IFLA WLIC 2018
KUALA LUMPUR

Submitted on: 10.07.2018

Exploring Bibliometric Mapping in NUS using BibExcel and VOSviewer

Irine Tanudjaja
Scholarly Communication, National University of Singapore Libraries, Singapore, Singapore.
E-mail address: irine.tanudjaja@nus.edu.sg

Gerrie Yu Kow
Scholarly Communication, National University of Singapore Libraries, Singapore, Singapore.
E-mail address: gerrie@nus.edu.sg

Abstract:

Citation analysis is a widely-used bibliometric method for evaluating research output. The Bibliometrics Team, a part of the Scholarly Communication Department in National University of Singapore (NUS) Libraries, has been supporting the NUS community with citation analysis since 2010, be it for Promotion and Tenure (P&T) or departmental benchmarking purposes. With the advancement of technology, bibliometrics can now be applied to other areas, for example, the mapping of the collaborations between institutions. A pilot project was initiated by the NUS Libraries Bibliometrics Team in collaboration with the NUS School of Business to map institutions which have collaborated with and which have cited the school's publications for the past 5 years (2012 to 2017) using Web of Science data. The aims of this project were twofold; to explore the new applications of bibliometrics and to deepen our existing partnerships with the research community. In this paper, we will share about this project, starting from the community engagement, methodology, prototype created, as well as the limitations and future plans.

Keywords: bibliometric mapping, bibliometric analysis, collaboration analysis.

1. Introduction

The Scholarly Communication department in NUS Libraries is comprised of three teams, namely Research Data Management, Bibliometrics and ScholarBank@NUS. The department supports research and publishing practices across disciplines, ranging from research planning, publishing, open access and the deposit of scholarly outputs in our institutional repository ScholarBank@NUS, to research impact measurement. The Bibliometrics team is responsible for the latter and has been conducting citation analysis for various purposes (e.g. Promotion and Tenure (P&T), Annual Review, departmental benchmarking, etc.) and workshops on citation analysis since 2010.

Citation analysis measures how often a publication is cited, acting as a proxy of its quality and this is just one of the many applications of bibliometrics. Narin, Olivastro and Stevens (1994) highlighted other applications of bibliometrics beyond citation analysis, such as the measuring of scientific activity in terms on the number of publications written in a certain subject area, the tracing of knowledge transfer by analysing citation, as well as the identifying of the underlying social networks in those communities through collaboration or citation.

The advancement of technology has enabled more data to be collected across all walks of life, from astronomical observatories, laboratory experiments to sales data. However, the availability of such big data would be futile if it cannot be processed for further interpretation or analysis. This need has since triggered a trend to visualize data. The notion of communicating through visuals is a very attractive one as it could help audience process large amount of data in a more accessible manner (Keim, Qu, & Ma, 2013). The visuals would capture the audience's interest and help them glean some significant patterns or other useful insights. Some readers may not consider this a novel concept as data or information have, for the longest time been presented in pie charts, x-y charts, bar charts and others. There are however some differences between data visualization and these traditional graphs. In data visualization, more indicators can be utilized to relay information, such as distance proximity, color and size. There is also an interactive component to it that lets users zoom in and out or mouse over certain parts of the visualization. This interactivity increases and changes the nature of the engagement users have with the data. Traditional graphs on the other hand still rely heavily on numbers and tend to be more static, i.e. not interactive.

The phenomenon of big data is also observed in the scientific world, where a large number of publications are being published daily. Bornmann and Mutz (2015) mentioned that the growth of scientific publications has tripled to 8 to 9% in the 2010s, as compared to 2 to 3% between the early parts of the 20th century to 2010s. Furthermore, bibliographic databases such as Web of Science and Scopus can now provide cited reference information (i.e. works listed in the bibliographies of indexed publications) from as far back as early 1900s (Clarivate Analytics, 2018) and 1970s (Elsevier, 2017) respectively. It is therefore not surprising that there is a lot of interest in visualizing or mapping these publications' metadata. There are various objects from these publications' data that can be mapped out, for e.g. journal titles, publications listed in the bibliographies, keywords, authors, institutions and even countries. There are also various relationships that can be analysed using bibliometric methods that are based on citation, bibliographic coupling, co-citation, co-authorship or co-occurrence of keywords. In this paper, the mapping of various objects from publications' metadata using bibliometric methods is referred to as bibliometric mapping.

