We examine this by means of a simple regression analysis using OLS.[10] We regress the knowledge space centrality measure of each patent on its PageRank and Alpha centrality scores as well as its inward citations, controlling for important patent characteristics that may influence the number

[9]Erdi et al. (2013) propose an alternative similarity indicator based on a distance measure obtained from an inward citation vector for each patent across 36 technological subcategories. This measure is less granular and focuses on "long jumps" in the knowledge space, whereas the indicator used here is able to capture also local recombinant developments. For the level of aggregation chosen for this study, this method seem not so well suited as the number of clusters resulting from the analysis of the citation vectors is extremely large, and it not possible to identify an objective criterion to determine the correct number of cluster.

[10]For an analysis of inward citations count data models would be more adequate than OLS. PageRank scores do not show properties of count data for this reason OLS was used. For the EPO sample PageRank scores ($\alpha = 0.5$; PR $\times 10^6$) range between a minimum of 2158 and a maximum of 192478 with an average of 2600.



(a) 1980

(b) 1995

(c) 2010

Figure 4: The evolution of PageRank scores in the biotechnology knowledge space. Network constructed from global sample of patents in biotechnology. PageRank scores of vertices are based on the EPO sample only. Authors' calculations using EPO PATSTAT data base. Bubble sizes reflects the sum of PageRank scores in each subperiod. Linkages for proximity > 0.1. Graph generated using the Spring Force algorithm. IPC class description listed in the appendix.

of citations a patent receives. These are the number of claims in a patent (claims), an indicator showing whether the patent has been granted and the (average) number of subsequent renewals (renewal), whether the patent is part of a triadic patent family or not (triadic), and finally, whether the patent has been opposed or not (opposition). The renewal indicator is 0, if a patent has not been granted and one if it has been granted. It is larger than one and corresponds to the (average) number of renewals plus one if the the patent after having been granted has also been renewed.[11]

Important contributions such as Lanjouw and Schankerman (2001) or Bessen (2008), have argued that these indicators are also proxies for patent quality. Hence, the regression analysis on the one hand permits to examine to what extent centrality in knowledge space correlates with PageRank scores. At the same time, by controlling for important determinants influencing the number of citations a patent receives, it allows also to examine to what extent the scores obtained from PageRank, Alpha centrality and inward citation counts correlate with indicators for patent quality. The regression analysis is carried out for EPO patent applications only, to ensure data consistency as to what concerns the legal and procedural aspects of patent applications. We then split this sample into smaller subsamples capturing the tail parts of the distribution of the scores in order to examine whether the relationship changes in the tails of the distribution.

Tables 4 and 5 show the results. The two tables differ insofar as Table 5 includes an interaction effect between *centrality* and *age*, to examine how the impact of knowledge space centrality changes with patent age. All

[11]As an EPO patent has to be validated by national partner patent offices and patent owners can then decide on a country-by-country basis whether to renew or not the patent the database contains several renewal dates. This is true for all other patents extended to foreign patent offices through the Paris convention well. We take therefore the average number of renewals.

Table 4: The relationship between patent position in the knowledge space and PageRank: Regression analysis for EPO patent applications.

Dependent variable:									
	ALL			TOP 10%			TOP 1%		
	Inward Cit.	PageRank $\alpha = 0.50$	Alpha C. $\alpha = 0.50$	Inward Cit.	PageRank $\alpha = 0.50$	Alpha C. $\alpha = 0.50$	Inward Cit.	PageRank $\alpha = 0.50$	Alpha C. $\alpha = 0.50$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Centrality (knowledge space)	0.012***	-0.019***	0.012***	0.362***	0.266***	0.162***	1.070***	0.756***	0.676***
	(0.003)	(0.002)	(0.003)	(0.031)	(0.022)	(0.025)	(0.199)	(0.135)	(0.225)
Age	0.154***	0.258***	0.042***	0.230***	0.615***	0.267***	0.202	0.741***	2.227***
	(0.003)	(0.003)	(0.003)	(0.037)	(0.026)	(0.03)	(0.3)	(0.229)	(0.477)
Opposition	0.081***	0.092***	0.050***	0.239***	0.290***	0.166***	0.358***	0.530***	0.572***
	(0.003)	(0.002)	(0.003)	(0.019)	(0.014)	(0.016)	(0.084)	(0.062)	(0.109)
Claims	0.037***	0.048***	0.002	0.136***	0.243***	0.045*	-0.001	0.167	-0.098
	(0.003)	(0.003)	(0.003)	(0.032)	(0.023)	(0.026)	(0.178)	(0.121)	(0.21)
Renewal Mean	0.038***	0.030***	0.003	0.204***	0.181***	0.023	0.177	0.116	0.049
	(0.003)	(0.003)	(0.003)	(0.029)	(0.02)	(0.023)	(0.185)	(0.126)	(0.213)
Triadic	-0.009***	-0.009***	-0.003	-0.02	0.02	0.017	-0.195	0.01	0.323
	(-0.003)	(-0.003)	(-0.003)	(0.037)	(0.025)	(0.029)	(0.272)	(0.182)	(0.288)
Constant	0	0	0	0.897***	0.658***	-0.225***	4.161***	4.557***	-2.857***
	(0.002)	(0.002)	(0.002)	(0.061)	(0.042)	(0.049)	(0.536)	(0.462)	(0.977)
Observations	160,172	160,172	160,172	11,079	14,196	15,696	1,518	1,602	1,602
R^2	0.046	0.094	0.005	0.048	0.105	0.017	0.038	0.088	0.043
Adjusted R^2	0.046	0.094	0.005	0.047	0.104	0.017	0.034	0.084	0.039
Residual Std. Error	0.977	0.952	0.997	3.347	2.631	3.159	7.943	5.61	9.543
	(df = 160165)	(df = 160165)	(df = 160165)	(df = 11072)	(df = 14189)	(df = 15689)	(df = 1511)	(df = 1595)	(df = 1595)
F Statistic	1,274.860***	2,777.731***	142.187***	92.034***	276.422***	45.943***	9.929***	25.546***	11.868***
	(df = 6; 160165)	(df = 6; 160165)	(df = 6; 160165)	(df = 6; 11072)	(df = 6; 14189)	(df = 6; 15689)	(df = 6; 1511)	(df = 6; 1595)	(df = 6; 1595)
Note:									β- coefficients: *p<0.1; **p<0.05; ***p<0.01

