

1 **BibeR: A Web-based tool for bibliometric analysis in scientific**
2 **literature**

3 Yang Liu^{1,2,3*}, Meng Li¹

4 ¹ Institute for Advanced Study, Shenzhen University, Shenzhen 518060, China

5 ² Key Laboratory of Optoelectronic Devices and Systems of Ministry of Education and Guangdong Province,
6 College of Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, China

7 ³ EcoLab, Université de Toulouse, CNRS, INPT, UPS, Toulouse, France

8

9 Corresponding author: yang.liu.shsf@gmail.com

10

11 **Abstract**

12 Bibliometric analysis is a statistical method to summarize the amount of scientific activity in a domain. Insights
13 can be derived from bibliometrics to understand the development trend of the research domain. R is an open
14 sourced programming language specialized for statistical computing and graphic visualization. To benefit from
15 the convenience of R and the outcome of bibliometric analysis, we here introduce BibeR, a web-based
16 application for the visualization of bibliometric analysis. An example of bibliometric analysis on the articles
17 published in journal *Scientometrics* is used to illustrate the usage of BibeR. The development of BibeR is still in
18 progress, and future possibly improvements on BibeR are discussed.

19 **Keywords:**

20 Bibliometrics; scientometrics; R; shiny; packages; BibeR; web app; workflow

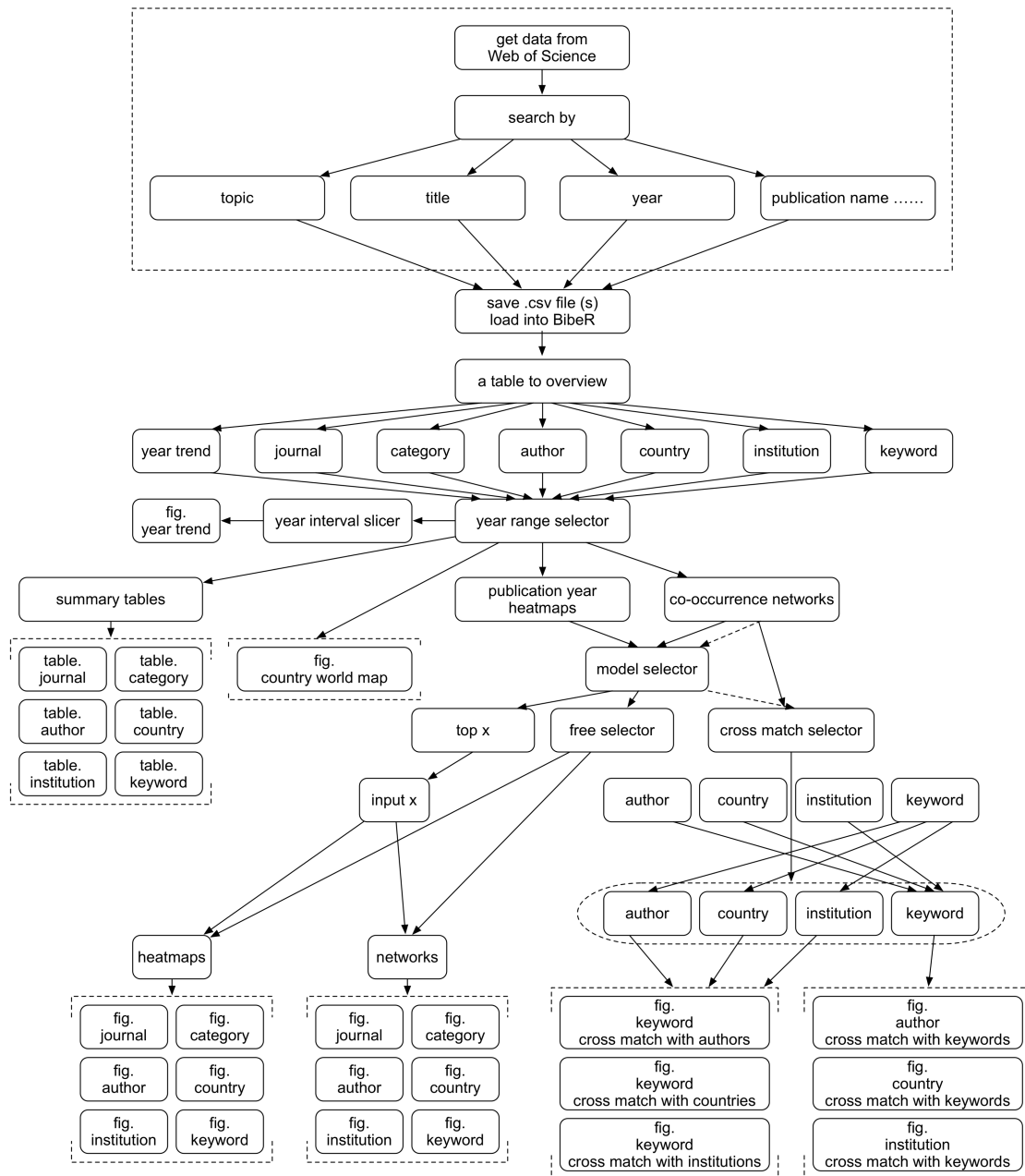
21

22 Introduction

23 Bibliometrics is the statistical analysis of bibliographic data including books, articles or other
24 publications mainly in the scientific and technical literatures. It measures the amount of scientific activity in a
25 subject category, journal, country or other area of interest (Henry Small, Thomson Reuters). Individual
26 researchers could use bibliometric methods to promote their research in a certain scientific field or domain.
27 Bibliometric methods help researchers to understand the developments in the field within the scope of the
28 worldwide research community by answering some typical questions e.g. (1) Does the numbers of publications
29 in your field increase recently? (2) Who published the most in your field? (3) What are the top journals in your
30 field? There are superior software already exist can answer the above questions, e.g. Histcite (www.histcite.com)
31 and Citespace II (Chen, 2006). However, obstacles still lay in front of users e.g. who are unable or have no
32 intend to install Java dependencies or whose computers have no Windows installed. Besides, the comprehensive
33 functions and complicated parameters setting could confuse some of the beginning users, and make it less user-
34 friendly. For users who have no experience in bibliometrics and are interested in having a glance at the
35 bibliometric analysis, a light-weighted, less sophisticated platform with rich and extendable functions is strongly
36 needed.

37 R is an open sourced programming language for statistical computing and graphics visualization
38 supported by the R Foundation for Statistical Computing. R community provides extensive packages written by
39 both statisticians and various developers (e.g. web, front-end). Package “shiny” is a web application framework
