Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2017 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

Summer 7-19-2017

Research complexity of Australian universities

Ivan Lee
University of South Australia, ivan.lee@unisa.edu.au

Research complexity of Australian universities

Completed Research Paper

Ivan Lee
University of South Australia
Mawson Lakes, SA, Australia

Dalian University of Technology
Dalian, Liaoning, China

ivan.lee@unisa.edu.au

Abstract

Strategic research direction and prioritisation is crucial for decision making in universities. Analysis of research diversification and sophistication helps differentiating universities according to their research attributes. Based on the Microsoft Academic Graph data set, this paper conducts research complexity analysis for all Australian universities, and examines the ubiquity and diversity of the research output. This paper also investigates research complexity indices of Australian universities, with further discussions for universities with research leadership, technological and practical focuses, and young research universities.

Keywords: Academic research, Complexity analysis, Scientometrics

Introduction

Research output assessment is an essential component in most university ranking systems (Al-Juboori et al., 2011). In a study of identifying factors of national innovative capacity, it has been found that the degree of technological specialization plays an important role (Furman et al., 2002). Strategic research direction and prioritization is therefore essential decisions made by the senior management to ensure an academic institute staying competitive in securing research and education funding allocation, the attracting industrial organization's interests in procuring applied research services, as well as the attention from prospective students.

Various scientometrics algorithms have been developed to quantify and assess an academic institutes research outputs and provide rankings (Bornmann et al., 2013; Dobrota et al., 2016; Jiang et al., 2012), using proprietary or public accessible publication records (Haley, 2014). Among these ranking systems, they frequently offer disciplinary based rankings, with the overall university ranking representing the sum of all disciplinary outputs. Identifying the research specialization of universities by analyzing the association across different disciplinary research areas is missing in these ranking systems. Exploring the specialization attributes help identifying research areas that are close in proximity in the *research space* (Guevara et al., 2016) and with most potential to grow, which can be considered an alternative approach to predict the research trends (Dwivedi et al., 2011).

Research output is not limited to research articles. Consider universities are large enterprise, funded by public or private sectors, other intellectual properties, such as patents and copyrights, are also subject to the corporate management. These intellectual properties, often associate to the university's entrepreneurship in the form of spin-off companies or knowledge transfer to the industry, is highly popular in the recent past (Etzkowitz, 2013; Payumo et al., 2014). While intellectual property is mostly associate to the research outcome, according to the study by Wong and Singh (2010), "patenting output of the leading universities in the world are indeed significantly related to their research output quantity." (Wong and Singh, 2010)

Information systems plays a crucial role in facilitating multidisciplinary research. While Pervasive digitalization has introduced a new paradigm with the shift from the age of modularity to the age of generativity, breeding joint social and technological innovations which requires "new theoretical models and insights that guide management practices in the age of generativity" (Yoo, 2013). Wu et al. (2012) discuss the role of iSchool, representing information system utilization and its applications toward interdisciplinary research and graduate education, with "the vision and mission of working on relationships between information, people and technology" to provide "diverse subject areas to study this interdisciplinary integration" (Wu et al., 2012).

Various attempts have been conducted to identifying university specialization, including the studies for academic institutes in Spain (Ortega et al., 2011), China (Li et al., 2015), and European universities in general (Pastor and Serrano, 2016). This paper examining the research complexity of universities by applying a complexity model developed for analyzing economies around the world (Hausmann et al., 2008). The original model uses yearly export records as the input, to evaluate the ubiquity and diversity of the economy's industrial sectors. Instead of export records, research complexity of highly ranked universities in the world can be assessed using the research publication records (Lee et al., 2017). Economic complexity has also been investigated at the subnational level, investigates intra-country and inter-country trade records and compares economic complexity at state's level (Reynolds et al., 2017). This paper adopts similar model and examines the research complexity across Australian universities.

Strategic direction as its research concentration differentiates itself among competition. Quantitative analysis of the research output provides a method for senior management to assess research productivity and competitiveness against competitors. The study of this paper hence represents a building block for assisted information management for making decisions on research priorities.

