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

Woltmann, S.L.; Piccolo, S.A.; Kreye, M.

STI 2018 Leiden

23rd International Conference on Science and Technology Indicators

"Science, Technology and Innovation Indicators in Transition"

STI 2018 Conference Proceedings

Proceedings of the 23rd International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Chair of the Conference

Paul Wouters

Scientific Editors

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

Layout

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at https://hdl.handle.net/1887/64521

ISBN: 978-90-9031204-0

23rd International Conference on Science and Technology Indicators (STI 2018)
"Science, Technology and Innovation indicators in transition"
12 - 14 September 2018 | Leiden, The Netherlands
#STI18LDN

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;*
[***] *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). Universities are knowledge centres and provide expertise, solutions or innovations and inventions (Etzkowitz & Leydesdorff 1997). Accordingly, a key function of universities is knowledge dissemination through different research output types, such as (journal) publications, patents, newspaper articles 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 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 period 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 this 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. For patents in particular, the integration has 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.).

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

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).

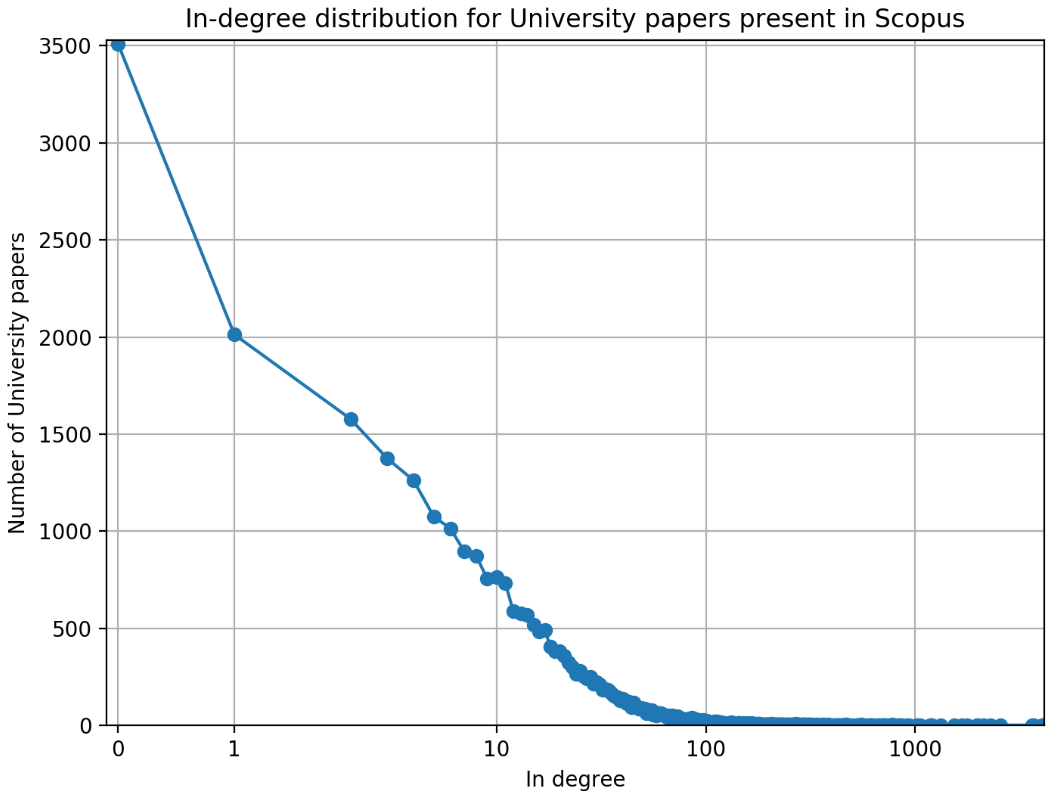

Figure 1: In-degree for University papers present in Scopus

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).