40 for R. It could turn analysis performed by R script into interactive web applications (e.g. Ziegler, Hartsock &
41 Baxter, 2015). Advantages can be taken from R and “shiny” package to make contribution to bibliometric
42 community. Recently, Guler et al. (2016) discussed about possible scientific workflows for bibliometrics. They
43 pointed out that the Taverna Workbench software is competent in this field by integrating the multi-step analyses
44 including automated data import via Web, data extraction from XML and statistical analysis and visualization
45 with R. Interestingly, it coincided with the design concept of BibeR which also aims to provide a one-stop
46 service for bibliometric analysis. Until now, except for data source highly relying on the most commonly used
47 scientific database Web of Science (Thompson Reuters), BibeR can fully capable for data extraction, statistical
48 analysis and visualization, and have high potential to extend its functionalities with the help of R community
49 (Guler, Waaijer & Palmblad, 2016).

50 BibeR, a web-based application, is designed to make it easier for individual researchers to perform
51 bibliometric analysis and visualization interactively, characterized with a friendly and easy user-interface, cross-
52 platform, free of installation, open sourced and flexible functions with the supports benefited from R developer
53 community. Before the born, bibliometric researches have been done by using the prototype of BibeR (Guo et al.,
54 2015; Nabil Majdi, 2015; Wang et al., 2015). In this paper, we proudly introduce BibeR to readers, and hope to
55 draw attentions from both of users and developers worldly. Fig. 1 is the flow diagram to show how BibeR works
56 and what BibeR can provide. We will start with an exemplary analysis, which is to analyze the research trend of
57 all articles published in journal *Scientometrics*.



58

59 **Fig 1** A diagram to show how BibeR works and what BibeR can provide. The arrows indicate the workflow directions.
 60 The closed dotted square box includes the steps should be done before diving into BibeR. The vertical dotted brackets
 61 include the outputs of bibliometric analysis e.g. figures and tables. The squashed dotted rectangle covers the
 62 parameters for cross-match selector.

63 Exemplary analysis

64 The data source file could be download from Web of Science as the following ways: (1) select “Web of
 65 Science™ Core Collection” and enter your search query (in this case: select “Publication name” and input
 66 “scientometrics”) and click search; (2) choose and click “Save to Other File Formats”; (3) select the number of

67 records, e.g. all records on page or records from 1 to 500; (4) in “Record Content” dropdown menu, select and
68 click “Full Record and Cited References”; (5) in “File Format” dropdown menu, select and click “Plain Text”.
69 Several tips should be noted here, (1) making your search in “Web of Science™ Core Collection” database is
70 mandatory because full records could be saved only in this database; (2) the limit to save maximum 500 records
71 is set by Web of Science, for one who need more than 500 records, please repeat the above steps and remember
72 to select the range continuously (repeated 8 times in this case).