Bibliometric mapping would benefit both the scientific community and the public in general as it could help transform the large body of publications' metadata into maps or visualizations, which are more manageable to process in order to get useful insights. The following are some case studies that use bibliometric mapping for such purposes. For e.g. Zhu and Guan (2013) and Sinkovics (2016) visualized keywords to identify research themes or clusters on specific disciplines. Eddy and Solomon (2017) mapped out the affiliations of authors from a particular journal to identify the journal's geographical coverage. Li, Zhou, Xue and Huang (2016) mapped out institutional collaborations and international collaborations as part of their framework to identify emerging technologies.

That being said, Gelman and Unwin (2013) cautioned against relying heavily on maps or visualizations to understand the data. They recommended supplementing such maps with traditional graphs and tables. The maps would trigger the initial thinking or analysis on a big picture level, while the traditional graphs or tables would provide complementary evidence to support the initial findings and enable deeper analysis.

In an effort to improve our support of the NUS research community, the team is continuously exploring new and relevant services. In the early 2018, the team collaborated with the NUS School of Business to map out institutions which have collaborated with and cited the school's publications for the past five years (2012 to 2017). In this paper, we will share about this initiative, starting from the engagement, methodology, prototype created, as well as the limitations and future plans.

2. Pilot Project

2.1. Engagement

The team aims to support the various stakeholders of the NUS community which consist of researchers, research administrators, and students. One of our goals in 2017 was to build a closer relationship with research administrators. As such, the team met up with several research administrators from various faculties to understand their roles as well as their needs and challenges. This effort culminated in an inaugural boot camp designed with their needs in mind, which were to understand the meaning of various research metrics and how to generate these metrics. Through this connection, the team had a chance to present an analysis of one of the departments in the School of Business to the school's Vice Dean of Research & PhD in January 2018. The analysis used bibliometric mapping to get insights on the department's strongest collaborators, potential collaborators, as well as funding agencies. The Vice Dean found the analysis to be interesting and potentially useful and gave the greenlight for the team to generate similar analyses for the whole school and the remaining five departments. In the midst of the project, he also provided feedback that were incorporated in the prototype submitted. For example, he recommended that the report included a list of the top business schools which have collaborated with the NUS Business School. The information will enable the school to evaluate their collaborations with the best schools.

2.2. Methodology

This section will be further segmented to discuss various aspects, starting from the software used, the dataset and the interpretation of the map.

2.2.1. Software Used

There are various software tools that have been developed for the purpose of bibliometric mapping. Pradhan (2016) provides a comprehensive overview on these various software. Two of such software were utilized in this project for specific reasons stated below:

- **BibExcel** (Persson, n.d.) is an open source tool developed by Professor Olle Persson from Umeå University, Sweden. It is specifically designed for analysing bibliographic relationships, i.e. bibliographic coupling, co-occurrence, co-authorship. It is useful as it accepts data downloaded directly from bibliographic databases, and can be used to extract the various objects available in the publication metadata. Those objects extracted will be uniquely tagged to the respective publication, and will be saved in a file format that can be imported to other tools (such as Excel, Pajek, etc.) for further processing. For the prototype, BibExcel was used for the preparation of a thesaurus file that can be inputted to VOSviewer for standardization of institution names.
- **VOSviewer** (Van Eck and Waltman, 2009) is also an open source tool developed by Nees Jan van Eck and Ludo Waltman at the Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands. It is designed to construct and visualize bibliometric maps such as co-authorship, co-occurrence, and citation based maps. Similar to BibExcel, it can also accept data from bibliographic databases and can be integrated with other tools. For the prototype, VOSviewer was used to visualize the collaborating institutions and the citing institutions maps. It was selected over other tools because of its ability to calculate a score called Average Normalized Citation that can be used as an indication of impact. This score will be discussed further in section 2.2.3.

2.2.2. Data

This project made use of data from Web of Science because of its capability to provide a more standardized format for institution names as compared to Scopus. That being said, the institution names from Web of Science also required a certain amount of cleaning up. One of the main challenges in this project was to standardize these name variations, which will be discussed in the subsequent paragraphs.