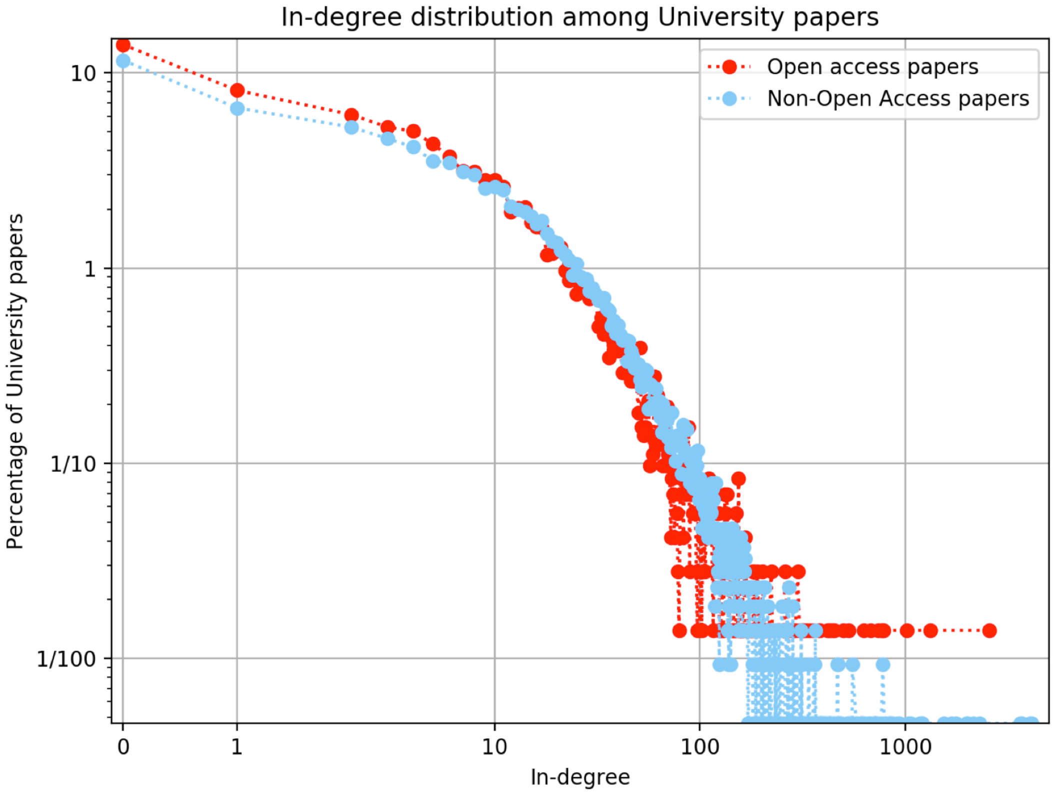

Figure 2: In-degree for University papers present in Scopus based on access type

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. This 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.

Year	**Open Access Nodes**			**Non-Open Access Nodes**		
	Number of nodes	In-edges	Average in-degree/node	Number of nodes	In-edges	Average in-degree/node
2005	203	551	**2.71**	1514	3997	2.64
2006	249	582	2.34	1632	4254	**2.61**
2007	308	846	**2.75**	1871	4568	2.44
2008	373	1137	**3.05**	1814	4182	2.31
2009	583	1381	**2.37**	1882	4359	2.32
2010	480	1177	**2.45**	2135	4552	2.13
2011	751	2139	**2.85**	2257	4273	1.89
2012	851	1385	**1.63**	2106	3203	1.52
2013	966	1352	**1.40**	2178	2335	1.07
2014	1051	784	**0.75**	2188	1491	0.68
2015	1377	320	**0.23**	1965	423	0.22
2005-2015	7192	11654	**1.62**	21542	37637	**1.75**

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

Ankrah, S. N. & Burgess, T. F., Grimshaw, P. & Shaw, N. E. (2013), 'Asking both university and industry actors about their engagement in knowledge transfer: What single-group studies of motives omit', *Technovation,* 33, 50–65.

Barabási, A. L. (2016), 'Network science', *Cambridge University Press*.

Boyack, K. W. (2015), 'Locating an astronomy and astrophysics publication set in a map of the full Scopus database', *ISSI*.

Buckland, M. & Gey, F. (1994), 'The relationship between recall and precision', *Journal of the American Society for Information Science*, 45, 12-19.

Cheah, S. (2016), 'Framework for measuring research and innovation impact', *Innovation,* 18, 212–232.

Cohen, W. M., Nelson, R. R. & Walsh, J. P. (2002), 'Links and impacts: the influence of public research on industrial R&D', *Management science* 48, 1– 23.

