

Bibliometric Analysis using R for Non-Coders

A practical handbook in conducting bibliometric analysis studies using Biblioshiny
for Bibliometrix R package.



Aidi Ahmi



Table of Contents

List of Figures	ix
List of Tables	xiv
List of Abbreviations	xv
Preface	xvii
About Author	xix
Chapter One	1
Introduction	1
An Overview of Bibliometrics Analysis	3
Examining the Trends	5
Evaluating the Impacts	5
Analysing the Structure	6
History of Bibliometrics	7
Study's Aims, Objectives and Questions	10
Chapter Two	15
Getting the Dataset	15
Identifying the Topic	17
Search Strategy	18
Phrase Searching	21
Loose/approximate phrase	21
Exact Phrase	21
Boolean Operators	21
Wildcard Symbols	23
Proximity Operators	23
Refining and Exporting the Documents	27
Data Sources	31
Scopus	33
Web of Science (WoS)	34
Dimensions	36

Lens	37
PubMed	39
Cochrane Library	41
Materials and Methods	42
Chapter Three	48
Installing R and Uploading the Dataset	48
Installing R	48
Installing RStudio	49
Installing Bibliometrix Package	50
Biblioshiny for Bibliometrix	54
Features	55
Workflow	55
Uploading Data	56
Harmonising the Data	61
Filter	64
Troubleshoot Error Messages	65
Chapter Four	68
Harmonising Data using OpenRefine	68
Installing OpenRefine	69
Creating a New Project	71
Split Multi-Valued Cells	76
Faceting Data	78
Join Multi-Valued Cells	81
Moving Column	83
Exporting the File	84
Increase OpenRefine Memory	85
Chapter Five	88
Research Trend and Impact	88
Dataset	89
Main Information	89
Main Information About Data	89
Document Types	91
Document Contents	94
Authors	95
Author's Collaboration	96

Annual Scientific Production	98
Average Citations per Year	100
Three-Fields Plot	104
Sources Analysis	106
Most Relevant Sources	107
Most Local Cited Sources	108
Bradford's Law	111
Source Impact	113
H-index	115
G-index	116
M-index	117
Total Citation	118
Source Dynamics	118
Authors Analysis	119
Most Relevant Authors	120
Most Local Cited Authors	122
Authors' Production Over Time	122
Lotka's Law	123
Authors Impact	125
Affiliations Analysis	127
Most Relevant Affiliations	127
Countries Analysis	128
Corresponding Author's Country	129
Country Specific Production	131
Most Cited Countries	134
Documents Analysis	136
Most Global Cited Documents	136
Most Local Cited Documents	141
Cited References	147
Most Local Cited References	147
References Spectroscopy	148
Words Analysis	152
Most Frequent Words	152
Word Cloud	157
Tree Map	162
Word Dynamics	165
Trend Topics	167
Chapter Six	170
Knowledge Structures	170
Clustering	170

Clusters by Documents Coupling	172
Clusters by Authors Coupling	177
Clusters by Sources Coupling	178
Conceptual Structure	179
Co-occurrence Network	180
Thematic Map	186
Thematic Evolution	193
Factorial Analysis	200
Word Map	201
Topic Dendrogram	202
Factorial Map of Most Contributing Papers	205
Factorial Map of Most Cited Papers	205
Words by Cluster	206
Articles by Cluster	207
Intellectual Structure	208
Co-citation Network	209
Historiography	214
Social Structure	216
Collaboration Network	217
Collaboration World Map	222
References	224
Index	237



List of Figures

Figure 1.1	Bibliographic data that can be extracted from the Scopus database	3
Figure 2.1	Topic identification in educational technologies area of study	17
Figure 2.2	Scopus document search	20
Figure 2.3	Refining search results	27
Figure 2.4	Filter by year	28
Figure 2.5	Green procurement literature from 2011-2021	29
Figure 2.6	Export document settings	30
Figure 2.7	Scopus preview	34
Figure 2.8	Web of Science login page	36
Figure 2.9	app.dimensions.ai	37
Figure 2.10	lens.org scholar search	39
Figure 2.11	pubmed.gov	41
Figure 2.12	cochranelibrary.com	42
Figure 2.13	Research protocol	44
Figure 2.14	Methodological approach	45
Figure 2.15	Flowchart of data collection and refinement process	46
Figure 3.1	Installation of Bibliometrix R Package	51
Figure 3.2	Run command library(bibliometrix)	52
Figure 3.3	Run command biblioshiny()	53
Figure 3.4	biblioshiny: The shiny app for bibliometrix	54
Figure 3.5	Workflow of biblioshiny as suggested by Aria and Cuccurullo (2017)	56
Figure 3.6	Import of Load Files	57
Figure 3.7	Load bibliometric file(s)	60
Figure 3.8	Filter	64
Figure 3.9	Select document type	65
Figure 4.1	OpenRefine 3.5.0 Welcoming Page	71

