Understanding The Diversity Of University Research Knowledge Structures And Their Development Over Time

Woltmann, Sabrina; Piccolo, Sebastiano; Kreye, Melanie

Publication date: 2018

Document Version
Publisher’s PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Understanding The Diversity Of University Research Knowledge Structures And Their Development Over Time

Sabrina Woltmann*, Sebastiano A. Piccolo**, Melanie Kreye***

*swol@dtu.dk; Management Engineering, Technical University of Denmark, Centrifugevej 372, Lyngby, 2800 (Denmark)
**sebpi@dtu.dk; Management Engineering, Technical University of Denmark, Produktionstorvet 424, Lyngby, 2800 (Denmark)
***mkreye@dtu.dk; Management Engineering, Technical University of Denmark, Produktionstorvet 424, Lyngby, 2800 (Denmark)

Introduction
Public research in universities is today under high pressure to contribute to society and economic development (D’Este & Patel 2007, Tijssen et al. 2009). Universities are seen as knowledge centres, which means they create new knowledge (Ankrah et al. 2013, Perkmann et al. 2013), provide expertise, and foster innovation (Etzkowitz & Leydesdorff 1997). Accordingly, a key function of universities is knowledge dissemination through different research output types, such as (journal) publications, patents, license agreements and so on. This dissemination is often measured through various proxy indicators. Two main approaches can be distinguished: one focusing on research output from academics for academics, such as (journal) publications (Tijssen et al. 2002, Waltman 2016), and the other investigating research output that fosters university-industry exchange, including patents, license agreements and spin-outs (Drucker & Goldstein 2007). However, current methods and empirical studies often focus only on academic or non-academic implications. This separation leads to the absence of recognition of the inter-relation between the different types of research output, resulting in an underassessment of the true impacts of research (Cohen et al. 2002).

This study explores the different types of research output by examining the overall structure of research output of one technical university in Europe over time. The goal is to identify the internal development, relevant key features and their integration into the university knowledge structure (Jensen et al. 2003, Geuna & Muscio 2009). By investigating the structure and changes over time, this study identifies the different dissemination strategies in the light of changing paradigms. Our objectives are to investigate the distribution of different output types, to identify their potential content overlap and understand the relevance of these different types. To achieve the objectives we utilize tools from social network analysis and bibliometrics.

Literature
Current studies try to unveil the underlying structures of knowledge transfer from and between universities. This led to highly interdisciplinary research (Gherardini & Nucciotti 2017), focusing either on economic and societal implications (Drucker & Goldstein 2007, Cheah 2016) or on a purely academic perspective (Tartari et al. 2014). The former focuses on
commercially relevant indicators like patents or license agreements (Erdi et al. 2013), while the latter examines academic transfer through citation networks. There has been limited attempt to investigate their relationship (Salter et al. 2017). A recent development is the introduction of 'patent-paper pairs', which uses empirically the combination of patents and their related academic publications (Magerman et al. 2015, Roach & Cohen 2013). For our purpose we draw from the two streams to get a full picture of knowledge structures within one institution. This approach highlights the overall relevance of university research output types. We expect the following outcomes:

**Hypothesis 1a:** There is an observable change in the distribution of the different output types produced by the university over time.

**Hypothesis 1b:** Non-journal output becomes more integrated into the network over time.

Furthermore, it is important to identify the overlaps in knowledge between the different types to show the importance of a combined assessment.

**Hypothesis 1c:** Patent-Paper Pairs differ, but overlap, in their references and are bridges to the different partitions in the knowledge network.

**Data & Method**

This research utilises a network analytic approach because of its suitability for the purpose of this study. Many network analytic approaches are used to grasp the structures and development of knowledge, identifying linkages and emerging topics in various scientific areas (Su & Lee 2010, Zhang et al. 2012, Zhu et al. 2015).

Our sample of research output is collected from one technical university, which has the explicit aim to foster knowledge transfer. We utilised university’s own publication database (ORBIT), where all university written output is registered. Our sample contains only entries from the years 2005-2015, since this is the time frame with most complete data. All entries in ORBIT are registered with a type label, which enables us to distinguish between the different output types like patents, papers, book chapters and a label for the scientific fields (in our case these are classified into 20 different scientific fields). The total number of entries for the time period is 77920. We start out with a common citation network created from the Scopus publication database (Boyack 2015, Kamdem et al. 2017), which we generated based on the registered entries from ORBIT. We identify the documents by using string matching for all tiles available. To follow our objectives we add the other types of research output and expand the knowledge network. However, this expansion is by no means trivial and requires quite some additional data processing.

We later add the commercially relevant indicators: patents and their citations, additional open access papers and newspaper articles using additional full-text publications and reference lists. To include these items we need to develop for each new type ways to computationally identify their citations and references. With regard to patents we examine whether these use also internal (university publications) or only external knowledge sources.