MAG Data Set

In order to perform analysis of research output, Microsoft Academic Service (MAG) (Sinha et al., 2015) is chosen due to its coverage across a wide range of research disciplines compares to other open data set such DBLP[1] for computer science articles and APS[2] for physics articles. The MAG dataset has

[1] http://dblp.uni-trier.de/xml/

been utilized in predictive analytics, as recommended in 2016 KDD CUP competition (Sandulescu and Chiru, 2016).

While the complexity analysis of different economies around the world utilizes the export record in different industries, classification such as Standard International Trade Classification (SITC) has been chosen in the studies by Hidalgo et al. (Hausmann et al., 2008) (Hidalgo and Hausmann, 2009). The original MAG dataset does not include classifications, and in this paper the SCImago Journal Classification, provided in sciMAG2015, is used (De Domenico et al., 2016). The 27 area code provided by sciMAG2015 is shown in Table 1.

Agricultural and Biological Sciences	Health Professions
Arts and Humanities	Immunology and Microbiology
Biochemistry, Genetics and Molecular Biology	Materials Science
Business, Management and Accounting	Mathematics
Chemical Engineering	Medicine
Chemistry	Multidisciplinary
Computer Science	Neuroscience
Decision Sciences	Nursing
Dentistry	Pharmacology, Toxicology and Pharmaceutics
Earth and Planetary Sciences	Physics and Astronomy
Economics, Econometrics and Finance	Psychology
Energy	Social Sciences
Engineering	Veterinary
Environmental Science	

Table 1 Research areas in SCImago classification (De Domenico et al., 2016)

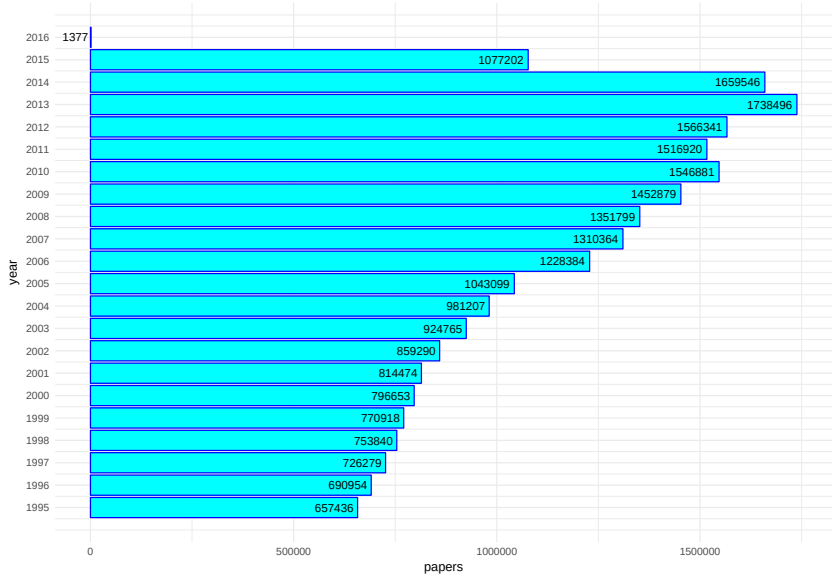

Figure 1 MAG's journal collection between 1995 and 2015

[2] http://journals.aps.org/datasets

MAG is one of the largest collection of academic publication records across all research fields. Figure 1 shows the journal publications across years included in the MAG data set. The collection increases consistently between 1995 and 2013, reflecting the expansion of academia and growth in research publications. However, there are noticeable decline in 2014 and 2015 (whereas the 2016 record is incomplete and should be discarded), and the observation coincides the reduced popularity of MAG in the academic community (Orduna-Malea et al., 2014). In this paper, the peak collection in 2013 is chosen.

Data processing

MAG applies algorithms to automatically retrieval bibliographical information such as author and institution details. However, the lack of a universal format for research articles among different publisher, as well as the lack of standard to re-enforce details provided by authors are correct, degrade the quality of the data set. For instance, some academic institutes have multiple well perceived (not necessarily official) names, such as "The University of Melbourne", "University of Melbourne", and "Melbourne University" appear as separated entries in the MAG data set. In our study, we take only the official institution name which may omit some wanted entry while filtering out some unwanted entries. Lastly, affiliation with 300 characters or more are likely parsed from biography and these entries are removed in our study. The following list shows the pre-processing applied to the data set to improve the search results.

