Abstract
The Arabic Citation Index (ARCI) was launched in 2020. This study gives an overview of the scientific literature available in this new database. By using metadata available in scientific publications, I analyse ARCI to characterize the scientific literature published in Arabic. First, I describe the data and the methods used in the analyses. As of October 2020, ARCI indexed 65,208 records covering the 2015-2019 period. Second, I explore the literature distributions at various levels (research domains, countries, languages, open access). Close to 99% of documents indexed are articles. Results reveal the concentration of publications in the Arts & Humanities and Social Sciences fields. Most journals indexed in ARCI are currently published from Egypt, Algeria, Iraq, Jordan and Saudi Arabia. Around 7% of publications in ARCI are published in languages other than Arabic. Then, I use an unsupervised machine learning model, LDA (Latent Dirichlet Allocation) and the text mining algorithms of VOSviewer to uncover the main topics in ARCI. These methods are particularly useful to better understand the topical structure of ARCI. Finally, I suggest few research opportunities after discussing the results of this study.

Keywords
Arabic Citation Index (ARCI), Web of Science, Middle-East North Africa, Regional analysis

Introduction
Research excellence is often equated to publishing in English in high impact factors journals, as stated in the third principle of the Leiden Manifesto (Hicks, Wouters, Waltman et al., 2015). This is problematic for the humanities and social sciences where research tends to be more engaged on national issues and published in local languages. Identifying the peer-reviewed journals of regional relevance and importance is a major issue for all scientific stakeholders. Protecting excellence in locally relevant scientific research is key to preserve fields which have regional or national dimensions.

In 2020, the Arabic Citation Index (ARCI) was launched in the Web of Science platform, first in Egypt and later in the rest of the Arab World. Clarivate Analytics partnered with the Egyptian Knowledge Bank (EKB), as part of the Egyptian Ministry of Education, to power the first Arabic Citation Index. This launch is also part of Egypt’s Vision 2030 where knowledge, innovation and scientific research are key pillars to achieve scientific excellence (Egyptian Government, 2016). As mentioned by Dr. Shawki, Minister of Education & Technical Education in Egypt and President of the EKB Project, the “aim is to work toward becoming a more knowledgeable Egyptian community that encourages learning as a part of everyday life. We look forward to building our economy and exporting our sciences globally in the Arabic language.” (Clarivate Analytics, May 2018).

The focus of ARCI is on the 22 countries of the Arab League and their scholarly research. ARCI joins other regional citation indexes, using the same core features with an Arabic language interface. The criteria for inclusion in ARCI are a subset of the Web of Science Core Collection criteria (Clarivate Analytics, 2019). The journals covered in ARCI are selected by a newly established Editorial board with members from Arab League countries who provide subject knowledge and regional insights. The selection process for ARCI is based on traditional scientific publishing standards and the research norms of the Arab region. First, there is an
initial triage to confirm content accessibility and format for all titles considered for indexation in ARCI. All journals must have an ISSN. Several elements are evaluated in this first step: journal title, publisher information, URL for online journals, content access, DOI/pagination and timeliness/volume. Next, the journals are reviewed from an editorial perspective. In this second step, each journal is evaluated to confirm it provides scholarly content, with a clear scope statement, article abstracts, cited references, content relevance with the stated scope or mission, quality of language consistent with scientific communications and an editorial board reflective of the field of the journal.

Such indexing will provide more visibility to research published in Arabic facilitating contribution to local and international research efforts. ARCI is still relatively new and very little is known about this new database. It is still unfamiliar to many researchers. Therefore, the purpose of this study is to characterize the literature available in this new citation index. First, I describe the data and methods used in the analyses. Next, I explore few content distributions at various levels (research domains, countries, languages, open access). Then, I analyse the main topics covered in ARCI by using the Latent Dirichlet Allocation model and the text mining algorithms of VOSviewer (van Eck & Waltman, 2010). Finally, I discuss the results of this study, identify its limitations and suggest few research directions.

**Data and Methods**

*Data*

ARCI has a coverage back to 2015. ARCI data was extracted on October 5th, 2020. I have excluded 2020 since the year was not complete yet. Full records were exported from the Web of Science platform. The dataset under study consists of 65,208 records for the 2015-2019 period.