73 One or several .txt file (s) should be available on your computer for further steps. The uploaded data file
74 is required to be a .txt file or a .zip file compressed with several .txt files. When upload complete, a brief table
75 with selected default columns is shown for a general review (Fig. 2). You can download the full table as a .csv
76 file. You can also upload the .csv file again for analysis. Hereafter, you can enjoy all the figures and tables
77 BibeR provides to you. And, you can filter the data source by selecting the range of publication years or select
78 the ones as you type. The figures/tables would be shown simultaneously and interactively, which is the beauty of
79 BibeR.

BibeR Home Load file Basic results Advanced results About

Show 10 entries Search:

Table: selected columns of the input .txt file

year	domain	author	title	journal	country	link
2014	COMPUTER SCIENCE; INFORMATION SCIENCE & LIBRARY SCIENCE	SONG, M;HEO, GE;KIM, SY	ANALYZING TOPIC EVOLUTION IN BIOINFORMATICS: INVESTIGATION OF DYNAMICS OF THE FIELD WITH CONFERENCE DATA IN DBLP	SCIENTOMETRICS	SOUTH KOREA	Link
2014	COMPUTER SCIENCE; INFORMATION SCIENCE & LIBRARY SCIENCE	ZHU, J	MEASURING RECENT RESEARCH PERFORMANCE FOR CHINESE UNIVERSITIES USING BIBLIOMETRIC METHODS	SCIENTOMETRICS	PEOPLES R CHINA;PAKISTAN;SAUDI ARABIA	Link
2014	COMPUTER SCIENCE; INFORMATION SCIENCE & LIBRARY SCIENCE	MARAUT, S;MARTINEZ, C	IDENTIFYING AUTHOR-INVENTORS FROM SPAIN: METHODS AND A FIRST INSIGHT INTO RESULTS	SCIENTOMETRICS	SPAIN	Link
2014	COMPUTER SCIENCE; INFORMATION SCIENCE & LIBRARY SCIENCE	PEZZONI, M;LISSONI, F;TARASCONI, G	HOW TO KILL INVENTORS: TESTING THE MASSACRATORA (C) ALGORITHM FOR INVENTOR DISAMBIGUATION	SCIENTOMETRICS	SWITZERLAND;ITALY;FRANCE	Link
2014	COMPUTER SCIENCE; INFORMATION SCIENCE & LIBRARY SCIENCE	KAZAKIS, NA	BIBLIOMETRIC EVALUATION OF THE RESEARCH PERFORMANCE OF THE GREEK CIVIL ENGINEERING DEPARTMENTS IN NATIONAL AND EUROPEAN CONTEXT	SCIENTOMETRICS	GREECE	Link
2014	COMPUTER SCIENCE; INFORMATION SCIENCE & LIBRARY SCIENCE	SCHOEN, A;HEINISCH, D;BUENSTORF, G	PLAYING THE 'NAME GAME' TO IDENTIFY ACADEMIC PATENTS IN GERMANY	SCIENTOMETRICS	GERMANY	Link
2014	COMPUTER SCIENCE; INFORMATION SCIENCE & LIBRARY SCIENCE	BOGOCZ, J;BAK, A;POLANSKI, J	NO FREE LUNCHES IN NATURE? AN ANALYSIS OF THE REGIONAL DISTRIBUTION OF THE AFFILIATIONS OF NATURE PUBLICATIONS	SCIENTOMETRICS	POLAND	Link
2014	COMPUTER SCIENCE; INFORMATION SCIENCE & LIBRARY SCIENCE	ABRIZAH, A;ERFANMANESH, M;ROHANI, VA;THELWALL, M;LEVITT, JM;DIDEGAH, F	SIXTY-FOUR YEARS OF INFORMETRICS RESEARCH: PRODUCTIVITY, IMPACT AND COLLABORATION	SCIENTOMETRICS	MALAYSIA;IRAN;United Kingdom	Link
2014	COMPUTER SCIENCE; INFORMATION SCIENCE & LIBRARY SCIENCE	KIM, J;DIESNER, J	A NETWORK-BASED APPROACH TO COAUTHORSHIP CREDIT ALLOCATION	SCIENTOMETRICS	USA	Link
2014	COMPUTER SCIENCE; INFORMATION SCIENCE & LIBRARY SCIENCE	IVANOVIC, D;HO, YS	INDEPENDENT PUBLICATIONS FROM SERBIA IN THE SCIENCE CITATION INDEX EXPANDED: A BIBLIOMETRIC ANALYSIS	SCIENTOMETRICS	SERBIA;TAIWAN	Link

Showing 1 to 10 of 3,652 entries

Previous 1 2 3 4 5 ... 366 Next

Upload file here:

Choose File Upload complete

Here accept single .txt or compressed .zip file saved from Web of Science.