- James Cook University: exclude James Cook University Hospital in the UK;
- Torrens University: include its former name Torrents College;
- University of New England: exclude University of New England in the USA;
- University of Newcastle: exclude University of Newcastle in the UK and UPON TYNE, USA;
- University of Notre Dame: only include the one in Australia;
- University of South Australia: exclude Flinders University of South Australia;
- Victoria University: exclude Victoria University in Toronto, Wellington, and Manchester.

Note: research publications from offshore campuses, such as James Cook University's Singapore campus, are counted towards the publications of their parent universities.

The complete record of published journal articles by Australian universities found in the MAG data set can be found in Appendix I of this paper.

Research Complexity Analysis

According to the ATLAS of Economic Complexity, Economic Complexity Index is obtained by analysing the export classified in industrial/product sectors from an economy/country. In this paper, the export value is replaced by the number of published journal articles, the product sectors are replaced by the research fields, and the countries are replaced by the universities.

It should be noted that while export values are typically associated to products developed in the recent past, research articles, on the other hand, may be subject to variable lengths of review time. The duration of the review is not available in the MAG bibliographic data set; hence, the review time is not considered in this study.

Let i denote an academic institute (university), and f denotes a research field. For a paper $\vec{P}_{if}$ published in journals labelled with m disciplines, a normalisation factor $m_{p(i,f)}$ needs to be applied. The publication matrix $\vec{P}_{if}$ is defined as:

$$\vec{P}_{if} = \sum_{p(i,f)} \frac{1}{m_{p(i,f)}}.$$

To evaluate whether a university has revealed comparative advantage (RCA) (Balassa, 1965) in a particular research field, scientific output in terms of journal publication is a popular quantitative measure. While comparing universities with different scales, fair comparison should take account of

the ratio of the research publications of a research area over the value of all research areas from a university, and the ratio of total publication in a research area over all research areas from all universities. Thus, a smaller scale university with fewer number of publications can be adequately assess its research concentration or the level of specialisation (as oppose to the competitiveness of the university in a research field.) Mathematically, RCA for research institute i and research field f is formulated as:

$$\vec{RCA}_{if} = \frac{\vec{P}_{if} \sum_{i,f} \vec{P}_{if}}{\sum_{i} \vec{P}_{if} \sum_{f} \vec{P}_{if}}.$$

To determine whether the specialisation of a university in a research area is significant or not, $\vec{M}_{if}$ contains a Boolean matrix with 0's indicating insignificance and 1's indicating significance. The threshold of 1 is inherited from the study in Economic Complexity analysis (Hausmann et al., 2008).

$$\vec{M}_{if} = \begin{cases} 0 & \text{if } \vec{RCA}_{if} < 1, \\ 1 & \text{if } \vec{RCA}_{if} \geq 1. \end{cases}$$

Diversity and ubiquity measures of an institute can be determined from analysing $\vec{M}_{if}$. Diversity $\vec{D}_i$ is the measure of an institute's research fields with reveal comparative advantage, whereas ubiquity $\vec{U}_f$ shows how many research institutes has revealed comparative advantage in a research field. Mathematically, $\vec{D}_i$ and $\vec{U}_f$ are defined as:

$$\vec{D}_i = k(i,0) = \sum_{f} \vec{M}_{if}.$$

$$\vec{U}_f = k(f,0) = \sum_{i} \vec{M}_{if}.$$

Then, the values of $k(i,n)$ and $k(f,n)$ may be iteratively obtained using the method of reflections (Hidalgo and Hausmann, 2009):

$$k(i,n) = \frac{1}{k(i,0)} \sum_{f} \vec{M}_{if} k(f,n-1).$$

$$k(f,n) = \frac{1}{k(f,0)} \sum_{i} \vec{M}_{if} k(i,n-1).$$

Subsequently, by replacing the $k(f,n-1)$ component in $k(i,n)$, the $k(f,n)$ term can be eliminated with and

$$k(i,n) = \frac{1}{k(i,0)} \sum_{f} \vec{M}_{if} \frac{1}{k(f,0)} \sum_{i} \vec{M}_{if} k(i,n-2) = \sum_{i'} \tilde{M}_{ii'} k(i',n-2).$$