Table 5: The relationship between patent position in the knowledge space and PageRank: Regression analysis for EPO patent applications with interaction effect $Centrality \times Age$

Dependent variable:		ALL			TOP 10%			TOP 1%	
	Inward Cit.	PageRank $\alpha = 0.50$	Alpha C. $\alpha = 0.50$	Inward Cit.	PageRank $\alpha = 0.50$	Alpha C. $\alpha = 0.50$	Inward Cit.	PageRank $\alpha = 0.50$	Alpha C. $\alpha = 0.50$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Centrality (knowledge space)	-0.008***	-0.040***	-0.001	0.063	-0.104***	-0.184***	0.188	-0.204	-2.452***
	(0.003)	(0.003)	(0.003)	(0.053)	(0.037)	(0.043)	(0.451)	(0.386)	(0.857)
Centrality × Age	0.090***	0.095***	0.059***	0.240***	0.294***	0.277***	0.592**	0.536***	1.633***
	(0.003)	(0.003)	(0.003)	(0.035)	(0.024)	(0.028)	(0.272)	(0.202)	(0.432)
Age	0.177***	0.282***	0.057***	0.237***	0.646***	0.281***	-0.043	0.569**	1.315**
	(0.003)	(0.003)	(0.003)	(0.037)	(0.026)	(0.03)	(0.32)	(0.237)	(0.533)
Opposition	0.078***	0.088***	0.048***	0.233***	0.283***	0.159***	0.352***	0.526***	0.563***
	(0.003)	(0.002)	(0.003)	(0.019)	(0.014)	(0.016)	(0.084)	(0.062)	(0.108)
Claims	0.039***	0.050***	0.003	0.132***	0.236***	0.04	0.002	0.17	-0.097
	(0.003)	(0.003)	(0.003)	(0.031)	(0.023)	(0.026)	(0.177)	(0.12)	(0.209)
Renewal Mean	0.035***	0.027***	0.001	0.198***	0.175***	0.017	0.181	0.11	0.078
	(0.003)	(0.003)	(0.003)	(0.029)	(0.02)	(0.023)	(0.185)	(0.126)	(0.212)
Triadic	-0.007***	-0.006**	-0.001	0.003	0.044*	0.04	-0.159	0.038	0.4
	(0.003)	(0.002)	(0.003)	(0.037)	(0.025)	(0.029)	(0.272)	(0.182)	(0.288)
Constant	-0.017***	-0.018***	-0.011***	0.903***	0.644***	-0.217***	4.522***	4.906***	-1.05
	(0.002)	(0.002)	(0.003)	(0.061)	(0.042)	(0.049)	(0.56)	(0.479)	(1.084)
Observations	160,172	160,172	160,172	11,079	14,196	15,696	1,518	1,602	1,602
R^2	0.052	0.102	0.008	0.052	0.114	0.023	0.041	0.092	0.051
Adjusted R^2	0.052	0.102	0.008	0.051	0.113	0.023	0.037	0.088	0.047
Residual Std. Error	0.974	0.948	0.996	3.34	2.617	3.149	7.933	5.6	9.503
	(df = 160164)	(df = 160164)	(df = 160164)	(df = 11071)	(df = 14188)	(df = 15688)	(df = 1510)	(df = 1594)	(df = 1594)
F Statistic	1,259.471***	2,586.032***	187.877***	85.949***	260.442***	53.543***	9.210***	22.986***	12.301***
	(df = 7; 160164)	(df = 7; 160164)	(df = 7; 160164)	(df = 7; 11071)	(df = 7; 14188)	(df = 7; 15688)	(df = 7; 1510)	(df = 7; 1594)	(df = 7; 1594)
Note:									β coefficients: *p<0.1; **p<0.05; ***p<0.01

variables in both tables have been standardised. The coefficients therefore are β-coefficients that can be directly compared to one another. The first important observation from Table 4 is, that the centrality of a patent in the knowledge space is strongly and positively correlated with PageRank, Alpha centrality and inward citation scores of patents. The coefficient increases considerably, the more we move into the tail of the PageRank and inward citation distributions. This clearly indicates that patents that are in the more central parts of the knowledge space get also more direct and indirect citations. Hence, this suggests that inventive activity and technical progress in biotechnology are cumulative and cluster in knowledge space.

The results in Table 5 allow to further qualify this finding. As the interacted variables and the interaction effect are all included in the regression the interpretation of the coefficients for *centrality* and *age* change relative to Table 4. Now these coefficients represent the effect of one variable, conditional on the other equalling zero, whereas the total effect of a variable results from the linear combination of the coefficient of that variable and the coefficient for the interaction effect. Looking at the coefficient for *centrality*, one sees now that if the patent age is zero or very recent, then being positioned in a very central part of the network negatively affects the different scores, whereas the total effect, i.e. the linear combination of the coefficient for *centrality* and the interaction effect *centrality* $\times$ *age* is positive. In line with the evidence presented in Figure (4), this indicates that the effect of centrality on high scores decreases over time in the technology field. This means that over time biotechnological research has branched out of traditional knowledge fields and developed new trajectories positioned on more peripheral parts of the biotechnology knowledge space. It also implies that important later patents cumulatively build on earlier ones, but that as time goes by the influence of the early paradigmatic patents positioned in the very core of the knowledge space decreases for later patents. The different score measures therefore seem to capture this aspect of technological development well.