In Web of Science, the author names and their respective institution names are available in the Author Address - C1 field (Figure 1). In this project, collaborating institutions were defined as other institutions besides NUS that appeared in the C1 field of the school's publications. The citing institutions were defined as other institutions besides NUS that appeared in the C1 field of articles that cited the school's publications. In defining these collaborating and citing institutions, we noted that an author may have joint affiliations. For example, Prof. Ho listed both his affiliations (Univ Calif Berkeley and Natl Univ Singapore) in the publication below (Figure 1), and in this prototype, both affiliations received 1 count. This was one of the limitations of the prototype.

```
C1- [Ho, Teck-Hua] Natl Univ Singapore, Off Deputy President Res &
Technol, Singapore 119077, Singapore; [Ho, Teck-Hua] Univ Calif Berkeley,
Haas Sch Business, Berkeley, CA 94720 USA; [Chong, Juin Kuan] Natl Univ
Singapore, Sch Business, Singapore 119245, Singapore; [Xia, Xiaoyu]
Chinese Univ Hong Kong, Sch Business, Dept Decis Sci & Managerial Econ,
Shatin, Hong Kong, Peoples R China|
```

Figure 1 Author's address field (C1)

As mentioned above, the institution names in Web of Science could vary because some authors may list their main institution as their affiliation (i.e. Natl Univ Singapore), but some may list their department or faculty (i.e. NUS Business Sch – Figure 2). These variations were standardized, and the main institution was used.

```
C1- [Sun, Li] Zayed Univ, POB 19282, Dubai, U Arab Emirates; [Rajiv,
Surendra] Philind Mot Pictures Private Ltd, 1004 B,Hill
Grange,Hiranandani Estate, Thane W 400607, India; [Chu, Junhong] NUS
Business Sch, BIZ 1,8-34,15 Kent Ridge Dr, Singapore 119245, Singapore|
RP- Sun, L (reprint author), Zayed Univ, POB 19282, Dubai, U Arab
Emirates|
```

Figure 2 Institution's variations

Two types of identifiers were used in combination to get a comprehensive publication lists for the analysis: (i) Web of Science's article IDs of the staff publications from our research information management system called NUS Elements, and (ii) Scopus IDs of the existing staff. NUS Elements relies heavily on author identifiers, such as: Scopus Author ID, ResearcherID, etc., to claim staff's publications from various databases. However ResearcherID, Web of Science's author identifier, requires a researcher to manually create and maintain it. Scopus, on the other hand, automatically creates a Scopus Author ID on behalf of the researcher, and has an algorithm that tries to match new publications to existing Scopus Author IDs. Therefore, there is a higher probability that a Scopus Author ID would contain a more comprehensive publication list of a researcher, compared to a ResearcherID. As per the school's request, only publications classified as "Journal article" were included as part of the publication list.

2.2.3. Interpreting the map

As mentioned in the introduction, data visualization utilizes several indicators to relay information. In this prototype, the distance proximity between each node, the thickness of lines, the color and size of the nodes relay various information about the collaboration or about the citing institutions. Each node is representing an institution.

The size of the nodes in the map represents the institutions' number of publication within the publication set. The higher the number of publications within the set, the larger the nodes are. For example, University of Wisconsin has the highest number of publications compared to the rest within this publication set (Figure 3).

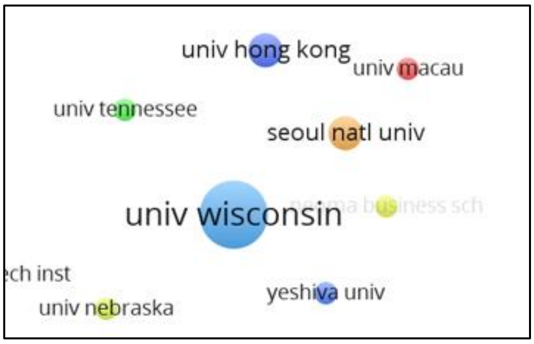

Figure 3 Size of the nodes

The strength of collaboration between two institutions in the map is portrayed by:

- Thickness of lines between two nodes, which represents the number of publications co-authored by the two institutions. It corresponds to a score called Link strength. The higher this score, the thicker the lines is.
- Closeness between nodes; distance between nodes represents the relatedness of the institutions. The closer two nodes are located, the stronger their collaboration is. In VOSviewer, the distance does not correspond to a score that is readily available.

Based on a small area of the map below (Figure 4), it can be concluded that NUS has a stronger collaboration with Hong Kong University of Science and Technology (HKUST), in comparison with University California of Berkeley (UC Berkeley). This is because the line between NUS and HKUST is thicker than the line between NUS and UC Berkeley. Similarly, the distance between NUS and HKUST is closer than the distance between NUS and UC Berkeley. This is also supported by the Link strength score. The score between NUS and HKUST is 13, while the score between NUS and UC Berkeley is only 5.