Table 1 lists the distribution of records by publication year as well as the share they represent within ARCI.

| Publication Years | Records | Share (%) |
|-------------------|---------|-----------|
| 2019              | 14,198  | 21.8      |
| 2018              | 15,307  | 23.5      |
| 2017              | 14,774  | 22.7      |
| 2016              | 12,013  | 18.4      |
| 2015              | 8,916   | 13.7      |

This database is well structured with 48 fields of information in each record allowing multiple bibliometric analyses (e.g. Publisher Information, Funding Information, Research Area, Open Access Indicator, Cited References, Citations, Usage Counts, ESI Highly Cited Paper/Hot Paper). In addition to essential metadata available in English as in the Web of Science Core Collection, ARCI has some specific information written in Arabic such as authors names, article title, publication name, author keywords, abstract and author address. ARCI records also show the ARCI times cited and the Total Times Cited Count (Web of Science Core Collection, Arabic Citation Index, BIOSIS Citation Index, Chinese Science Citation Database, Data Citation Index, Russian Science Citation Index, SciELO Citation Index) as well as the Cited References and the Cited Reference Count.
Methods

In this study, I use bibliometric methods to characterize the literature indexed in ARCI. The objective is to examine the local research landscape from various perspectives. Such analyses can help research managers and policy makers to better understand local research activity. I conducted a bibliometric analysis to study the journal distribution across countries and scientific productivity of countries. Then, I explored the distribution of publications by research fields, languages and access types.

Many machine learning algorithms have also been developed to understand, group or search information from large text databases. Such approaches have been frequently used to examine the structure of an aggregated literature. In natural language processing, a topic model is a statistical model to discover the abstract hidden semantic structures or topics that occur in a collection of documents. Latent Dirichlet Allocation (LDA) is a model proposed by Blei, Ng, and Jordan (2003) used to classify text in a document to a topic. Previous studies have shown that LDA performed well to understand the topical structure of a scientific corpus (Han, 2020; Suominen & Toivanen, 2016; Yau, Porter, Newman et al., 2014).

LDA is a generative probabilistic model of a corpus. The basic idea is that publications are composed of groups of words with no sequential relationship between them. As documents can include multiple topics, there is a probability of topic distribution for each topic. Each record can be described by a distribution of topics. Each topic is characterized by a distribution over words, described as a distribution of terms in a fixed vocabulary. Blei et al. (2003) represented the LDA topic model with the graphical model shown in Figure 1.

There are many tools available for LDA. In this study, I applied the LatentDirichletAllocation model available in the Scikit-learn.org python library to perform LDA.

Results

In this section, I report the main findings of the study. First, I analyse the research domains in ARCI by number of records and the proportion they represent in the database. Next, I present the journals distribution by countries. Then, the most productive countries are examined, followed by an analysis of the languages of publications and their access types. Finally, I focus on the main topics covered in the Arabic scientific literature indexed in ARCI.

Research areas distribution

Research Areas constitute a subject categorization scheme that is shared by all Web of Science product databases. This is particularly helpful when analysing documents from multiple databases related to the same research areas. All 153 research areas in the Web of Science are grouped into five broad categories: Arts & Humanities, Life Sciences & Biomedicine, Physical Sciences, Social Sciences, and Technology.

I relied on the journal category and not on the topic covered in the individual publications to analyse the disciplinary coverage in ARCI. These categories or areas, which are defined at the journal level, are used as proxies for scientific fields.

The ARCI records relate to 21 research areas in the dataset under study. Currently, 11,070 records (around 17% of ARCI), do not contain data in the Research Area field. In Figure 1, I summarize the number of records within each of research area and the share of the database they represent. I have limited to the 16 research areas with a share higher than 0.5%.
I have also summarized the shares of the number of papers within each of the five broad domains in Figure 2.

This figure shows ARCI contains mainly journals in the Arts & Humanities, and Social Sciences categories. These categories represent 79% of ARCI total coverage. Journals in Life Sciences & Biomedicine account for 4% of the coverage. As mentioned earlier, 17% of records
retrieved do not contain information about the research information. It is worth noting there is no journal related to Technology or Physical Science categories. This confirms the current focus of ARCI. Local issues in Arts & Humanities as well as Social Sciences dominate the ARCI coverage.