A second important observation is that controls for patent properties that reflect the quality of a patent, are positively correlated with the scores, with the exception of the indicator for triadic patent families. While for inward citations this is an expected result, it confirms that also the PageRank and Alpha centrality scores can be taken as a good proxy for the quality (and by implication the commercial value) of a patent. However, looking at the F and the R^2 model statistics it is apparent that inward citations and Alpha centrality scores have a higher variation and do not correlate as strongly with indicators for patent quality as PageRank scores. The model with PageRank scores as dependent variable explains about twice as much of the variation than the other models. The highly significant effect of *age* in all regressions clearly points at an age bias for all three scores. However, controlling for age, PageRank scores are still positively and significantly related to controls for patent quality. Overall, this evidence indicates that PageRank scores should be a preferred measure when it comes to characterise both the importance of a patent for technical progress in a field as well as its economic value.

A final observation that can be made on the results in Table 4 is that in the tails of the PageRank and inward citation distributions the models qualitatively differ from the total sample and the subsample limited to the top decile. The indicators on claims, renewal/grant and triadic filings get insignificant indicating that in the tail these indicators are no longer able to explain differences in citation based scores across patents.This is the case, as almost all patents in the tail are very similar in these characteristics. Only the control variable for oppositions remains significant and positive. This is in line with earlier findings that have found this to be a particularly important indicator for patent quality (cf. Bessen, 2008).

To conclude, the regression analysis lends support to the view that PageRank scores capture important aspects of technological development well. They strongly correlated with centrality measures in the knowledge space, and its importance becomes weaker over time, which hints at a gradual process of diversification of the technological field over time. Hence, PageRank scores capture the clustering of inventive activity in the knowledge space of biotechnology over time well, which is an important characteristic of technical progress in that technological field. The results also indicate that PageRank scores seem to be a better proxy for patent quality than inward citations or Alpha centrality.

4.3 Country rankings of inventive activities in biotechnology based on PageRank scores in a moving time window

As the analysis in the prior sections shows that PageRank scores capture important aspects of technological progress in a technology field in this final section, we examine the use of PageRank scores in the construction of country rankings that capture the participation of countries in the technological development of this field over time. Instead of countries, the unit of observation could be companies or inventors. The logic of the construction of the indicator would not change.

The indicators have been calculated as follows:

$$C_i^d = \sum_j s_i^j \mathbf{x}_j^{\mathrm{SW}(d)} p_j^{\mathrm{SW}(d)} = \sum_{j \text{ newer than } d} s_i^j \mathbf{x}_j^{\mathrm{PR}}, \tag{5}$$

where s_i^j denotes the share of country i in patent j (by inventor or applicant) to get a fractional count. Equation (5) corresponds to a standard citation weighted patent count as introduced by Trajtenberg (1990), with the exception that for each 'date' d the sum runs over newer patent applications than that date only, whereas the score vector $\mathbf{x}_j^{\mathrm{SW}(d)}$ is calculated once for all patents j and all dates.

Figure (5) shows the rankings for the top 20 nations based on the country of residence of the inventors listed in the patents.[12] They are based on counts of patent applications at the EPO weighted either through PageRank scores or inward citations in moving time windows. They have been calculated on the basis of patents in the top decile of each distribution only. Panel (a) and (c) show the rankings obtained using the PageRank, whereas panel (b) and (d) show for comparison the rankings obtained using inward citations. Panel (c) and (d) in addition show population adjusted counts to control also for country size.

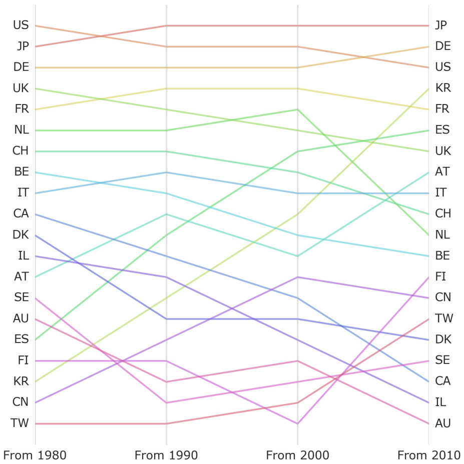
To understand how such a ranking should be interpreted, consider that if the moving time window comprises only patents with priority identical to the most recent 'date', $d = t_0$, where patents do not yet receive citations from later patents, this indicator collapses to a simple fractional patent count. In the case of the PageRank, it will be adjusted by the outward citations $1/k_j^{\mathrm{out}}$ listed in each patent j. As d stretches now back in time, the count for each country comprises only the patents entering the global patent ranking constructed from equation (2) between d and the the most recent 'date', t_0. The country ranking at 'date' d therefore shows the most important contributors to the technological field in time window $t_0 - d$. The

[12]The country codes used in Figures (5) and (6) follow the International Standard for country codes ISO 3166. Codes can be accessed at `https://www.iso.org/iso-3166-country-codes.html`.

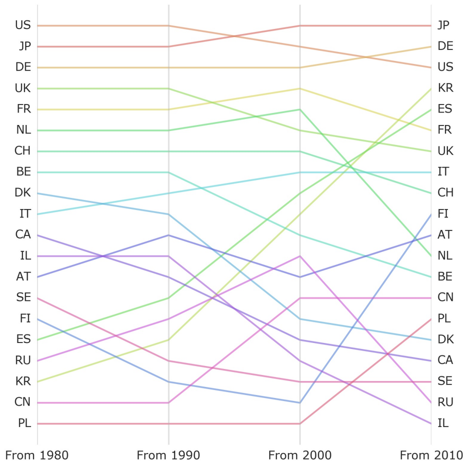
longer the time window one observes, the more the ranking corresponds to a historical ranking of technological leadership of a country in the technological field. The closer the time windows moves to more recent periods, the more the ranking captures important recent contributors to the technological field, and therefore reflects the timeliness of the inventive activities in the technological field in a country.