The research complexity index of an institute can then be found:

$$RCI(i) = \frac{\vec{K}_i - \text{mean}\left(\vec{K}_i\right)}{\sigma(\vec{K})},$$

where $\vec{K}_i$ denotes the eigenvector of $\tilde{M}_{ii'}$ (Hausmann et al., 2008) and $\sigma(\vec{K})$ denotes the standard deviation of $\vec{K}_i$.

Results and Analysis

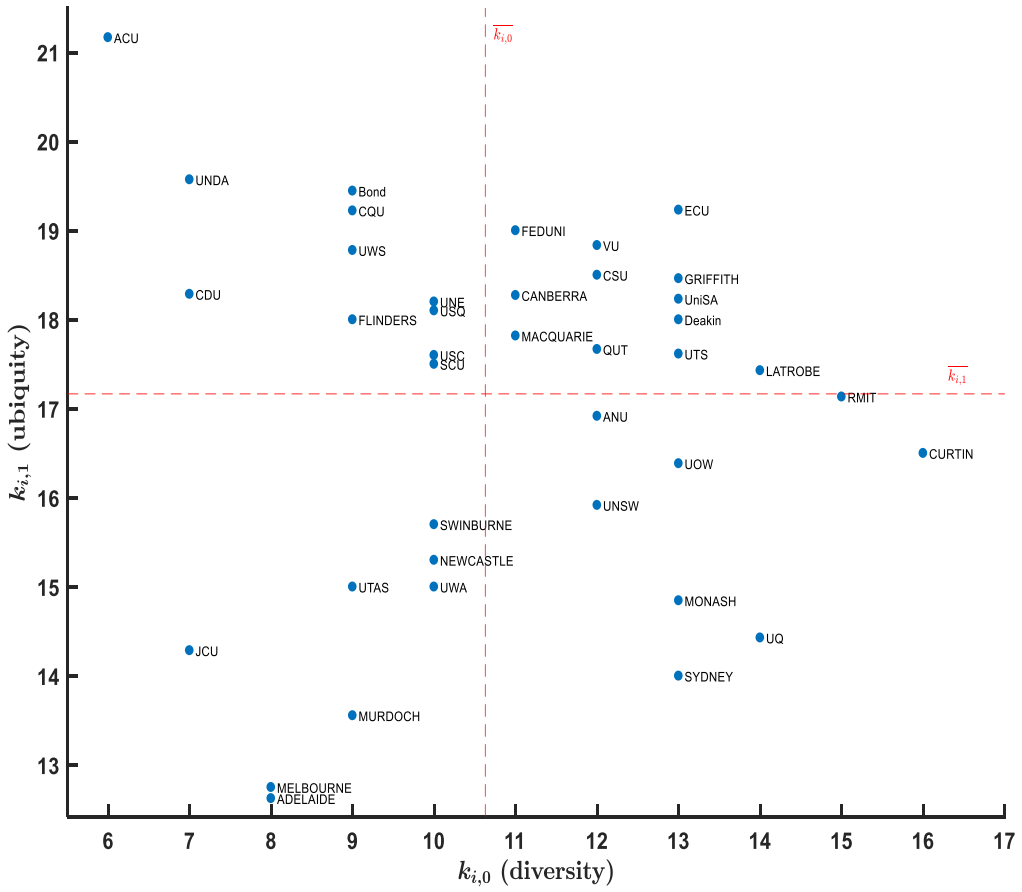

Figure 2 Research ubiquity and diversity of Australian universities

Figure 2 shows the distribution of research ubiquity and diversity among Australian universities. The mean diversity line and the mean ubiquity line divides the distribution into four quadrants. Universities appear on the left of the mean diversity line indicate that they are less diversified in their research field (in terms of the number of RCA values greater or equal to 1), vice versa for the universities appear on the right of the diversity mean. Universities sit above the ubiquity mean indicate that they are publishing in standard areas, as oppose to publishing in exclusive areas if they sit below the mean ubiquity line.