ARCI also offers its own research categories. We retrieve similar results. However, some differences emerge. When analysing the records with the ARCI classification, only 242 records do not contain information about the research categories, representing less than 0.4% of the total database. Some categories standout such as Islamic Studies, Islamic Jurisprudence, Islamic Creed, Poetry and Hadith which are fields well studied in the Arab region.

Content coverage by countries
In this section I analyse the coverage by country. First, I examine the types of documents indexed in ARCI. Table 2 lists the number of documents per type and they share they represent in the database.

| Document Type       | Records | Share (%) |
|---------------------|---------|-----------|
| Article             | 64,467  | 98.864    |
| Review              | 432     | 0.662     |
| Editorial           | 104     | 0.159     |
| Art and Literature  | 97      | 0.149     |
| Other               | 62      | 0.095     |
| Bibliography        | 43      | 0.066     |
| Meeting             | 3       | 0.005     |

ARCI is primarily composed of articles. Close to 99% of documents indexed are articles. Other document types all represent less than 1% of the database.

Now, I focus on the distribution of journals over countries, where each journal is assigned to a country based on the country in which the publisher is located. Before analysing the country distribution in ARCI, I examined the coverage of Arab journals in the various citation indices in the Web of Science Core Collection (WoS CC): Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI), Arts and Humanities Citation Index (AHCI) and Emerging Sources Citation Index (ESCI). This coverage is represented in Figure 3.

As of October 2020, 21,419 journals were indexed in WoS CC. 145 journals (or 0.68%) are published in 13 of the 22 Arab League countries: 67 in SCIE, 1 in SSCI and 77 in ESCI. The United Arab Emirates (UAE), Egypt and Saudi Arabia are the three most represented Arab countries in WoS CC with a total of 113 journals and a share of 78% of all journals published in the Arab region and indexed in WoS CC).

The distribution by country of publisher in ARCI is represented in Figure 4. ARCI indexes content from 19 of the 22 Arab League countries. Content for Djibouti, Comoros and Somalia is not indexed yet. Journals published in Egypt, Algeria, Iraq, Jordan and Saudi Arabia represent 79% of the journals indexed in ARCI.
Figure 3. Number of journals by Arab country in WoS CC citation indices (September 2020)

Figure 4. Share (%) of journals by country in ARCI (2015-2019)
Egypt and Algeria cumulate more than half of the journals indexed in ARCI. There is currently a concentration of journals published from these two countries in ARCI. The submission process is managed by the Egyptian Knowledge Bank (www.arcival.ekb.eg) and journals are evaluated according to the ARCI selection process as explained in the introduction of this paper. Several Arab countries have also set up national journal platforms and initiatives. For example, the Algerian Scientific Journal Platform (www.asjp.cerist.dz) and the portal of Moroccan scientific journals (www.revues.imist.ma) have been developed by their respective Ministry of Higher Education and Research. The common goal of such initiatives is to improve the visibility of local journals by improving their publishing standards.

ARCI is still new and is still growing. It will be interesting to analyse how this new citation index evolves over time in terms of content coverage by journals’ countries.

Languages coverage
Table 3 shows the coverage of records in terms of language of publications in ARCI. Arabic obviously dominates the database with 60,439 publications, representing a share of around 93%.

| Language  | Records | Share (%) |
|-----------|---------|-----------|
| Arabic    | 60,439  | 92.69     |
| English   | 3,221   | 4.94      |
| French    | 1,406   | 2.16      |
| Spanish   | 35      | 0.05      |
| German    | 32      | 0.05      |
| Persian   | 19      | 0.03      |
| Hebrew    | 18      | 0.03      |
| Italian   | 16      | 0.03      |
| Russian   | 11      | 0.02      |
| Chinese   | 15      | 0.01      |
| Turkish   | 5       | 0.01      |
| Amazigh   | 1       | -         |

The second most represented language is English with 3,221 records (4.94%) then followed by 1,406 publications in French (2.15%). 142 publications in 9 other languages represent 0.22% of this database.