Since it may take long time for web server to parse your raw .txt files, it is encouraged to save the parsed file as .csv format. Please click the button below.

Then, upload the .csv file, it just takes milli seconds to start the analysis!

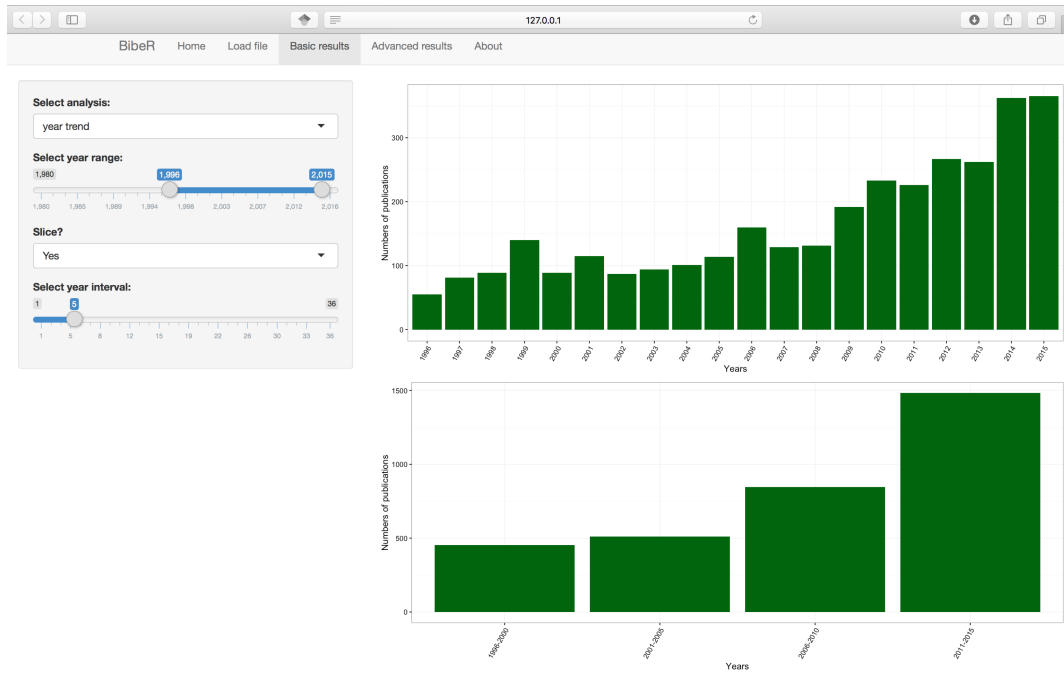
Save as:

csv

80

81 **Fig 2** The screenshot of BibeR interface under the tab “Load file”. A table of all records parsed from users’ uploaded
 82 files is shown, including the columns “year”, “domain”, “author”, “title”, “journal”, “country” and the “link” to find
 83 the record on Google Scholar.