RCI is calculated according to the equation defined in the Research Complexity Analysis section. The experiment shows that Torrens University has no publication in year 2013, which yields a zero row vector in the $\vec{M}_{if}$ matrix and the Eigenvalue cannot be found. While Torrens University (formerly known as Torrents College) publishes 2 paper in Mathematics and Social Sciences in other years according to MAG, these two areas are changed to the lowest number in the RCA matrix such that Eigenvalues can be obtained with minor impact to the RCI calculations.

University	RCI
AustralianCatholicUniversity	0.12
AustralianNationalUniversity	-0.8
BondUniversity	-0.15
CentralQueenslandUniversity	-0.96
CharlesDarwinUniversity	0.67
CharlesSturtUniversity	0.34
CurtinUniversity	-1.04
DeakinUniversity	-0.54
EdithCowanUniversity	-0.98
FederationUniversity	-0.84
FlindersUniversity	0.17
GriffithUniversity	0.2
JamesCookUniversity	2.05
LaTrobeUniversity	0.18
MacquarieUniversity	-0.78
MonashUniversity	0.24
MurdochUniversity	1.96
QueenslandUniversityofTechnology	-1.05
RMITUniversity	-0.87
SouthernCrossUniversity	-0.43

University	RCI
SwinburneUniversityofTechnology	-1.07
TorrensUniversity	-0.95
UniversityofAdelaide	2.28
UniversityofCanberra	0.12
UniversityofMelbourne	2.43
UniversityofNewEngland	0.76
UniversityofNewSouthWales	-0.45
UniversityofNewcastle	0.97
UniversityofNotreDame	0.42
UniversityofQueensland	0.92
UniversityofSouthAustralia	-0.73
UniversityofSouthernQueensland	-1.13
UniversityofSydney	1.02
UniversityofTasmania	0.16
UniversityofTechnologySydney	-1.04
UniversityoftheSunshineCoast	0.86
UniversityofWesternAustralia	0.36
UniversityofWesternSydney	-0.55
UniversityofWollongong	-1.21
VictoriaUniversity	-0.66

Table 2 RCI of Australian universities

1	UniversityofMelbourne	2.43
2	UniversityofAdelaide	2.28
5	UniversityofSydney	1.02
7	UniversityofQueensland	0.92
12	UniversityofWesternAustralia	0.36
14	MonashUniversity	0.24
23	UniversityofNewSouthWales	-0.45
29	AustralianNationalUniversity	-0.8

Table 3 RCI rank and RCI scores of Go8 universities

27	UniversityofSouthAustralia	-0.73
31	RMITUniversity	-0.87
35	CurtinUniversity	-1.04
36	UniversityofTechnologySydney	-1.04
37	QueenslandUniversityofTechnology	-1.05

Table 4 RCI rank and RCI scores of ATN universities

3	JamesCookUniversity	2.05
4	MurdochUniversity	1.96
10	CharlesDarwinUniversity	0.67
15	GriffithUniversity	0.2
16	LaTrobeUniversity	0.18
17	FlindersUniversity	0.17

Table 5 RCI rank and RCI scores of IRU universities

To compare RCI of Australian universities, seminal examples grouped in well-known alliances are chosen to be examined. The Group of Eight (Go8) includes eight leading research Universities in Australia. The Australian Technology Network (ATN) includes five universities with their focus on the practical application of tertiary studies and research. The Innovative Research Universities (IRU) includes fix research universities younger in age compares to the Go8. The ranking and the RCI scores of the Go8, ATN and IRU leagues are summarized in Table 3-Table 5, with the average RCI of 0.75, -0.95, and 0.87, respectively.

According to the definition, the higher the RCI value reflects diverse and sophisticated research outcome produced by the university. For ATN with low (and negative) number, it is not a surprise as these are highly specialized in (mostly) technological research disciplines. Go8 and IRU both have relatively high RCI values. As discussed previously, the RCI values provide insight of the combined diversity and ubiquity of a university's research outcome, normalized according to the scale of the university. RCI is not a reflection of university rankings, although, it may provide some indication of strategic research priority. Exploring whether the research priority will influence the university ranking is considered a future work of this research.