As ARCI aims to provide more exposure to journals published in the Arab League countries, it is no surprise to see Arabic as the dominant language. It is also worth reminding many journals indexed in ARCI provide publication in multiple languages. Several countries from the Arab League are former British or French colonies which explains why English and French are the main non-Arabic languages in ARCI. Other languages suggest research published in ARCI journals might also tackle regional issues of interest with neighbour countries.

Open access
The last few years have seen the development of several open access (OA) options (Bosman & Kramer, 2018; Lewis, 2012). Several scholars have studied the advantage of OA in terms of readership as well as citation impact (Basson, Blanckenberg, & Prozesky, 2020; Cintra, Furnival, & Milanez, 2018; Morillo, 2020; Piwowar, Priem, Lariviere et al., 2018; Riera &
Since 2014, Web of Science has provided information to identify publications from OA journals. I use this information to analyse the access type of records indexed in ARCI. The statistics for various OA types and non-OA records in ARCI are presented in Table 4.

Table 4. Number and share of records by Open-Access type in ARCI (2015-2019)

| Open access type     | Records | Share (%) |
|----------------------|---------|-----------|
| Non-OA               | 49,405  | 75.77     |
| All Open Access      | 15,772  | 24.19     |
| Bronze               | 11,029  | 16.91     |
| DOAJ Gold            | 2,927   | 4.49      |
| Other Gold           | 1,190   | 1.82      |
| Green Published      | 623     | 0.96      |
| Green Accepted       | 3       | 0.00      |
| Unknown (no DOI)     | 31      | 0.00      |

Close to 24% of papers indexed in ARCI and published between 2015 and 2019 are openly accessible. This is below the share of 31.8% of OA documents in the Web of Science Core Collection for the same period. We notice the various OA types have different shares in ARCI. Bronze is the main OA type with 11,029 papers representing close to 17% of ARCI. DOAJ Gold has the second highest OA share (4.49%) in this database with 2,927 papers published with this OA type.

Main topics

Before applying LDA, one must define the number of topics for the corpus. One option is to examine the performance of text clustering on a small dataset. Another way is to choose the number of topics based on judgments or tests (Blei et al., 2003). In this study, the corpus is organized into 10 topics which are listed in Table 5.

The model is applied on the combinations of words available in the title, abstract and author keywords of all records indexed in ARCI. I limited my study to words in English. All records have at least the title written in English. Titles, abstracts and keywords written in Arabic and other languages are not analysed in this paper.

Table 5. 10 topics found in ARCI (2015-2019)

| #  | Topic                                | #  | Topic                                |
|----|--------------------------------------|----|--------------------------------------|
| 1  | design islamic role                  | 6  | students skills learning             |
| 2  | social media study                   | 7  | algeria economic study               |
| 3  | iraq study administrative             | 8  | physical effect players              |
| 4  | saudi arabia education                | 9  | comparative study law                |
| 5  | children psychological relationship   | 10 | international model heritage         |
It is straightforward to interpret the topics generated by the LDA model. The results are useful to understand the topical structure of ARCI by highlighting the main topics covered in the Arabic scientific literature indexed in ARCI.

**Term map**

When applying the LDA model on a corpus, it is assumed one document can address multiple topics. As shown in Table 5, this is helpful to have a precise understanding of the topical structure of a large corpus. However, it does not map the relationships between topics. The purpose of building a so-called term map of the publications in ARCI is to further clarify their contents. I used VOSviewer (van Eck & Waltman, 2010) to create such map.

Titles, abstracts as well as author keywords have been concatenated into a single string which has been used by the text mining algorithms of VOSviewer. I have limited this analysis to the terms which occur at least 15 times. 954 keywords have satisfied this threshold. For each of the 954 keywords, the total strength of the co-occurrence links with other keywords has been calculated by VOSviewer. Figure 5 shows the co-occurrence network for all the terms indicating for each pair of terms the number of papers in which these terms appear together.

![Figure 5. Term map highlighting the main topics in ARCI (2015-2019)](image)

The terms have been clustered into 9 clusters with different colours. The map confirms a broad coverage of scientific literature as shown previously in the topic analysis. The clustering is useful in delineating the topics covered as well as highlighting the relatedness between them.