Figures (5a) and (5b) show that for the three nations on top of the list, the US, Japan (JP) and Germany (DE), the rankings based on PageRank and inward citations weights do not differ. The US starting as the leading nation in patenting activities in bio tech in the 1980s has lost ground to Japan and Germany in more recent years. After the three top positions we observe that the rankings and their dynamic changes over time differ between the two indicators. PageRank seems to identify relevant changes earlier than inward citation based counts. For instance, the rise of Spain (ES) and China (CN) as countries with important inventive activities in bio tech are visible earlier in the PageRank based ranking than in the inward citation based one. The rise of South Korea (KR) instead shows a more dynamic development until the year 2000 and a more dampened one afterwards in the PageRank based ranking relative to the one based on inward citations.

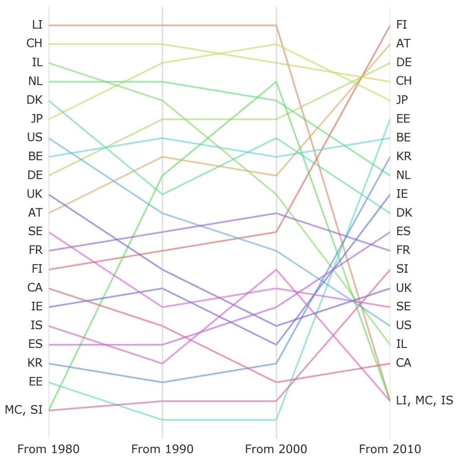
Looking at countries losing ground a similar observation can be made. For instance, the falling behind of the UK, Canada (CA) or Israel (IL) is stronger in the PageRank based ranking. The citation weights of indirect counts seem to accentuate dynamics that can be observed in simple citation based counts. Figures (5c) and (5d) show similar country specific patterns, but on the one hand due to the country size adjustment the ranking now favours smaller countries with a strong patenting record relative to larger ones, and on the other hand it also further accentuates up- and downward



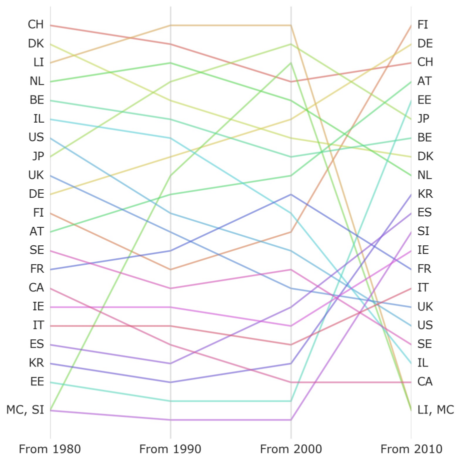
(a) PageRank



(b) Inward Citations

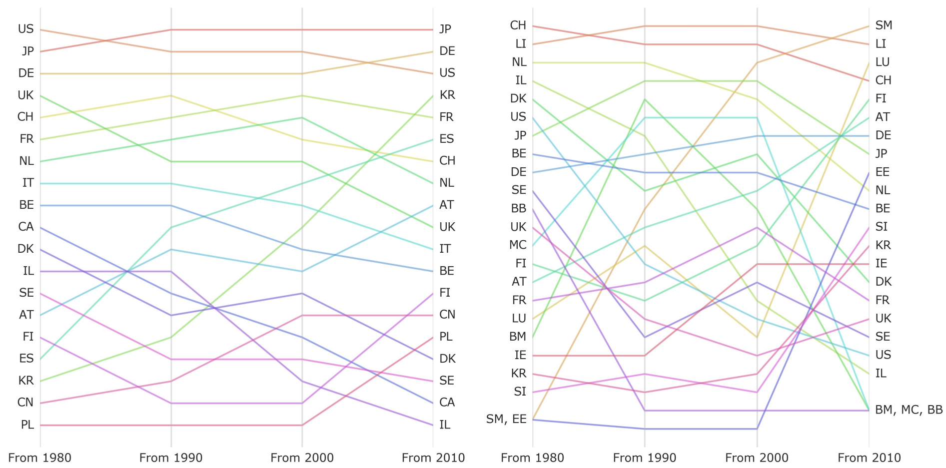


(c) PageRank,
country size adjusted



(d) Inward Citations,
country size adjusted

Figure 5: Country ranking by inventor residence over time based on the top decile of patents ranked according to PageRank and inward citations in a sliding time window. Authors' calculations based on EPO PATSTAT data and EPO patent applications.



(a) Unadjusted (b) Country size adjusted

Figure 6: Country ranking by applicant location over time based on inward citations in top PageRank decile. Authors' calculations based on EPO PATSTAT data and EPO patent applications.

developments of countries in the ranking.

Figure (6) shows now fractional counts based on the geographical location of applicants, and not inventors, and thus do not represent the geographical location of inventive activity but rather the geographical location of patent ownership. While panel (6a) largely mirrors the evidence presented in Figure (5), panel (6b) now shows how patent ownership has increased in countries reputed for being tax heavens in which larger international business groups establish offshore IP subsidiaries with the purpose to optimise their corporate tax burden. Barbados (BB), the Bermudas (BM), Monaco (MC), and Luxembourg (LU) rank now very highly. The same holds true for Lichtenstein (LI). While Monaco, the Barbados and the Bermudas were important locations in the 1990s, their importance has fallen in the more recent past. The reverse development can be observed for San Marino (SM). Luxembourg instead has constantly gained importance over time and Lichtenstein was on top throughout the period of observation. Given that tax