Figure 4.2	OpenRefine Data Preview – scopus.csv file	72
Figure 4.3	OpenRefine Data Preview – Bibliometric Exported File	73
Figure 4.4	OpenRefine Project – scopus.csv file	74
Figure 4.5	OpenRefine Project – Bibliometrix Exported File	74
Figure 4.6	Split multi-valued cells menu	76
Figure 4.7	How to split multi-valued cells	77
Figure 4.8	Separate row being created for each author’s keyword	78
Figure 4.9	Facet menu	78
Figure 4.10	Text facet of DE’s column	79
Figure 4.11	Cluster and edit column “DE”	80
Figure 4.12	Sample of facet	81
Figure 4.13	Join multi-valued cells	82
Figure 4.14	Move column right	83
Figure 4.15	AU column now in the second column	84
Figure 4.16	Exporting the file	85
Figure 5.1	Main information about the collection	98
Figure 5.2	Annual Scientific Production	100
Figure 5.3	Average Citations per Year	101
Figure 5.4	Total publication and total citations per year	103
Figure 5.5	Three-Fields Plot of authors, references and keywords (before harmonising)	105
Figure 5.6	Three-Fields Plot of authors, references and keywords (after harmonising)	106
Figure 5.7	Most relevant sources	107
Figure 5.8	Most local cited sources (from references list) plot – uncleaned data	109
Figure 5.9	Most local cited sources (from references list) table – uncleaned data	110
Figure 5.10	Top 20 of the most local cited sources – cleaned data	111
Figure 5.11	Bradford’s law	112
Figure 5.12	Bradford’s law zone	113
Figure 5.13	h-index	116
Figure 5.14	h-index and g-index chart	117
Figure 5.15	Source growth	119
Figure 5.16	Most relevant authors	120
Figure 5.17	Most local cited authors	122

Figure 5.18	Authors' Production over Time	123
Figure 5.19	Author productivity through Lotka's law	124
Figure 5.20	Author local impact by h-index	125
Figure 5.21	A sample of unstandardised affiliation	127
Figure 5.22	Most relevant affiliations	128
Figure 5.23	Corresponding author's country	129
Figure 5.24	Country specific production	131
Figure 5.25	Alternative map presentation	133
Figure 5.26	Most cited countries (based on total citations)	134
Figure 5.27	Most cited countries (based on average citations per year)	134
Figure 5.28	Search documents	137
Figure 5.29	Most global cited documents	137
Figure 5.30	Cited Documents	141
Figure 5.31	Most local cited documents	142
Figure 5.32	Most local cited references	147
Figure 5.33	CR that required data cleaning	148
Figure 5.34	References publication year spectroscopy	149
Figure 5.35	RPY table	150
Figure 5.36	Cited references table	150
Figure 5.37	CRExplorer sample output of RPYS	151
Figure 5.38	Most frequent Keywords Plus	153
Figure 5.39	Most frequent author's keywords (required data cleaning)	153
Figure 5.40	Most frequent author's keywords (after data cleaning) – Plot view	154
Figure 5.41	Most frequent author's keywords (after data cleaning) – Table view	154
Figure 5.42	Most frequent words based on the article's titles (Unigrams)	155
Figure 5.43	Most frequent words based on the article's titles (Bigrams)	156
Figure 5.44	Most frequent words based on the article's titles (Trigrams)	156
Figure 5.45	Word cloud of the Keywords Plus	158
Figure 5.46	Word cloud of the Author's Keywords	158
Figure 5.47	Word cloud of the article title (Trigrams)	159
Figure 5.48	Word cloud of the abstract (Trigrams)	159
Figure 5.49	Word cloud of the author's keyword for the period before from 1975-2010	160