84 Hereby, we present the selected screenshots of interfaces of BibeR. When finishing data upload, user
85 can click “Basic results” tab, and select “year trend” under “Select analysis” dropdown menu, fix the year range
86 to “1996 – 2015” or any other possible options. If user needs to slice the year interval, just select “Yes” under
87 “Slice?” dropdown menu, then select the interval as 5 (Fig. 3). It is seen that the number of publications in 1999
88 stands out of the period from late 1990s to mid 2000s. The sum of publications in the period 2011 – 2015 is
89 almost twice of that in the period 2006 – 2010. Besides, user can acquire a table with summary of authors (Fig.
90 4). The table gives the information on (1) author names (last name with the initial of first name); (2) frequencies
91 of each author; (3) ranks by the frequency and percentages of total number of publications; (4) numbers and
92 ranks of author when as first author; (5) numbers and ranks of author when as corresponding author. Fig. 4
93 shows that Glanzel W. published the most articles not only in total but also as both first author and
94 corresponding author. Under the tab “Basic results”, user can also get a world map in a selected year range. As
95 shown in Fig. 5, the colors indicate the numbers of articles published by countries. It illustrates that USA is the
96 most productive country on the topic of bibliometrics, followed by China, Spain, Belgium and Netherlands that
97 one can distinguish with naked eyes. If user need the detailed numbers of country productivities, just select
98 “country table” would get answered. Furthermore, for heatmaps, user needs to click “Advanced results” tab. Fig.
99 6 gives the top 25 most productive institutions in the period 1996 – 2015. Each cell represents the rate to the total
100 frequency of each institution at certain year, and rows are ranked by total frequencies of institutions decreasingly.
101 Katholieke Universiteit Leuven in Belgium published the most articles, however, its productivity spreads out of
102 the two decades. For both Wuhan University and Chinese Academy of Science in China, their contributions to
103 bibliometric literature are increasing quickly during recent years. Last but not the least, the co-occurrence
104 network analysis is also an important feature of BibeR. Co-occurrence analysis is an established bibliometric
105 method to explore the network of terms based on their paired presence (co-occurrence) within a specified content
106 of document in scientific discipline (Callon, Courtial & Laville, 1991; Cheng et al., 2014). BibeR can provide
107 co-author, co-keyword, co-country and co-institution analyses. User can select three modes to present the co-
108 occurrence networks. The first two modes are (1) to select top number of items and (2) select freely within the
109 items no matter their frequencies, which is as same as for heatmap visualization. The third mode is “cross match”
110 that allow user to select freely in other criteria. Taking Fig. 7 as an example, it shows the relationships among the
111 authors who published articles with keywords “Bioinformatics”, “Topic evolution”, and “Medical subject
112 headings (MESH)”. In the example, “keyword” is the cross-matching criterion for the author network. Other
113 directions of cross-matching are also available (Fig. 1).



114

115 **Fig 3** The screenshot of BibeR interface for “year trend” analysis under the tab “Basic results”. The analysis “year
 116 trend” is selected, the year range is set to 1996 – 2015, and the year interval is set to 5. Two figures are presented with
 117 x-axis “Years” and y-axis “Numbers of publications”.

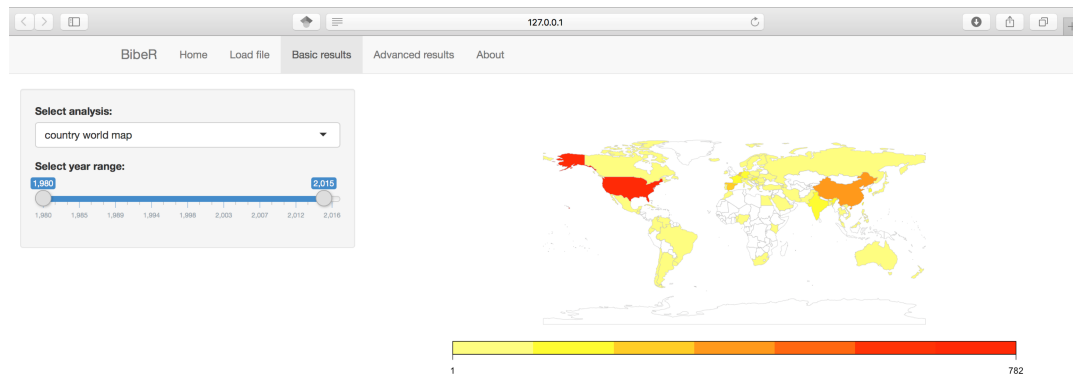
The table shows the following data:

author	Freq	TA R (%)	FA (FR)	RP (RP R)
GLANZEL, W	98	1 (2.7)	54 (1)	63 (1)
SCHUBERT, A	72	2 (2)	43 (3)	44 (2)
ROUSSEAU, R	56	3 (1.6)	14 (17)	24 (9)
LEYDESDORFF, L	49	4 (1.4)	34 (4)	34 (4)
EGGHE, L	48	5 (1.3)	48 (2)	44 (2)
BRAUN, T	38	6 (1.1)	28 (5)	23 (10)
BORNMANN, L	34	7 (0.9)	24 (7)	25 (7)
THELWALL, M	30	8 (0.8)	7 (46)	10 (30)
GUPTA, BM	29	9 (0.8)	18 (9)	15 (19)
HUANG, MH	29	9 (0.8)	10 (28)	13 (23)