The horizontal and vertical axes have no meaning. The size of a term reflects the number of records in which this specific term is mentioned. The proximity of two terms is an indicator of how these terms are related based on the number of co-occurrences. In general, groups of terms closely located together can be interpreted as topics. For readability purpose, labels are shown only for selected terms to avoid overlapping labels.
Discussion

The main objective of this study was to examine the structure of ARCI, the first Arabic Citation Index. 290 Arabic journals are now indexed in ARCI as of October 2020. This indexation brings several benefits to the scientific community. This new index will improve the visibility of Arabic journals by making them more accessible. All journals indexed in ARCI need to meet selection criteria and essential publications metadata are provided. Such a database could greatly enhance the scholarly literature search. As a result, this will also help researchers to identify critical and influential research published in Arabic by providing access to highly curated peer-reviewed scientific content. Since research evaluation increasingly implies the bibliometric analysis of research output (Wilsdon, Allen, Belfiore et al., 2015; Wouters, Thelwall, Kousha et al., 2015), ARCI could also provide useful bibliometric data sources to research managers for science assessment and research analysis. This would be helpful to identify and reward excellence in locally relevant research (Hicks et al., 2015).

Now, I discuss in detail the main findings identified in this analysis. This study reveals that ARCI contains mainly journals in the Arts & Humanities, and Social Sciences categories. These categories represent 79% of ARCI’s total coverage. ARCI has a strong focus on local issues in Arts & Humanities as well as Social Sciences. It is important to keep in mind the well documented limitation on subject delineation where I used research areas of journals as proxies of categories to characterize the subject coverage.

As of October 2020, ARCI indexes content from 19 of the 22 Arab League countries, with more than half of the journals indexed in ARCI being published in Egypt and Algeria. Since ARCI is still new and under development, it will be interesting to track its evolution over time. It would also be interesting to analyse the contribution of each Arab League country to ARCI by using the author affiliations.

Unsurprisingly, ARCI has a great share of papers published in Arabic (93% of the database). However, English and French are two other languages well represented in ARCI. Other languages suggest research published in ARCI journals may also tackle regional issues of interest with neighbour regions such as Europe and Asia.

Around 24% of the content indexed in ARCI is openly accessible. This share is below the proportion of Open Access (OA) documents in the Web of Science Core Collection for the same period. The OA information available in ARCI is particularly useful to better share scientific knowledge as well as to track the adoption of local OA mandates by research managers.

The topic analysis as well as the term map are helpful to better understand the underlying structure of ARCI. Such techniques provide a great overview of the topics covered in this database. Overall, the clusters found with VOSviewer seem to be closely related and show a broad coverage of ARCI. The title, abstract and keywords in Arabic were not included in the topic analysis. It might be interesting to characterize the literature in ARCI by focusing on the Arabic content as well.

In conclusion, this paper offers a brief profile of the newest citation index in the Web of Science. This paper contributes also to the literature on regional citation indices (Huang, Zhu, Lv et al., 2017; Jin & Wang, 1999; Leydesdorff & Jin, 2005; Moskaleva, Pislyakov, Sterligov et al., 2018; Pajic, 2015; Seol & Park, 2008; Velez-Cuartas, Lucio-Arias, & Leydesdorff, 2016). One common purpose of such regional databases is to provide more visibility to local journals. ARCI
is likely to have positive effects on regional research discovery as well as research management in the Arab region. Future research may seek to propose detailed mappings of ARCI to better understand its structure and impact. Finally, it will also be interesting to track its development and evolution by using dynamic topic models to study the time evolution of topics by using the text available in English as well as Arabic.

**Competing Interests**
The author is an employee of Clarivate Analytics, the provider of the Web of Science and the Arabic Citation Index.

**References**
Basson, I., Blanckenberg, J. P., & Prozesky, H. (2020). Do open access journal articles experience a citation advantage? Results and methodological reflections of an application of multiple measures to an analysis by WoS subject areas. *Scientometrics*, 26.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3*(Jan), 993-1022.

Bosman, J., & Kramer, B. (2018). Open access levels: a quantitative exploration using Web of Science and oaDOI data. PeerJ.

Cintra, P. R., Furnival, A. C., & Milanez, D. H. (2018). The impact of open access citation and social media on leading top Information Science journals. *Investigacion Bibliotecologica, 32*(77), 117-132.