*Internal Network & External Network*
We build an *internal citation network* using only the entries from the university and the links between them. Crucial hereby is to incorporate most available output types and their citations. The identification has to be exercised by another title string matching via the Scopus application programming interface (API). This works satisfactory, in particular for longer titles.

We could identify 28,734 entries from the orbit database in Scopus. These matched entries build the nodes of the internal publication network. Further, we identified in the university database more than 1500 patent applications and retrieved their non-patent literature (NPL). This structure allows capturing the most important and interdisciplinary entries (within the university) of the internal network. On the basis of this internal network we generate also an *external citation network* based on additional Scopus references, which are not output of the university. These are used as measures of external relevance of the publications. This is to assess whether the network structure within the university reflects also the global importance of specific output.

The NPL of the patents shall be used as outward edges, but we also aim to include the patent citations, which show the importance of the inventions. We also aim to investigate the overlap between commercialized and non-commercialized output types of the university research. However, some of the citation identification approaches need improvement, in particular for patents the integration not yet been reliable.

**Preliminary results**

The preliminary results for this study are based solely on calculations that are applied to the basic internal and external Scopus networks. This provides first insights into features of relevant and high quality research items, since these are typically present in the Scopus database. Furthermore, the citations and references are verified and comparatively complete. The overall ratio between registered entries in ORBIT and Scopus is around 40%. The yearly distribution between 2005 and 2015 is not uniform (see Table 1.).

<table>
<thead>
<tr>
<th>Year</th>
<th>Total university items</th>
<th>Internal network nodes/external nodes*</th>
<th>Internal network: In edges/aver. node degree</th>
<th>Internal network: Out edges/aver. node degree</th>
<th>External network: In edges</th>
<th>External network: Out edges/aver. node degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>5907</td>
<td>1717 / 48435</td>
<td>4548 / 2.65</td>
<td>301/0.18</td>
<td>62053</td>
<td>44106 / 0.91</td>
</tr>
<tr>
<td>2006</td>
<td>6236</td>
<td>1881 / 55408</td>
<td>4836 / 2.57</td>
<td>1025/0.54</td>
<td>67433</td>
<td>50834 / 0.92</td>
</tr>
<tr>
<td>2007</td>
<td>6775</td>
<td>2179 / 68047</td>
<td>5414 / 2.48</td>
<td>1767/0.81</td>
<td>76381</td>
<td>62917 / 0.92</td>
</tr>
<tr>
<td>2008</td>
<td>6650</td>
<td>2187 / 70074</td>
<td>5319 / 2.43</td>
<td>2527/1.16</td>
<td>76431</td>
<td>65036 / 0.93</td>
</tr>
<tr>
<td>2009</td>
<td>6986</td>
<td>2465 / 79742</td>
<td>5740 / 2.33</td>
<td>3410/1.38</td>
<td>75907</td>
<td>74437 / 0.93</td>
</tr>
<tr>
<td>2010</td>
<td>6830</td>
<td>2615 / 87398</td>
<td>5729 / 2.19</td>
<td>4429/1.69</td>
<td>74913</td>
<td>82132 / 0.94</td>
</tr>
<tr>
<td>2011</td>
<td>7185</td>
<td>3008 / 102628</td>
<td>6412 / 2.13</td>
<td>6159/2.05</td>
<td>78194</td>
<td>97278 / 0.95</td>
</tr>
<tr>
<td>2012</td>
<td>7244</td>
<td>2957 / 97430</td>
<td>4588 / 1.55</td>
<td>6150/2.08</td>
<td>54832</td>
<td>93351 / 0.96</td>
</tr>
<tr>
<td>2013</td>
<td>7439</td>
<td>3144 / 110493</td>
<td>3687 / 1.17</td>
<td>7103/2.26</td>
<td>50809</td>
<td>107382 / 0.97</td>
</tr>
<tr>
<td>2014</td>
<td>7391</td>
<td>3239 / 113894</td>
<td>2275 / 0.70</td>
<td>7690/2.37</td>
<td>42749</td>
<td>112212 / 0.99</td>
</tr>
<tr>
<td>2015</td>
<td>7459</td>
<td>3342 / 126416</td>
<td>743 / 0.22</td>
<td>8730/2.61</td>
<td>30950</td>
<td>126391 / 1.00</td>
</tr>
</tbody>
</table>

*Table 1: ORBIT papers registered in Scopus per year*
External network nodes have edges with university nodes from the actual year, but no year filtering is applied on the external network nodes.

In our case the use of established basic calculations help to identify structural changes. To compare the networks we apply first simple measures like the average node degree, meaning the average number of links (edges) that a node has. We also distinguish between inwards links (in edges) and outwards links (out edges) generating a directed network. All nodes, including the university entries that were not found in Scopus, build a large sparse network with 661,859 nodes. Here over 47,000 single nodes have no (identified) connections (the average node degree is then 1.41). Due to this sparsity we remove all unattached nodes. The total number of all remaining nodes is 614,372 with 934,034 edges (1.52 average node degree). The total amount of identified nodes from the university in Scopus from 2005-2015 is 28,734 with 49,291 edges between them (1.72 average node degree).