Crespi, G., D'Este, P., Fontana, R. & Geuna, A. (2011), 'The impact of academic patenting on university research and its transfer', *Research Policy,* 40, 55– 68.

Drucker, J. & Goldstein, H. (2007), 'Assessing the regional economic development impacts of universities: A review of current approaches', *International regional science review* 30, 20–46.

D'Este, P. & Patel, P. (2007), 'University–industry linkages in the uk: What are the factors underlying the variety of interactions with industry?', *Research Policy,* 36, 1295–1313.

É rdi, P., Makovi, K., Somogyv´ari, Z., Strandburg, K., Tobochnik, J., Volf, P. & Zal´anyi, L. (2013), 'Prediction of emerging technologies based on analysis of the us patent citation network', *Scientometrics* 95(1), 225–242.

Etzkowitz, H. & Leydesdorff, L. (1997), 'Introduction to special issue on science policy dimensions of the triple helix of university-industry-government relations'.

Geuna, A. & Muscio, A. (2009), 'The governance of university knowledge transfer: A critical review of the literature', *Minerva,* 47, 93–114.

Gherardini, A. & Nucciotti, A. (2017), 'Yesterday's giants and invisible colleges of today. a study on the 'knowledge transfer'scientific domain', *Scientometrics* 112, 255–271.

Jensen, R. A., Thursby, J. G. & Thursby, M. C. (2003), The disclosure and licensing of university inventions, *Technical report, National Bureau of Economic Research*

Kamdem, J. P., Fidelis, K. R., Nunes, R. G., Araujo, I. F., Elekofehinti, O. O., da Cunha, F. A., de Menezes, I. R., Pinheiro, A. P., Duarte, A. E. & Barros, L. M. (2017), 'Comparative research performance of top universities from the northeastern brazil on three pharmacological disciplines as seen in Scopus database', *Journal of Taibah University Medical Sciences,* 12, 483–491

Magerman, T., Van Looy, B. & Debackere, K. (2015), 'Does involvement in patenting jeopardize one's academic footprint? an analysis of patent-paper pairs in biotechnology'*, Research Policy,* 44, 1702–1713.

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Brostr¨om, A., D'Este, P., Fini, R., Geuna, A., Grimaldi, R., Hughes, A. et al. (2013), 'Academic engagement and commercialisation: A review of the literature on university– industry relations', *Research Policy*, 42, 423–442.

Roach, M. & Cohen, W. M. (2013), 'Lens or prism? patent citations as a measure of knowledge flows from public research', *Management Science,* 59, 504–525.

Schwarz, A. W., Schwarz, S. & Tijssen, R. J. (1998), 'Research and research impact of a technical university—a bibliometric study', *Scientometrics,* 41, 371–388.

Su, H.-N. & Lee, P.-C. (2010), 'Mapping knowledge structure by keyword cooccurrence: a first look at journal papers in technology foresight', *Scientometrics* 85, 65–79.

Tartari, V., Perkmann, M. & Salter, A. (2014), 'In good company: The influence of peers on industry engagement by academic scientists', *Research Policy* 43, 1189–1203.

Tijssen, R. J., Van Leeuwen, T. N. & Van Wijk, E. (2009), 'Benchmarking university-industry research cooperation worldwide: performance measurements and indicators based on co-authorship data for the world's largest universities', *Research Evaluation,* 18, 13–24.

Tijssen, R. J., Visser, M. S. & Van Leeuwen, T. N. (2002), 'Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference?', *Scientometrics* 54, 381–397.

Waltman, L. (2016), 'A review of the literature on citation impact indicators', *Journal of Informetrics* ,10, 365–391.

Zhang, J., Xie, J., Hou, W., Tu, X., Xu, J., Song, F., Wang, Z. & Lu, Z. (2012), 'Mapping the knowledge structure of research on patient adherence: knowledge domain visualization based co-word analysis and social network analysis', *PloS one,* 7

Zhu, L., Liu, X., He, S., Shi, J. & Pang, M. (2015), 'Keywords co-occurrence mapping knowledge domain research base on the theory of big data in oil and gas industry', *Scientometrics,* 105, 249–260.