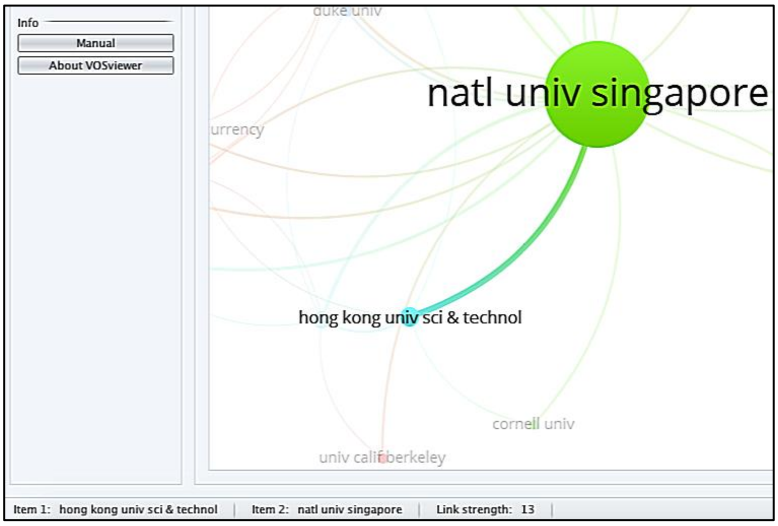

Figure 4 Strength of collaboration

The impact of publications resulting from the collaboration between NUS and other institution is represented by a score called Average Normalized Citations (Ave. Norm. Citations). It is calculated by VOSviewer and defined as the average number of citations received by publications which are co-authored by NUS and the other institution. It is also normalized by taking into consideration the age of the publication set which is inputted to VOSviewer. This normalization tries to correct the age bias since older publications tend to have more citations. Similarly, the impact of the citing publications is also represented by this score.

When the Ave. Norm. Citations has a value above 1, the value is above expectation. This means the value is above the average of all the publications used for the analysis. In the map, Ave. Norm. Citations is represented by color. The color bar available on each map indicates how scores are mapped, and the colors typically range from blue (lower than average) to green (average) to red (higher than average).

To illustrate based on a small area of the map below (Figure 5), it can be concluded that the collaboration between NUS and Leuphana University is yielding impact that is higher than average because Leuphana's node is Red in color. This is also supported by its Ave. Norm. Citations score that is above 1, i.e. 1.79.

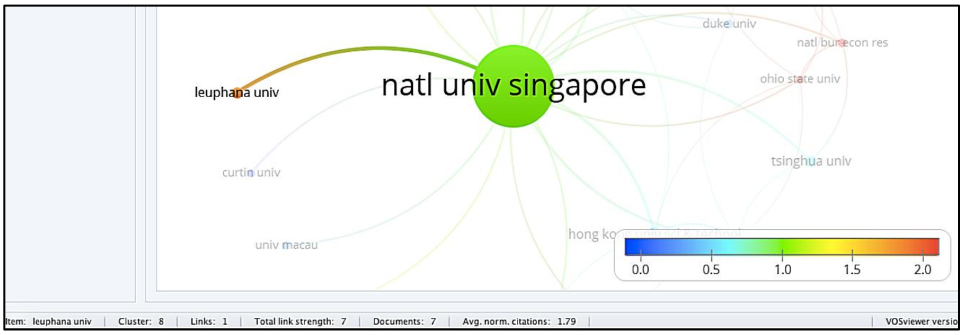

Figure 5 Impact of collaboration

2.4. Prototype

In the prototype created, two bibliometric maps were created for each department and the school itself. They are the collaborating institutions and the citing institutions maps. These maps were created to provide an overview of the school's scientific network and to provide some preliminary insights. For example, Figure 6 is the collaboration map of the school and from it, viewers can get a sense that researchers from NUS do not co-author many publications with researchers from New York University but when they do, those publications typically received higher citation counts.