We examine the development of the network over time by taking snapshots of the different years, calculating specific network properties and compare them. The yearly average in-degree of the internal network show a decrease in the last few years, which makes sense since it takes time before newer publications get cited by new research. The out-degree shows pretty much the opposite trend with a more steady increase in the final years, meaning that the university keeps on using their previous work (see Table 1.). The development of the external network shows similar trends.

An insight provided by the Scopus database is the actual in-edges of each paper. We did not retrieve a full external network and considered only out-degrees from the university entries, but took the overall importance of the papers into account by using their citation scores (Figure 1).
We investigated the changes within the different fields and publication types, like for instance for Open Access. Approximately 25% of university publications in Scopus are Open Access (7192 out of 28734). We looked at the citation count, differentiating for instance Open Access and non-Open Access papers as different types of publications (Figure 2).
Figure 2: In-degree for University papers present in Scopus based on access type
In the external network Open Access papers do not seem to be more cited, while in fact, it seems that the average non-Open Access publications is usually more often cited. These difference between Open Access and Non-Open Access tends to disappear with highly cited papers (network hubs). When looking at the internal network only, we see a different picture. Table 2 shows the in-degree node ratios. Here Open Access papers are more central. The average is lower for Open Access due to the low score in the last 2 years and the significant increase in the number of nodes.

Thanks to the comparatively small size of the networks, displaying only one university, a more in-depth insight into network changes is possible. We can see that the total number of open access publications increases from 2011, this is a change as stated in hypothesis 1a) as it shows a clear change in importance of certain output types.

<table>
<thead>
<tr>
<th>Year</th>
<th>Open Access Nodes</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of nodes</td>
<td>In-edges</td>
<td>Average in-degree/node</td>
<td>Number of nodes</td>
</tr>
<tr>
<td>2005</td>
<td>203</td>
<td>551</td>
<td><strong>2.71</strong></td>
<td>1514</td>
</tr>
<tr>
<td>2006</td>
<td>249</td>
<td>582</td>
<td>2.34</td>
<td>1632</td>
</tr>
<tr>
<td>2007</td>
<td>308</td>
<td>846</td>
<td><strong>2.75</strong></td>
<td>1871</td>
</tr>
<tr>
<td>2008</td>
<td>373</td>
<td>1137</td>
<td><strong>3.05</strong></td>
<td>1814</td>
</tr>
<tr>
<td>2009</td>
<td>583</td>
<td>1381</td>
<td><strong>2.37</strong></td>
<td>1882</td>
</tr>
<tr>
<td>2010</td>
<td>480</td>
<td>1177</td>
<td><strong>2.45</strong></td>
<td>2135</td>
</tr>
<tr>
<td>2011</td>
<td>751</td>
<td>2139</td>
<td><strong>2.85</strong></td>
<td>2257</td>
</tr>
<tr>
<td>2012</td>
<td>851</td>
<td>1385</td>
<td><strong>1.63</strong></td>
<td>2106</td>
</tr>
<tr>
<td>2013</td>
<td>966</td>
<td>1352</td>
<td><strong>1.40</strong></td>
<td>2178</td>
</tr>
<tr>
<td>2014</td>
<td>1051</td>
<td>784</td>
<td><strong>0.75</strong></td>
<td>2188</td>
</tr>
<tr>
<td>2015</td>
<td>1377</td>
<td>320</td>
<td><strong>0.23</strong></td>
<td>1965</td>
</tr>
<tr>
<td><strong>2005-2015</strong></td>
<td><strong>7192</strong></td>
<td><strong>11654</strong></td>
<td><strong>1.62</strong></td>
<td><strong>21542</strong></td>
</tr>
</tbody>
</table>

*Table 2: Open Access vs. Non-Open Access paper in-degrees*

**Current Challenges**
Current challenges are mainly the improvement of title detection in the different data sets. The data sample has the clear advantage that we are only searching for a limited amount of publications and do not have to rely on the detection of all references in general, which would be even more challenging. However, each of the types has own challenges, which need to be addressed. In particular the detection of citations in the full-texts remains difficult for short titles leading potentially to an under representations of the actual citations.
Discussion

Although we need more research to investigate hypotheses H1b and H1c, we found a difference in trends between open access and non-open access papers, in the internal network. Since 2011, the number of non-open access papers has not been growing, while the number of open access publications has been growing steadily, so we can already state the importance of the internal composition of different output types. The increase of average node degree over years shows an increased importance of the university research within the university itself. This is particularly evident, since the older items have an advantage to be cited also in the following years.

This shows interesting tendencies, but certainly need additional integration of the non-traditional output types into established network, which remains challenging. However, the numbers suggests that this might be highly beneficial. Conceptually, this approach aims to combine the notion of academic and industry knowledge transfer into a combined way of assessing both at the same time.

References


patent citation network’, *Scientometrics* 95(1), 225–242.