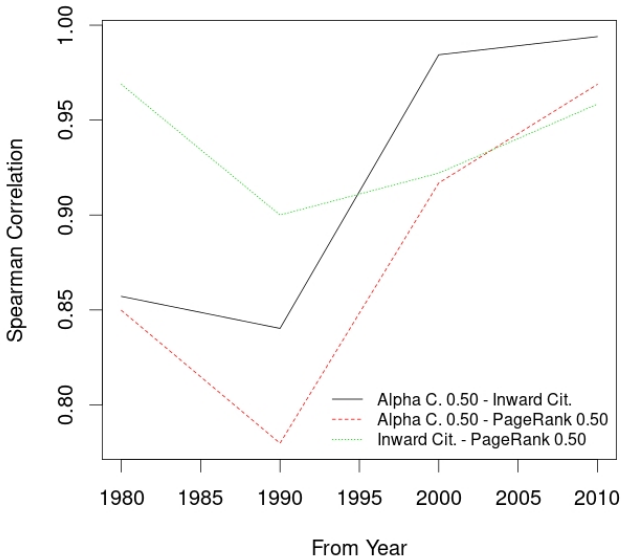


Figure 7: Spearman rank correlation between country rankings constructed from different network centrality measures based on inventor residence counts for the top decile of patents in top 20 countries. Authors' calculations based on EPO PATSTAT data and EPO patent applications.

optimisation generates the largest benefits for companies for the most valuable patents, one may take this evidence as an additional indication that PageRank scores seem to be a good proxy for the quality and the economic value of patents. As with the indicators explored in the previous section, more research using patent value data is however needed.

As in the previous section, Figure (7) presents a sensitivity analysis based on a Spearman rank correlation between cross country patent counts based on inward citation (Inward.Cit), PageRank and Alpha centrality (Alpha C. 0.50) weights. It shows the development of the rank correlation over time for the same sample evaluated in Figures (5) and (6). Despite a kink in the correlation around the 1990s that is due to a sudden surge in patent citations during that period, all indicators correlate highly. On the current edge the Alpha centrality and the inward citation based rankings almost perfectly correlate (Alpha C. 0.50 - Inward Cit), as would be expected. On the current edge indirect citations do not yet affect the ranking.

The rank correlation between inward citation and Alpha centrality based rankings on the one side and the PageRank based ranking on the other side is lower due to the outward citation weights in the PageRank. As the time window of observation stretches in the past, the rank correlations between the different approaches becomes increasingly weaker showing the influence of of both indirect and outward citations on the rank order of countries. Outward citations contribute to make Alpha and PageRank based rankings drifting apart. Indirect citations on the other hand drive apart inward citation based rankings and the rankings obtained through Alpha centrality and PageRank weights.

Comparing Figure (7) with Figure (3) the notable aspect is that while the patent rank correlation between Alpha centrality and inward citation based scores on the one hand and PageRank based is generally weak and decreases as the observed time window approaches the current edge, in Figure (7) the rank correlation increases. This indicates that patents with specific scores independently on how they are obtained cluster in specific countries and that the country rankings capture this well, leading to relatively consistent rank orders across methods. Given the results in Section 4.2, the PageRank based country rankings can be considered to capture the importance of the contribution of specific countries to the technological field historically and on the current edge best.

5 Conclusions

This paper has examined whether PageRank algorithms are a valid instrument for the analysis of technical progress by means of patent citation data. The appeal of PageRank rests in the circumstance that the importance of a patent in the development of a technological field is calculated recursively on the basis of direct and indirect citations of the cited and all citing patents. However, a number of recent papers have criticised the use of PageRank for the analysis of scientific citation networks. They argue that PageRank is biased towards historically important publications and that for this reason it is not able to properly identify recent important contributions and emerging fields. We show that this critique is partly unwarranted, if the objective of the ranking is to characterise progress in a technological field more broadly, and not just emergent fields in a specific technological domain.

The analysis focused on technological progress in modern biotechnology. The results show that with minor modifications PageRank can be used to construct valid time consistent rankings on the importance of inventions in a specific technological field and characterise the contribution of a country (or any other unit of observation) to its development over time. PageRank scores for patents capture the clustering process of inventive activities in time, where many minor and intermediate inventions are the consequences of major paradigmatic ones, and the cumulativeness of knowledge generation in a technological field well. They closely correlate with centrality measures of the knowledge space that captures the co-appearance of specific technological classes across patents in a technological field. Indeed, patents that are close to the core of the knowledge space are more likely to score high in PageRank. This points to a clustering of inventive activities also in specific knowledge domains with the development of new trajectories in related sub domains over time.

The PageRank scores of patents strongly and positively correlate with important measures of patent quality and the commercial value of patents identified in the pertinent literature. PageRank scores are more consistent and show less variation than inward citation or Alpha centrality based scores, and are therefore better explained by patent quality measures. For this reason, PageRank scores for patents can be considered the preferred measure to proxy the technological and economic relevance of an invention.

Finally, we have shown that time consistent rankings of countries can be obtained that capture the contribution of countries to the development of a technological field over time. To this end, PageRank has to be used in a moving time window framework as otherwise the critique that it is biased towards historically important inventions is valid. In a moving time window setting, instead the longer the time window one observes, the more the ranking corresponds to a historical ranking of technological leadership of a country in the technological field. The closer the time windows moves to more recent periods, the more the ranking captures important recent contributors to the technological field. PageRank based country rankings seem to accentuate cross country dynamics in inventive activities in a technological field over time more than rankings based on inward citation based counts.

Our sensitivity analyses show, that PageRank based country rankings of patenting in a technological field are largely consistent with rankings constructed from simple inward citations. From this the question arises, whether we gain much by using computationally considerably more burdensome PageRank based patent scores over simple inward citation based ones. We would argue that given that PageRank based scores appear to capture key characteristics of technical progress better, it is worth making this additional effort.

Acknowledgements: The research leading to this paper has received funding in the context of the Austria 2025 Research Project funded by the Federal Ministry of Transport, Innovation and Technology (BMVIT), the Federal Ministry Economic Affairs and Research (BMWFW), and the Austrian National Bank (OeNB).