The current study may be constrained due to several factors. While the MAG dataset automatically parses the affiliation details, errors may be introduced due to different writing or formatting as discussed previously. While additional information such as city or country are used to filter out unwanted entries, it is not always possible to properly identify these entries for papers lack of relevant details to differentiate universities with identical or similar names.

Another limitation is that the author attribute is missing from the MAG dataset, hence this analysis cannot associate a paper to an affiliation according to its first author or its corresponding author. As a workaround, each paper is associate to all its authors multiple times. The subject areas of a journal are also considered a limitation, as many journals include papers under more than one discipline. Due to the lack of weighting of area classification for each journal provided by Scopus, same weightings are used for each area classification.

Conclusion

In this paper, we conducted research complexity analysis for Australian universities using the journal publication record in 2013 in the MAG data set. Among the three well-known university leagues, Go8 and IRU universities are found to have higher research complexities, meaning their research outputs are diversified and sophisticated. ATN universities are found to have lower research complexities, reflecting most of their research are specialized in focused research areas. While RCI is not a reflection

of university rankings, exploration of its suggested research priority and the associated influence to future university ranking can be considered a future work of this study.

References

Al-Juboori, A.F.M.A., Na, Y., and Ko, F. (2011). University ranking and evaluation: Trend and existing approaches. In The 2nd International Conference on Next Generation Information Technology, pp. 137–142.

Balassa, B. (1965). Trade Liberalisation and "Revealed" Comparative Advantage1. Manch. Sch. *33*, 99–123.

Bornmann, L., Mutz, R., and Daniel, H.-D. (2013). Multilevel-statistical reformulation of citation-based university rankings: The Leiden ranking 2011/2012. J. Am. Soc. Inf. Sci. Technol. *64*, 1649–1658.

De Domenico, M., Omodei, E., and Arenas, A. (2016). Quantifying the Diaspora of Knowledge in the Last Century. Appl. Netw. Sci. *1*.

Dobrota, M., Bulajic, M., Bornmann, L., and Jeremic, V. (2016). A new approach to the QS university ranking using the composite I-distance indicator: Uncertainty and sensitivity analyses. J. Assoc. Inf. Sci. Technol. *67*, 200–211.

Dwivedi, Y.K., Venkitachalam, K., Sharif, A.M., Al-Karaghouli, W., and Weerakkody, V. (2011). Research Trends in Knowledge Management: Analyzing the Past and Predicting the Future. Inf. Syst. Manag. *28*, 43–56.

Etzkowitz, H. (2013). Anatomy of the entrepreneurial university. Soc. Sci. Inf. *52*, 486–511.

Furman, J.L., Porter, M.E., and Stern, S. (2002). The determinants of national innovative capacity. Res. Policy *31*, 899–933.

Guevara, M.R., Hartmann, D., Aristarán, M., Mendoza, M., and Hidalgo, C.A. (2016). The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. Scientometrics *109*, 1695–1709.

Haley, M.R. (2014). Ranking top economics and finance journals using Microsoft academic search versus Google scholar: How does the new publish or perish option compare? J. Assoc. Inf. Sci. Technol. *65*, 1079–1084.

Hausmann, R., Hidalgo, C.A., Bustos, S., Coscia, M., Chung, S., Jimenez, J., Simoes, A., and Yıldırım, M.A. (2008). The atlas of economic complexity: Mapping paths to prosperity. Obs. Econ. Complex. Harv. HKSCDIMIT Media Lab.

Hidalgo, C.A., and Hausmann, R. (2009). The building blocks of economic complexity. Proc. Natl. Acad. Sci. *106*, 10570–10575.

Jiang, X., Sun, X., and Zhuge, H. (2012). Towards an effective and unbiased ranking of scientific literature through mutual reinforcement. In Proceedings of the 21st ACM International Conference on Information and Knowledge Management, (ACM), pp. 714–723.