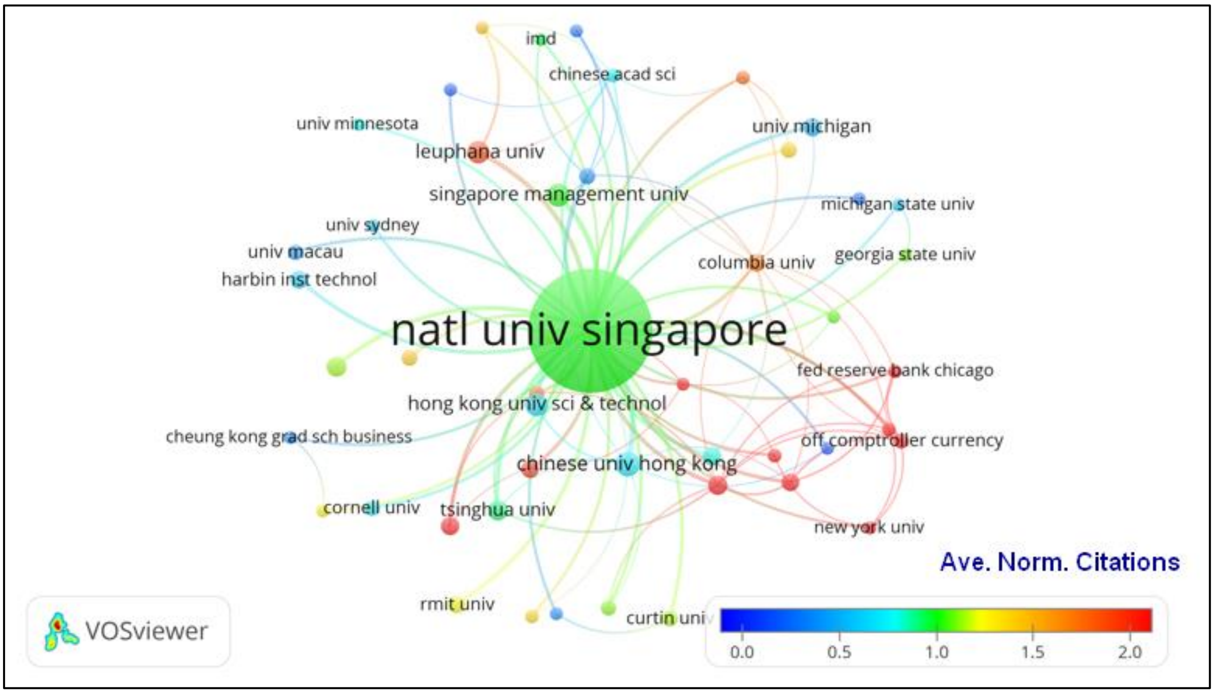

Figure 6 Collaborating institutions map

However in line with the suggestion provided by Gelman and Unwin (2013), traditional tables and charts, for example Table 1, were also included as supplementary materials for further analysis into the school's network. These are:

i. Collaborating institutions:
   - Top 10 institutions based on the strength of their collaboration with the NUS School of Business. The strength of collaboration refers to a score called Link strength, which has been discussed in section 2.2.3.
   - Top 10 institutions based on impact (Table 1). Impact refers to a score called Ave. Norm. Citations, which has been discussed in section 2.2.3.
   - Summary on the strength and impact of the collaboration with top business schools. The top 25 business schools were retrieved from the Quacquarelli Symonds (QS) world university ranking by subject: Business and Management Studies (Quacquarelli Symonds, 2018) and Times Higher Education (THE) world university ranking by subject: Business and Economics (Times Higher Education, 2018).

ii. Citing institutions:
   - Top 10 institutions based on the number of publications which have cited the school' publications.
   - Top 10 institutions based on the impact of the publications which have cited the school' publications.
   - Summary of the top business schools which have cited the school's publications.

Institutions	Link strength	Avg. norm. citations	Avg. pub. Year
caltech	2	8.5413	2016
oregon state univ	3	4.8276	2016.67
erasmus univ	2	4.6356	2014
univ florida	3	4.3359	2015.33
radboud univ nijmegen	2	4.0771	2015
univ chicago	6	3.8406	2014.67
nyu	5	3.411	2015.4
univ oxford	2	3.3617	2017
wuhan univ technol	2	3.3617	2017
fed reserve bank chicago	6	3.1883	2015.17

Table 1 Top 10 collaborating institutions based on impact

3. Limitations

The team acknowledges there are some limitations in the prototype created. These are (i) staff's joint affiliations were not taken into consideration, as such, when an author has joint affiliations, both affiliations will receive 1 count, (ii) the standardization of institution names was completed based on findings on the internet and may not be accurate, and (iii) the maps provided in the prototype were static; thus, readers won't be able to interact with it, i.e. to zoom in and out or to mouse over on certain parts of the map.