References

Abrams, D., Akcigit, U., and Popadak, J. (2013). Understanding the link between patent value and citations: Creative destruction or defensive disruption. mimeo., University of Pennsylvania.

Acemoglu, D., Akcigit, U., and Kerr, W. (2016). Innovation network. Working Paper 22783, National Bureau of Economic Research.

Alcacer, J., Gittelman, M., and Sampat, B. (2009). Applicant and examiner citations in u.s. patents: An overview and analysis. *Research Policy*, 38:415–427.

Atallah, G. and Rodriguez, G. (2006). Indirect patent citations. *Scientometrics*, 67:437–65.

Avrachenko, L., Litvak, N., and Pham, K. (2008). A singular pertubation approach for choosing the pagerank damping factor. *Internet Mathematics*, 5:47–70.

Bessen, J. (2008). The value of u.s. patents by owner and patent characteristics. *Research Policy*, 37:932–945.

Boldi, P., Santini, S., and Vigna, S. (2005). Pagerank as a function of the damping factor. Proceedings of the 14tj international conference on the world wide web, Association for Computing Machinery, New York.

Bonacich, P. and Lloyd, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23:191–201.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30:107–117.

Bruck, P., Rethy, I., Szente, J., Tobochnik, J., and Erdi, P. (2016). Recognition of emerging technology trends: class-selective study of citations in the u.s. patent citation network. *Scientometrics*, 107:1465–1475.

Castaldi, C., Frenken, K., and Los, B. (2014). Related variety, unrelated variety and technological breakthroughs: An analysis of US state-level patenting. *Regional Studies*.

Chen, P., Xie, H., S., M., and S., R. (2007). Finding scientific gems with google. *Journal of Informetrics*, 8:8–15.

de Rassenfosse, G., Dernis, H., and G., B. (2014). Patent citation data in social science research: Overview and best practices. Working Paper 8/14, Melbourne Institute Working Paper.

Ding, Y., Yan, E., Frazho, A., and Caverlee, J. (2009). Pagerank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, 60:2229–2243.

Erdi, P., Makovi, K., Somogyvari, Z., Strandburg, K., Tobochnik, J., Volf, P., and Zalanyi, L. (2013). Predicition of emerging technologies based on analysis of the us patent citation network. *Scientometrics*, 95:225–242.

Hagedoorn, J. and Cloodt, M. (2003). Measuring innovative performance: is there an advantage in using multiple indicators? *Research Policy*, 32:1365–1379.

Hall, B., Griliches, Z., and Pakes, A. (1991). R&d, patents, and market value revisited: Is there a second (technological opportunity) factor? *Economics of Innovation and New Technology*, 1(1):183–202.

Hall, B., Jaffe, A., and Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*, 36:16–38.

Harhoff, D., Narin, F., and Vopel, K. (1999). Citation frequency and the value of patented inventions. *Review of Economics and Statistics*, 812:511–515.

Jaffe, A. and de Rassenfosse, G. (2016). Patent citation data in social science research: Overview and best practices. Working Paper 21868, National Bureau of Economic Research.

Katila, R. (2000). Using patent data to measure innovation performance. *International Journal of Business Performance Management*, 2.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, 18:39–43.

Kogler, D., D.L., R., and Tucker, I. (2013). Mapping knowledge space and technological relatedness in us cities. *European Planning Studies*, 21:1374–1391.

Langville, A. and Meyer, C. (2003). Survey: Deeper inside pagerank. *Internet Mathematics*, 1:335–380.

Lanjouw, J. and Schankerman, M. (2001). Characteristics of patent litigation: A window on competition. *RAND Journal of Economics*, 32:129–51.

Mariani, M. S., Medo, M., and Zhang, Y.-C. (2015). Ranking nodes in growing networks: When PageRank fails. *Scientific Reports*, 5:16181.

Maslov, S. and Redner, S. (2008). Promise and pitfalls of extending google's pagerank algorithm to citation networks. *Journal of Neuroscience*, 28(44):11103–11105.

Scherer, F. and Harhoff, D. (2000). Technology policy for a world of skew-distributed outcomes. *Research Policy*, 29:559–566.

Shaffer, M. (2011). Entrepreneurial innovation: Patent rank and marketing science. Phd dissertation, Washington State University.

Silverberg, G. (2002). The discrete charm of the bourgeoisie: quantum and continuous perspectives on innovation and growth. *Research Policy*, 31:1275–1289.

Silverberg, G. and Verspagen, B. (2007). The size distribution of innovation revisited: An application of extreme value statistics to citation and value measures of patent significance. *Journal of Econometrics*, 139:318–339.

Trajtenberg, M. (1990). A penny for your quote: Patent citations and the value of innovations. *RAND Journal of Economics*, 21:172–187.

Valverde, S., Sole, R. V., Bedau, M. A., and Packard, N. (2007). Topology and evolution of technology innovation networks. *Physical Review E*, 76(5):056118.

Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems*, 10:93–115.

Walker, D., Xie, H., Yan, K.-K., and Maslov, S. (2007). Ranking scientific publications using a model of network traffic. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(06):P06010.