Clarivate Analytics. (2019). Introducing the Arabic Citation Index. Retrieved on November 3, 2020 from: https://clarivate.com/webofsciencegroup/campaigns/arabic-citation-index/

Clarivate Analytics. (May 2018). Clarivate Analytics Partners with the Egyptian Knowledge Bank to Power the First Arabic Citation Index. Retrieved on November 3, 2020 from: https://www.prnewswire.com/news-releases/clarivate-analytics-partners-with-the-egyptian-knowledge-bank-to-power-the-first-arabic-citation-index-300655525.html

Egyptian Government. (2016). *Egypt's Vision 2030*. Retrieved on November 3, 2020 from: https://mped.gov.eg/EgyptVision?lang=en

Han, X. (2020). Evolution of research topics in LIS between 1996 and 2019: an analysis based on latent Dirichlet allocation topic model. *Scientometrics, 125*(3), 2561-2595.

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature, 520*(7548), 429-431.

Huang, Y., Zhu, D., Lv, Q., Porter, A. L., Robinson, D. K. R., & Wang, X. (2017). Early insights on the Emerging Sources Citation Index (ESCI): an overlay map-based bibliometric study. *Scientometrics, 111*(3), 2041-2057.

Jin, B., & Wang, B. (1999). Chinese science citation database: Its construction and application. *Scientometrics, 45*(2), 325-332.

Lewis, D. W. (2012). The Inevitability of Open Access. *College & Research Libraries, 73*(5), 493-506.

Leydesdorff, L., & Jin, B. H. (2005). Mapping the Chinese Science Citation Database in terms of aggregated journal-journal citation relations. *Journal of the American Society for Information Science and Technology, 56*(14), 1469-1479.

Morillo, F. (2020). Is open access publication useful for all research fields? Presence of funding, collaboration and impact. *Scientometrics, 125*(1), 689-716.

Moskaleva, O., Pisyakov, V., Sterligov, I., Akoev, M., & Shabanova, S. (2018). Russian Index of Science Citation: Overview and review. *Scientometrics, 116*(1), 449-462.

Pajic, D. (2015). The Serbian Citation Index: Contest and Collapse. In A. A. Salah, Y. Tonta, A. A. A. Salah, C. Sugimoto & U. Al (Eds.), *Proceedings of Issi 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference* (pp. 604-605). Leuven: Int Soc Scientometrics & Informetrics-Issi.

Piwowar, H., Priem, J., Lariviere, V., Alperin, J. P., Matthias, L., Norlander, B., et al. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *Peerj, 6*.
Riera, M., & Aibar, E. (2013). Does open access publishing increase the impact of scientific articles? An empirical study in the field of intensive care medicine. *Medicina Intensiva, 37*(4), 232-240.

Seol, S. S., & Park, J. M. (2008). Knowledge sources of innovation studies in Korea: A citation analysis. *Scientometrics, 75*(1), 3-20.

Suominen, A., & Toivanen, H. (2016). Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification. *Journal of the Association for Information Science and Technology, 67*(10), 2464-2476.

Tang, M., Bever, J. D., & Yu, F.-H. (2017). Open access increases citations of papers in ecology. *Ecosphere, 8*(7).

Torres-Salinas, D., Robinson-Garcia, N., & Moed, H. F. (2019). Disentangling gold open access *Springer handbook of science and technology indicators* (pp. 129-144): Springer.

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523-538.

Velez-Cuartas, G., Lucio-Arias, D., & Leydesdorff, L. (2016). Regional And Global Science: Publications From Latin America And The Caribbean In The Scielo Citation Index And The Web of Science. *Profesional De La Informacion, 25*(1), 35-46.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., et al. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*.

Wouters, P., Thelwall, M., Kousha, K., Waltman, L., de Rijcke, S., Rushforth, A., et al. (2015). The metric tide. *literature review, Supplementary Report I to the Independent Review of the Role of Metrics in Research Assessment and Management, HEFCE, London*.

Yau, C.-K., Porter, A., Newman, N., & Suominen, A. (2014). Clustering scientific documents with topic modeling. *Scientometrics, 100*(3), 767-786.

Young, J. S., & Brandes, P. M. (2020). Green and gold open access citation and interdisciplinary advantage: A bibliometric study of two science journals. *Journal of Academic Librarianship, 46*(2).