118

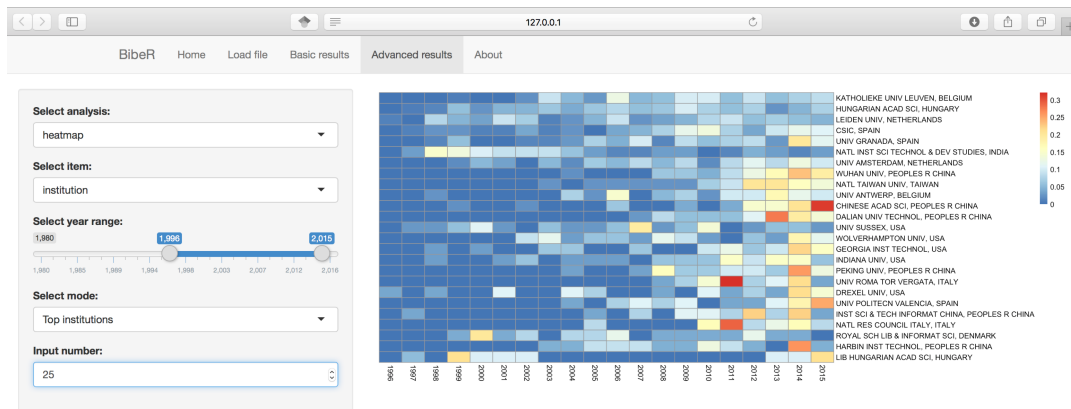
119 **Fig 4** The screenshot of BibeR interface for “author table” analysis under the tab “Basic results”. The analysis
 120 “author table” is selected, the year range is set to 1980 – 2015. A table is shown with five columns (“author”: authors’
 121 names; “Freq”: numbers of occurrence; “TA R(%)”: ranks of the occurrence and the percentages to total number of

122 publications; “FA (FR)” : numbers of occurrence as first author and their ranks; “FA (FR)” : numbers of occurrence
123 as corresponding author and the ranks).



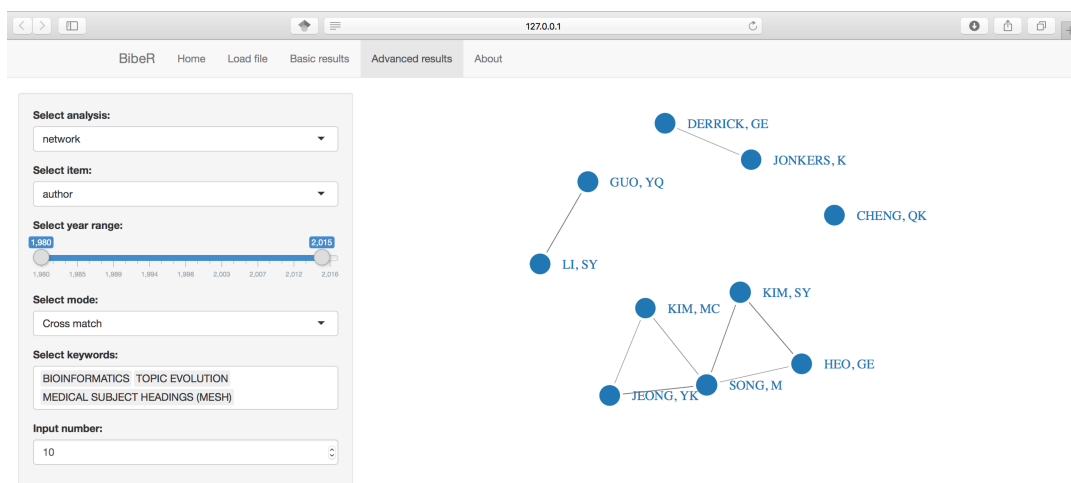
124

125 **Fig 5** The screenshot of BibeR interface for “country world map” analysis under the tab “Basic results”. The analysis
126 “country world map” is selected, the year range is set to 1980 – 2015. A world map is presenting the numbers of
127 publications contributed from each country. The legend in colors from yellow to red indicates the number scale.



128

129 **Fig 6** The screenshot of BibeR interface for “heatmap” analysis under the tab “Advanced results”. The analysis
 130 “heatmap” for the item “institution” is selected, the year range is set to 1996 – 2015. Analysis mode is set to “Top 25
 131 institutions”. A heatmap is shown. For example, the cell of “CHINESE ACAD SCI, PEOPLES R CHINA” on 2015 is
 132 roughly 0.3 which means 30 % of the total publication by this institution were achieved in the year 2015.