A Appendix

A.1 IPC Classes

List of IPC Labels of Classes used in the network graphs in Figure 4:

IPC Code	Label
A01H	NEW PLANTS OR PROCESSES FOR OBTAINING THEM; PLANT REPRODUCTION BY TISSUE CULTURE TECHNIQUES
A61K	PREPARATIONS FOR MEDICAL, DENTAL, OR TOILET PURPOSES (devices or methods specially adapted for bringing pharmaceutical products into particular physical or administering forms A61J0003000000; chemical aspects of, or use of materials for deodorisation of air, for disinfection or sterilisation, or for bandages, dressings, absorbent pads or surgical articles A61L)
C02F	TREATMENT OF WATER, WASTE WATER, SEWAGE, OR SLUDGE (processes for making harmful chemical substances harmless, or less harmful, by effecting a chemical change in the substances A62D0003000000; separation, settling tanks or filter devices B01D; special arrangements on waterborne vessels of installations for treating water, waste water or sewage, e.g. for producing fresh water, B63J; adding materials to water to prevent corrosion C23F; treating radioactively-contaminated liquids G21F0009040000)
C07G	COMPOUNDS OF UNKNOWN CONSTITUTION (sulfonated fats, oils or waxes of undetermined constution C07C0309620000)
C07K	PEPTIDES (peptides containing -lactam rings C07D; cyclic dipeptides not having in their molecule any other peptide link than those which form their ring, e.g. piperazine-2,5-diones, C07D; ergot alkaloids of the cyclic peptide type C07D0519020000; genetic engineering processes for obtaining peptides C12N0015000000)
C12M	APPARATUS FOR ENZYMOLOGY OR MICROBIOLOGY (installations for fermenting manure A01C0003020000; preservation of living parts of humans or animals A01N0001020000; brewing apparatus C12C; fermentation apparatus for wine C12G; apparatus for preparing vinegar C12J0001100000)
C12N	MICRO-ORGANISMS OR ENZYMES; COMPOSITIONS THEREOF (biocides, pest repellants or attractants, or plant growth regulators containing micro-organisms, viruses, microbial fungi, enzymes, fermentates, or substances produced by, or extracted from, micro-organisms or animal material A01N0063000000; medicinal preparations A61K; fertilisers C05F); PROPAGATING, PRESERVING, OR MAINTAINING MICRO-ORGANISMS; MUTATION OR GENETIC ENGINEERING; CULTURE MEDIA (microbiological testing media C12Q0001000000)
C12P	FERMENTATION OR ENZYME-USING PROCESSES TO SYNTHESISE A DESIRED CHEMICAL COMPOUND OR COMPOSITION OR TO SEPARATE OPTICAL ISOMERS FROM A RACEMIC MIXTURE
C12Q	MEASURING OR TESTING PROCESSES INVOLVING ENZYMES OR MICRO-ORGANISMS (immunoassay G01N0033530000); COMPOSITIONS OR TEST PAPERS THEREFOR; PROCESSES OF PREPARING SUCH COMPOSITIONS; CONDITION-RESPONSIVE CONTROL IN MICROBIOLOGICAL OR ENZYMOLOGICAL PROCESSES
C40B	COMBINATORIAL CHEMISTRY; LIBRARIES, e.g. CHEMICAL LIBRARIES, ; IN SILICO LIBRARIES
G01N	INVESTIGATING OR ANALYSING MATERIALS BY DETERMINING THEIR CHEMICAL OR PHYSICAL PROPERTIES (measuring or testing processes other than immunoassay, involving enzymes or micro-organisms C12M, C12Q)

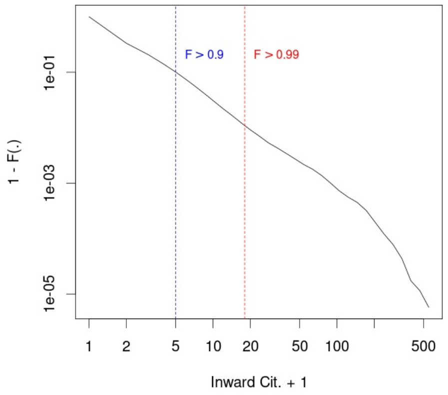
A.2 Centrality Measures

Alpha centrality is defined as follows:

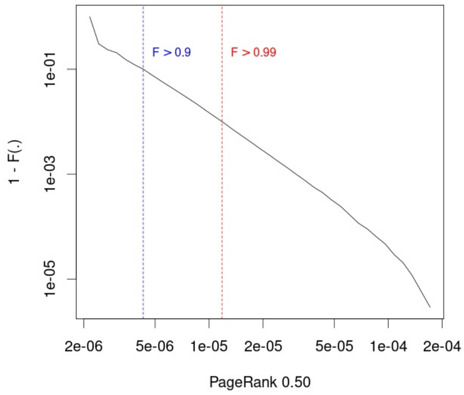
$$\mathbf{x} = \left(\alpha \mathbf{I} + \alpha A + \alpha^2 A^2 + \dots\right) \mathbf{1} = \sum_{i=0}^{\infty} (\alpha A)^i \, \mathbf{1} = (\mathbf{I} - \alpha A)^{-1} \, \mathbf{1}.$$

It corresponds to a PageRank without weight based on outward citations.

A.3 Pareto plots and the list of top 20 patents in biotechnology for EPO patent applications only - not for publication.



(a) Distribution of Inward Citations



(b) Distribution PageRank Scores

Figure A.1: Authors' calculations based on the EPO sample of patents in biotechnology extracted from the EPO PATSTAT database. Vertical lines indicate 90% and 99% of total number of patents in the sample.

Table A.1: Top 20 patents in biotechnology in the EPO sample (EPO patent applications only) from the EPO PATSTAT database