Figure 5.50	Word cloud of the author’s keyword for the period before from 2010-2021	160
Figure 5.51	The evolution of the “word cloud” of forest carbon sequestration literature, 1990–2018	161
Figure 5.52	Tree map of the keywords plus	163
Figure 5.53	Tree map of the author’s keyword	163
Figure 5.54	Tree map of the article’s title (Bigrams)	164
Figure 5.55	Tree map of the abstract (Trigrams)	164
Figure 5.56	Word dynamics representing the annual count for the top-10 repeated indexed keywords in the AIS research domain	165
Figure 5.57	Word dynamics representing the annual count for the top-10 repeated indexed keywords in the AIS research domain (regenerated using Microsoft Excel)	166
Figure 5.58	Word dynamics representing the cumulative count of the top-10 repeated indexed keywords in AIS research domain	166
Figure 5.59	Cumulative number of occurrences of the top-10 indexed keywords until 2021 in AIS studies	167
Figure 5.60	Trend topics of the author’s keywords from 2000-2021 in AIS research	168
Figure 6.1	Clustering by coupling analysis	172
Figure 6.2	Cluster by documents coupling map	173
Figure 6.3	Network of clusters by documents coupling	174
Figure 6.4	Data of clusters by documents coupling	176
Figure 6.5	Clusters of documents coupling	177
Figure 6.6	Clusters by authors coupling	178
Figure 6.7	Clusters by sources coupling	179
Figure 6.8	Co-occurrence network map of keywords plus	183
Figure 6.9	Details of the co-occurrence of keywords plus	185
Figure 6.10	Degrees node of the co-occurrences of keywords plus	185
Figure 6.11	Quadrant in a thematic map (also called as strategic diagram)	187
Figure 6.12	Thematic map	188
Figure 6.13	Thematic map (with three N. of labels)	189
Figure 6.14	Thematic network	191
Figure 6.15	Thematic network (zoom-in view of the red theme)	191
Figure 6.16	Thematic map data	192
Figure 6.17	Cluster table	193

Figure 6.18	Thematic evolution concept (Cobo et al., 2011b)	194
Figure 6.19	Thematic evolution map of Keywords Plus	196
Figure 6.20	Thematic map for the period from 1975 to 2000	199
Figure 6.21	Thematic map for the period from 2001 to 2010	199
Figure 6.22	Thematic map for the period from 2011 to 2021	200
Figure 6.23.	Conceptual structure map with Multiple Correspondence Analysis of Keywords Plus	202
Figure 6.24	Topic dendrogram	204
Figure 6.25	Factorial map of the documents with the highest contributes	205
Figure 6.26	Factorial map of the most cited documents	206
Figure 6.27	Words by cluster	207
Figure 6.28	Most contributing papers from Articles by Cluster's table	207
Figure 6.29	Most cited papers from Articles by Cluster's table	208
Figure 6.30	Co-citation network by papers	211
Figure 6.31	Co-citation network by authors	212
Figure 6.32	Co-citation network by sources	213
Figure 6.33	Papers with highest page rank	214
Figure 6.34	Degree Plot	214
Figure 6.35	Historical direct citation network of top-cited papers in AIS from 1975-2021	215
Figure 6.36	Historical direct citation table	216
Figure 6.37	Collaboration network by authors	219
Figure 6.38	Collaboration network by institutions	220
Figure 6.39	Collaboration network by countries	221
Figure 6.40	Collaboration table by countries	221
Figure 6.41	Degree plot for collaboration by countries	222
Figure 6.42	Country collaboration map	223



List of Tables

Table 1.1	Sample of Research Aims and Objectives	11
Table 1.2	Research Questions	12
Table 2.1	List of words and their synonyms	19
Table 2.2	Boolean operators	22
Table 2.3	Proximity rules	24
Table 2.4	Document results for each keyword	26
Table 2.5	Export file formats supported by Biblioshiny	32
Table 3.1	Field tags for data imported from the scopus.bib file	58
Table 3.2	Components of an object class in Bibliometrix R Package.	59
Table 3.3	Common errors in running Biblioshiny	66
Table 4.1	Increase OpenRefine memory	86
Table 5.1	Publication trends and impacts	102
Table 5.2	Most relevant sources	108
Table 5.3	Top 20 most of the highest influence sources based on total citations	114
Table 5.4	Top 10 most productive authors	121
Table 5.5	Productivity patterns of authors and research contributions	125
Table 5.6	Top 20 most influence authors	126
Table 5.7	Corresponding author's country	130
Table 5.8	Country specific production	132
Table 5.9	Most cited countries	135
Table 5.10	Most global cited documents (as per Biblioshiny)	138
Table 5.11	Most global cited documents (alternative presentation)	139
Table 5.12	Most local cited documents	143
Table 5.13	Most local cited documents (alternative presentation)	145
Table 6.1	Definition of terms	184
Table 6.2	Thematic evolution table	197