133

134 **Fig 7** The screenshot of BibeR interface for “network” analysis under the tab “Advanced results”. The analysis
 135 “network” for the item “author” is selected, the year range is set to 1980 – 2015. Analysis mode is set to “Cross match”
 136 with “keywords”. Three keywords “BIOINFORMATICS”, “TOPIC EVOLUTION” and “MEDICAL SUBJECT

137 HEADINGS (MESH)” are chosen from the library. A figure of co-occurrence network is shown for the top 10 authors
138 whose publications include these keywords.

139 Discussion

140 BibeR would have not been possible without these amazing R packages: “countrycode”, “dplyr”, “DT”,
141 “ggplot2”, “gsubfn”, “networkD3”, “pheatmap”, “rworldmap”, “shiny”, “stringr” and “tm” (Wickham, 2009;
142 South, 2011; Arel-Bundock, 2014; Grothendieck, 2014; Feinerer & Hornik, 2015; Gandrud, Allaire & Russell,
143 2015; Kolde, 2015; Wickham, 2015; Wickham & Francois, 2015; Xie, 2015; Chang et al., 2016). BibeR aims to
144 provide an entire non-stop pipeline for bibliometric analysis, in which users have no need to transfer data files in
145 different formats between various software. Thus, the supports for more functionalities are expected. The
146 development of BibeR is still in progress. However, to the best of my knowledge, BibeR is the first web platform
147 written in R language to perform bibliometric analysis interactively. Its extensibility allows to get the direct
148 support of a number of R packages for advanced analysis, and also allows experienced R users to custom their
149 own analysis and build their own BibeR since which is open-sourced. For instance, as stated by Guler et al.
150 (2016) recently, “bibtex” and “CITAN” packages are useful for BibeR to support more data sources/formats
151 besides Web of Science (Gagolewski, 2011; Francois, 2014). Advanced text mining technology to map and
152 cluster the terms extracted from texts is the built-in functionality for software e.g. VOSviewer (van Eck &
153 Waltman, 2011). Similar functions can be implemented into BibeR with the aid of the “openNLP” R package
154 (Hornik, 2016). Besides, statistical modeling is one the advantages of R, thus, to apply machine learning
155 algorithms to bibliometric analysis is also possible be realized by BibeR. Recently, Ebadi and Schiffauerova
156 (2016) developed Java program and performed a semi-automatic machine learning system to assess the impact of
157 research funding on the quantity and quality of scientific output in a specific field. Co-citation network analysis
158 can also be integrated into BibeR since the information of cited references are available for data manipulation.
159 Besides, an reproducible analysis report including graphical, tabular and textual outputs can be generated
160 automatically by R packages “ReporteRs” and “knitr” (Gohel, 2016; Xie, 2016). Microsoft and html-like
161 documents could be downloaded from BibeR for further purposes e.g. presentation and publication.

162 The future development of BibeR can not only focus on the application of emerging R packages
163 including both statistical methods and visualization options, but also on the improvement of user interface. The
164 original idea of BibeR is indeed to reduce users’ entrance level to perform bibliometric analysis as simple as
165 possible, e.g. the parameter settings should be limited. Thus, the developers (e.g. the authors of this paper) are
166 suggested to set up most commonly used parameters as default. An on-off switch could be considered as a
167 compromise, i.e. to permit experienced users to fully customize their analyses.

168 A demo of BibeR is now available on the free server provided by RStudio Inc.
169 (<https://yangliufr.shinyapps.io/BibeR/>), which has limited network traffic and computing ability. Any servers
170 with R (server version) and related packages (e.g. “shiny” server version) installed can host BibeR for internet or
171 intranet usages. Personal computer with local web server installed can also host BibeR for local use. Details for
172 server deployment could be find on the website (<https://www.rstudio.com/products/shiny/shiny-server2/>). BibeR
173 is open-sourced, under the MIT License. The source codes of BibeR can be found in the achieve
174 (<https://github.com/yangliu-shsf/BibeR/>).