Patent Nr.	Prio. Date	Title	Inventors	R. PR50	R. PR85	R. Inward
EP19840302368	1983-04-08	Recombinant immunoglobulin preparations	Cabilly, Shmuel (US) / Holmes, William Evans (US) / Wetzel, Ronald Burnell (US) / Heyneker, Herbert Louis (US) / Riggs, Arthur Dale (US)	1	1	3
EP20000301439	1999-02-25	Sequence-determined DNA fragments and corresponding polypeptides encoded thereby	Alexandrov, Nickolai (US) / Brover, Vyacheslav (US) / Chen, Xianfeng (US) / Subramanian, Gopalakrishnan (US) / Troukhan, Maxim E. (US) / Zheng, Liansheng (US) / Dumas, J. (FR)	2	13	4
EP19870302620	1986-03-27	Recombinant antibodies and methods for their production.	Winter, Gregory Paul (UK)	3	7	1
EP19860302298	1985-03-28	PROCESS FOR AMPLIFYING, DETECTING, AND/OR CLONING NUCLEIC ACID SEQUENCES	Saiki, Randall Keichi (US) / Erlich, Henry Anthony (US) / Mullis, Kary Banks (US) / Arnheim, Norman (US) / Horn, Glenn Thomas (US) / Scharf, Stephen Joel (US)	4	3	41
EP19820301804	1981-04-17	MODIFIED NUCLEOTIDES AND METHODS OF PREPARING AND USING SAME	Ward, David C. (US) / Waldrop, Alexander A., III (US) / Langer, Pennina R. (US)	5	2	42
EP19970100117	1996-01-05	Staphylococcus aureus polynucleotides and sequences	Rosen, Craig A. (US) / Kunsch, Charles A. (US) / Choi, Gil H. (US) / Barash, Steven C. (US) / Dillon, Patrick J. (US) / Fannon, Michael R. (US)	6	59	23
EP19880310922	1987-11-18	NANBV diagnostics	Houghton, Michael (US) / Choo, Qui-Lim (US) / Kuo, George (US)	7	17	24
EP19820303197	1981-07-01	RECOMBINANT MONOCLONAL ANTIBODIES	Reading, Christopher L. (US)	8	4	394
EP19870307433	1986-08-22	Purified thermostable enzyme and process for amplifying, detecting, and/or cloning nucleic acid sequences using said enzyme	Erlich, Henry Anthony (US) / Lawyer, Frances Cook (US) / Stoffel, Susanne (US) / Mullis, Kary Banks (US) / Horn, Glenn (US) / Saiki, Randall Keichi (US) / Gelfand, David Harrow (US)	9	9	37
EP19820303699	1981-07-17	HOMOGENEOUS NUCLEIC ACID HYBRIDIZATION DIAGNOSTICS BY NON-RADIATIVE ENERGY TRANSFER	Heller, Michael James (US) / Morrison, Larry Edward (US) / Prevatt, William Dudley (US) / Akin, Cavit (US)	10	8	35
EP19870302196	1986-03-13	Process for detecting specific nucleotide variations and genetic polymorphisms present in nucleic acids and kits therefor	Saiki, Randall Keichi (US) / Erlich, Henry Anthony (US) / Horn, Glenn Thomas (US) / Mullis, Kary Banks (US)	11	10	36
EP19870302367	1986-03-28	Anti-sense regulation of gene expression in plant cells.	Shewmaker, Christine K. (US) / Kridl, Jean C. (US) / Hiatt, William R. (US) / Knauf, Vic (US)	12	11	145
EP19820300418	1981-01-28	METHOD OF MANUFACTURING MONOCLONAL ANTIBODIES AND CELLS CAPABLE OF MANUFACTURING SUCH ANTIBODIES	Williamson, Alan Rowe (UK) / Stimson, William Howard (UK) / Dick, Heather May (UK) / Clark, Stuart Airdrie (UK)	13	5	2162
EP19850302096	1984-04-16	EMBRYOGENIC CALLUS AND CELL SUSPENSION OF INBRED CORN	Lowe, Keith Sands (US)	14	38	114
EP19860300037	1985-01-07	N-(OMEGA,OMEGA-1-DIALKOXY)- AND N-(OMEGA,OMEGA-1-DIALKENOXY)-ALK-1-YL-N,N,N-TRISUBSTITUTED AMMONIUM SURFACTANTS, THEIR PREPARATION AND PHARMACEUTICAL FORMULATIONS CONTAINING THEM	Jones, Gordon Henry (US) / Eppstein, Deborah Anne (US) / Felgner, Philips Louis (US) / Roman, Richard Bolton (US)	15	12	734
EP19860302299	1985-03-28	Process for amplifying nucleic acid sequences	Mullis, Kary Banks (US)	16	19	27
EP19850305604	1984-08-27	CHIMERIC RECEPTORS BY DNA SPLICING AND EXPRESSION	Morrison, Sherie L. (US) / Herzenberg, Leonard A. (US) / Oi, Vernon T. (US)	17	31	6
EP19790105015	1978-12-11	CARBOXYALKYL DIPEPTIDE DERIVATIVES, PROCESS FOR PREPARING THEM AND PHARMACEUTICAL COMPOSITION CONTAINING THEM	Patchett, Arthur A. (US) / Harris, Elbert E. (US) / Wyvratt, Matthew J. (US) / Tristram, Edward W. (US)	18	34	123
EP19830302501	1982-05-05	Human tissue plasminogen activator, pharmaceutical compositions containing it, processes for making it, and DNA and transformed cell intermediates therefor	Goeddel, David VanNorman (US) / Kohr, William Jack (US) / Pennica, Diane (US) / Vehar, Gordon Allen (US)	19	26	102
EP20000127688	1999-12-16	Novel polynucleotides	Nakagawa, Satochi, c/o Kyowa Hakko Kogyo Co., Ltd. (JP) / Mizoguchi, Hiroshi, c/o Kyowa Hakko Kogyo Co., Ltd. (JP) / Ando, Seiko, c/o Kyowa Hakko Kogyo Co., Ltd. (JP) / Hayashi, Mikiro, c/o Kyowa Hakko Kogyo Co., Ltd. (JP) / Ochiai, Keiko, c/o Kyowa Hakko Kogyo Co., Ltd. (JP) / Yokoi, Haruhiko, c/o Kyowa Hakko Kogyo Co., Ltd. (JP) / Tateishi, Naoko, c/o Kyowa Hakko Kogyo Co., Ltd. (JP) / Senoh, Akihiro, c/o Kyowa Hakko Kogyo Co., Ltd. (JP) / Ikeda, Masato, c/o Kyowa Hakko Kogyo Co., Ltd. (JP) / Ozaki, Akio, c/o Kyowa Hakko Kogyo Co., Ltd. (JP)	20	126	15