4. Future plans

The team learned many lessons over the completion of the project, such as learning how to use VOSviewer and BibExcel, but the most important one was to understand the reason why there was no follow up from the school upon submission of the prototype. In the future, the team would need to find an avenue (i.e. through a presentation) or to explore a better method (for e.g. through the use of data storytelling) to explain such visualizations and their potential applications. Nonetheless, the project also triggered positive outcome, colleagues from another team became interested in data visualization and is currently exploring another project that may benefit the research community in NUS. The team will continue to strive to look for other collaborative opportunities, be it internally or externally, so as to better serve the community in innovative and impactful ways.

Acknowledgments

The team would like to express our gratitude to Prof. Chong Juin Kuan and Ms. Lim Cheow Loo from the NUS School of Business for the opportunity to work together on this initiative and for providing insightful suggestions.

References

Bornmann, L., & Mutz, R. (2015), Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*, 2215-2222. https://doi.org/10.1002/asi.23329

Clarivate Analytics. (2018, February 26). *Web of Science: Summary Coverage*. Retrieved from https://clarivate.libguides.com/webofscienceplatform/coverage.

Eddy, M. A., & Solomon, D. (2017). Leveraging Librarian Liaison Expertise in a New Consultancy Role. *The Journal of Academic Librarianship*, *43*(2), 121-127. https://doi.org/10.1016/j.acalib.2017.01.001

Elsevier. (2017, August). *Scopus content coverage guide*. Retrieved from https://www.elsevier.com/__data/assets/pdf_file/0007/69451/0597-Scopus-Content-Coverage-Guide-US-LETTER-v4-HI-singles-no-ticks.pdf

Gelman, A., & Unwin, A. (2013). Infovis and Statistical Graphics: Different Goals, Different Looks. *Journal of Computational and Graphical Statistics*, *22*(1), 2-28. https://doi.org/10.1080/10618600.2012.761137

Keim, D., Qu, H., & Ma, K. L. (2013). Big-Data Visualization. *IEEE Computer Graphics and Applications*, *33*(4), 20-21. https://doi.org/10.1109/MCG.2013.54

Li, X., Zhou, Y., Xue, L., & Huang, L. (2016). Roadmapping an emerging industry with bibliometrics and patent analysis: A case of OLED industry in China. *International Journal of Technology Management*, *72*(1-3), 105-143. https://doi.org/10.1504/IJTM.2016.080538

Narin, F., Olivastro, D., & Stevens, K. A. (1994). *Bibliometrics/theory, practice and problems. Evaluation Review*, *18*(1), 65-76.

Persson, O. (n.d.). BibExcel [Computer software]. Retrieved from http://homepage.univie.ac.at/juan.gorraiz/bibexcel/

Persson, O., Danell, R., & Schneider, J. W. (2009). How to use Bibexcel for various types of bibliometric analysis. In F. Åström, R. Danell, B. Larsen, & J. Schneider, *Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday* (pp. 9–24). Leuven, Belgium: International Society for Scientometrics and Informetrics.

Pradhan, P. (2017). Science Mapping and Visualization Tools used in Bibliometric & Scientometric Studies: An Overview. Retrieved from http://ir.inflibnet.ac.in:8080/ir/bitstream/1944/2132/1/INFLIBNET%20NEWSLETTER%20Vol.23%20No.%204%20%28October-%20December%202016%29.pdf

Quacquarelli Symonds (QS). (2018). *University rankings for Business & Management studies*. Retrieved from https://www.topuniversities.com/university-rankings/university-subject-rankings/2018/business-management-studies

Sinkovics, N. (2016). Enhancing the foundations for theorising through bibliometric mapping. *International Marketing Review, 33*(3), 327-350. https://doi.org/10.1108/IMR-10-2014-0341

Times Higher Education (THE). (2018). *World University Rankings 2018 by subject: business and economics*. Retrieved from https://www.timeshighereducation.com/world-university-rankings/2018/subject-ranking/business-and-economics#!/page/0/length/25/sort_by/rank/sort_order/asc/cols/stats

Van Eck, N.J., & Waltman, L. (2009). VOSviewer [Computer software]. Retrieved from http://www.vosviewer.com/

Zhu, W., & Guan, J. (2013). A bibliometric study of service innovation research: based on complex network analysis. *Scientometrics*, *94*(3), 1195-1216. Retrieved from https://EconPapers.repec.org/RePEc:spr:scient:v:94:y:2013:i:3:d:10.1007_s11192-012-0888-1.