175 **Acknowledgements**

176 This work was funded by National Natural Science Foundation of China (No. 41506163) and the Shenzhen Peacock Plan (No.
177 KQCX2015032416053646). Yang Liu received a PhD grant by China Scholarship Council (CSC, grant No. 201208320230)
178 during his three years studying in France. Special thanks to Dr. Peng Zhao (University of Innsbruck, Austria) for his
179 encouragement to publish this work and valuable suggestions. The name “BibeR” is given by him, which means “beaver” in
180 German. Also thank Dr. Jie Pan (Shenzhen University) for his heavily use of BibeR and great comments as users’ point of
181 view.
182

183 **References**

- 184 Arel-Bundock V 2014.countrycode: Convert Country Names and Country Codes. Available at [https://CRAN.R-](https://CRAN.R-project.org/package=countrycode)
185 [project.org/package=countrycode](https://CRAN.R-project.org/package=countrycode) (accessed March 16, 2016).
- 186 Callon M, Courtial JP, Laville F 1991. Co-word analysis as a tool for describing the network of interactions
187 between basic and technological research: The case of polymer chemistry. *Scientometrics* 22:155–205.
- 188 Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J 2016. Web Application Framework for R [R package shiny
189 version 0.13.1].
- 190 Chen CM 2006. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific
191 literature. *Journal of the American Society for Information Science and Technology* 57:359–377.
- 192 Cheng B, Wang M, Mørch AI, Chen N-S, Kinshuk, Spector JM 2014. Research on e-learning in the workplace
193 2000–2012: A bibliometric analysis of the literature. *Educational Research Review* 11:56–72.
- 194 Ebadi A, Schiffauerova A 2016. How to boost scientific production? A statistical analysis of research funding
195 and other influencing factors. *Scientometrics* 106:1093–1116.
- 196 Feinerer I, Hornik K 2015. Text Mining Package [R package tm version 0.6-2].
- 197 Francois R 2014.bibtex: bibtex parser. Available at <https://CRAN.R-project.org/package=bibtex> (accessed March
198 16, 2016).
- 199 Gagolewski M 2011. Bibliometric impact assessment with R and the CITAN package. *Journal of Informetrics*
200 5:678–692.
- 201 Gandrud C, Allaire JJ, Russell K 2015. D3 JavaScript Network Graphs from R [R package networkD3 version
202 0.2.8].
- 203 Gohel D 2016. Microsoft Word, Microsoft PowerPoint and HTML Documents Generation [R package
204 ReporteRs version 0.8.6].
- 205 Grothendieck G 2014.gsubfn: Utilities for strings and function arguments. Available at [https://CRAN.R-](https://CRAN.R-project.org/package=gsubfn)
206 [project.org/package=gsubfn](https://CRAN.R-project.org/package=gsubfn) (accessed March 16, 2016).
- 207 Guler AT, Waaijer CJF, Palmblad M 2016. Scientific workflows for bibliometrics. *Scientometrics*:1–14.
- 208 Guo C, Park Y-S, Liu Y, Lek S 2015. Toward a new generation of ecological modelling techniques. In:
209 *Advanced Modelling Techniques Studying Global Changes in Environmental Sciences*. Developments in
210 Environmental Modelling. Elsevier, 11–44.
- 211 Hornik K 2016. Apache OpenNLP Tools Interface [R package openNLP version 0.2-6].
- 212 Kolde R 2015.pheatmap: Pretty Heatmaps. Available at <https://CRAN.R-project.org/package=pheatmap>
213 (accessed March 16, 2016).
- 214 Nabil Majdi WT 2015. Free-Living Nematodes in the Freshwater Food Web: A Review. *Journal of Nematology*
215 47:28.
- 216 South A 2011.rworldmap: A New R package for Mapping Global Data. Available at [http://journal.r-](http://journal.r-project.org/archive/2011-1/RJournal_2011-1_South.pdf)
217 [project.org/archive/2011-1/RJournal_2011-1_South.pdf](http://journal.r-project.org/archive/2011-1/RJournal_2011-1_South.pdf) (accessed March 16, 2016).
- 218 van Eck NJ, Waltman L 2011. Text mining and visualization using VOSviewer.
- 219 Wang C, Liu Y, Li X, Lai Z, Tackx M, Lek S 2015. A bibliometric analysis of scientific trends in phytoplankton
220 research. *Annales de Limnologie - International Journal of Limnology* 51:249–259.
- 221 Wickham H 2009. *ggplot2: Elegant Graphics for Data Analysis*.
- 222 Wickham H 2015. Simple, Consistent Wrappers for Common String Operations [R package stringr version
223 1.0.0].
- 224 Wickham H, Francois R 2015. A Grammar of Data Manipulation [R package dplyr version 0.4.3].
- 225 Xie Y 2015. A Wrapper of the JavaScript Library “DataTables” [R package DT version 0.1].
- 226 Xie Y 2016.knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.12.3
227 (accessed March 17, 2016).
- 228 Ziegler GR, Hartsock RH, Baxter I 2015. Zbrowse: an interactive GWAS results browser. *PeerJ Computer*
229 *Science* 1:e3.
- 230