Lee, I., Xia, F., and Roos, G. (2017). An observation of research complexity in top universities based on research publications. In Proceedings of the 26th International Conference on World Wide Web Companion, (Perth, Australia), p.

Li, F., Miao, Y., and Ding, J. (2015). Tracking the development of disciplinary structure in China's top research universities (1998–2013). Res. Eval. *24*, 312–324.

Orduna-Malea, E., Martin-Martin, A., M. Ayllon, J., and Delgado Lopez-Cozar, E. (2014). The silent fading of an academic search engine: the case of Microsoft Academic Search. Online Inf. Rev. *38*, 936–953.

Ortega, J.L., López-Romero, E., and Fernández, I. (2011). Multivariate approach to classify research institutes according to their outputs: The case of the CSIC's institutes. J. Informetr. *5*, 323–332.

Pastor, J.M., and Serrano, L. (2016). The determinants of the research output of universities: specialization, quality and inefficiencies. Scientometrics *109*, 1255–1281.

Payumo, J.G., Arasu, P., Fauzi, A.M., Siregar, I.Z., and Noviana, D. (2014). An entrepreneurial, research-based university model focused on intellectual property management for economic development in emerging economies: The case of Bogor Agricultural University, Indonesia. World Pat. Inf. *36*, 22–31.

Reynolds, C., Agrawal, M., Lee, I., Zhan, C., Li, J., Taylor, P., Mares, T., Morison, J., Angelakis, N., and Roos, G. (2017). A sub-national economic complexity analysis of Australia's states and territories. Reg. Stud. *0*, 1–12.

Sandulescu, V., and Chiru, M. (2016). Predicting the future relevance of research institutions - The winning solution of the KDD Cup 2016.

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.P., and Wang, K. (2015). An overview of Microsoft Academic Service (MAS) and applications. In Proceedings of the 24th International Conference on World Wide Web, (ACM), pp. 243–246.

Wong, P.K., and Singh, A. (2010). University patenting activities and their link to the quantity and quality of scientific publications. Scientometrics *83*, 271–294.

Wu, D., He, D., Jiang, J., Dong, W., and Vo, K.T. (2012). The state of iSchools: an analysis of academic research and graduate education. J. Inf. Sci. *38*, 15–36.

Yoo, Y. (2013). The Tables Have Turned: How Can the Information Systems Field Contribute to Technology and Innovation Management Research? J. Assoc. Inf. Syst. Atlanta *14*, 227–236.

Appendix I – Australian journal publications on the Microsoft Academic Graph dataset

University	Abbreviation	Journal Papers
Australian Catholic University	ACU	923
Australian National University	ANU	15767
Bond University	Bond	376
Central Queensland University	CQU	763
Charles Darwin University	CDU	893
Charles Sturt University	CSU	1573
Curtin University	CURTIN	4831
Deakin University	Deakin	3823
Edith Cowan University	ECU	1019
Federation University	FEDUNI	513
Flinders University	FLINDERS	6284
Griffith University	GRIFFITH	5217
James Cook University	JCU	4957
La Trobe University	LATROBE	5543
Macquarie University	MACQUARIE	6114
Monash University	MONASH	17142
Murdoch University	MURDOCH	1736
Queensland University of Technology	QUT	2228
RMIT University	RMIT	2115
Southern Cross University	SCU	829
Swinburne University of Technology	SWINBURNE	2210
Torrens University	TORRENS	2
University of Adelaide	ADELAIDE	12050
University of Canberra	CANBERRA	1260
University of Melbourne	MELBOURNE	26707
University of New England	UNE	3244
University of New South Wales	UNSW	18308
University of Newcastle	NEWCASTLE	11580
University of Notre Dame	UNDA	154
University of Queensland	UQ	21242
University of South Australia	UniSA	3688
University of Southern Queensland	USQ	875
University of Sydney	SYDNEY	29328
University of Tasmania	UTAS	4981
University of Technology Sydney	UTS	2799
University of the Sunshine Coast	USC	468
University of Western Australia	UWA	13100
University of Wollongong	UOW	4257
Victoria University	VU	1385
University of Western Sydney	UWS	3215