

1 **Bibliometric and sentiment analysis with machine learning on the scientific contribution of**
2 **Professor Srinivasa Sourirajan**

3

4 Mohamed Khayet ^{1,2,*}, Ersin Aytaç ^{1,3}, Takeshi Matsuura⁴

5

6 ¹Department of Structure of Matter, Thermal Physics and Electronics, Faculty of Physics,
7 University Complutense of Madrid, Avda. Complutense s/n, 28040, Madrid, Spain.

8 ²Madrid Institute for Advanced Studies of Water (IMDEA Water Institute), Calle Punto Net N° 4,
9 28805, Alcalá de Henares, Madrid, Spain.

10 ³ Zonguldak Bülent Ecevit University, Department of Environmental Engineering, 67100,
11 Zonguldak, Turkey

12 ⁴Department of Chemical and Biological Engineering, University of Ottawa, Ottawa, ON K1N
13 6N5, Canada

14

15 * Corresponding author:

16 khayetm@fis.ucm.es

17 Tel. +34-91-3945185

18

19

20

21

22

23 **Abstract**

24 P Prof. Srinivasa Sourirajan is remembered by the desalination and membrane community as the
25 “*Father of Reverse Osmosis*”. He passed away at the age of 98 peacefully in his beloved city
26 Ottawa (Canada). His legacy will be remembered by the scientific community “*membrane science,*
27 *membrane processes, desalination and engineering*”. His research studies were not only novel, but
28 also very creative and even visionary. He offered a priceless gift to humanity by bringing clean
29 water to all those in need through the presentation of reverse osmosis technology together with its
30 appropriate membranes for water treatment, including desalination. This technology has now
31 gained worldwide interest as it is able to produce clean water at a lower cost compared to other
32 separation processes. His scientific contribution also pioneered other research areas. He developed
33 novel research methodologies in geophysics while in catalysis he produced unleaded gasoline to
34 help with the smog issue. He was nominated for the Nobel Prize three times. Prof. Sourirajan had
35 also an exceptional humanitarian attribute. He played a significant role in bringing the Indian
36 community to Ottawa. In the present paper we apply machine learning for his extraordinary and
37 original scientific contribution. The results reveal how influential scientist he was.

38

39

40 **Keywords:** Biblioshiny; Exploratory Tool; text mining; VADER; word cloud.

41

42

43 **Highlights**

44 Prof. Srinivasa Sourirajan is known as the “*Father of Reverse Osmosis*”.

45 Prof. Sourirajan has been nominated three times for the Nobel Prize.

46 His research studies were not only novel, but also very creative and even visionary.

47 Prof. Sourirajan has filled 17 patents about reverse osmosis membranes and applications.

48 Machine learning results reveal how influential scientist he was.

49

50 **1. Introduction**

51 Prof. Srinivasa Sourirajan is known within the desalination and membrane community as the
52 pioneer of the membrane separation process, reverse osmosis (RO), used worldwide for
53 desalination of seawater and brackish water. This commenced when he announced the first
54 cellulose acetate membrane proposed for seawater desalination in 1960 at the University of
55 California (Los Angeles, USA), UCLA, together with Dr. Sidney Loeb [1]. Since then, a number
56 of membrane separation technologies have emerged such as nanofiltration (NF), ultrafiltration
57 (UF), microfiltration (MF), membrane gas and vapor separation and pervaporation (PV), etc.
58 These membrane technologies are nowadays applied in different industrial sectors such as
59 “pharmaceutical, medical, petrochemical, food processing, water, energy, etc.”.

60 Born in 1923 in a small rural village in southern India, Prof. Sourirajan graduated from
61 Annamali University in 1943, and received his Ph.D. in Chemistry from Bombay University in
62 1953. He was among the first generation of young Indians who received a Ph.D. in chemistry in
63 post-colonial India. His thesis research caught the attention of Prof. Paul Emmett of Johns Hopkins
64 University (USA), a world leader in the field of catalysis. This brought him to the United States in
65 the mid 1950's to Yale University (New Haven), where he studied Chemical Engineering for two
66 years and received an additional doctorate in Chemical Engineering. He used to work on
67 adsorption at high-pressure, and since RO is also a high-pressure operation, he moved to the
68 University of California Los Angeles campus (UCLA) in 1956 to conduct research on RO. He
69 used to say the fundamentals are not different [2]. He joined UCLA as a research scientist and
70 made seminal contributions in three key research areas: in geophysics he pioneered new research
71 methodologies and techniques; in catalysis his work led to the development of unleaded gas to
72 fight the growing smog problem; and in desalination and membrane science, perhaps the most

73 notably, together with Sidney Loeb he invented modern RO using a cellulose acetate polymer as
74 mentioned above. In 1961 he moved to Ottawa (Canada) to join the National Research Council
75 (NRC) where he continued to be an active researcher in the area of membrane science and related
76 technologies until he retired in 1986. In January 1987, he joined the Department of Chemical
77 Engineering of the University of Ottawa and founded the Industrial Membrane Research Institute
78 (IMRI). During the period October 1991 to June 1995, he served as a Visiting Professor to establish
79 the membrane research laboratories in the Chemical Engineering Department of the National
80 University of Singapore (NUS). Many researchers joined the two well recognized membrane
81 research laboratories (IMRI, NUS) from all around the world and later contributed to the
82 establishment of their membrane related industries, research laboratories/institutes or educational
83 centers. Some of the examples are; M. Khayet established his research group “*Membranes and*
84 *Renewable Energies*” at the University Complutense of Madrid (UCM), Spain; M.R. Qtaishat,
85 Arab Open University, Jordan; B. Kurczek, University of Ottawa, Canada; X. Feng, University of
86 Waterloo, Canada; J. Barzin, Iran Polymer and Petrochemical Institute, Iran; M.P. Chenar;
87 Ferdowsi University of Mashhad, Iran, M. Rafat, Linköping University, Sweden; S. Noh, Yonsei
88 University, Korea; J.W. Rhim, Hannam University, Korea; M. Ishiguro, Hokkaido University,
89 Japan; A. Hamza, Imtex, Canada, F. Baig, Petro Sep, Canada, M. Tabe, Ontario Ministry of
90 Environment, Canada: S. Mortazavi, Natural Resources Canada, Canada.

91 During his brilliant scientific career, Dr. Sourirajan was invited by many institutions to deliver
92 lectures. Some of them are as follows: Plenary lecture at the Symposium on the Chemistry for the
93 Welfare of Mankind at the 26th IUPAC Congress, Tokyo, 1977; The R.S. Jane Memorial Lecture
94 Award of the Canadian Society for Chemical Engineering (ChChE) at CSChE Conference, 1981;
95 A series of lectures in China at the Institute of Environmental Chemistry and the Chinese Academy

96 of Sciences, 1983; A series of lectures in India at the invitation of the Council of Scientific and
97 Industrial Research, New Delhi, under an UNDP project, 1985-1986; Plenary lecture at the
98 International Congress on Membranes and Membrane Processes, Tokyo, 1987; A series of lectures
99 at the Invitation of the Indian Membrane Society, Baroda, 1991; Plenary Lecture at the ACS-
100 Cellulose 91 Conference, New Orleans, 1991; etc. He also received many honors and awards such
101 as The International Desalination and Environmental Association Research Award, 1979; A
102 Symposium on Synthetic Membranes and Their Applications in his honor at the American
103 Chemical Society (ACS) Meeting, Las Vegas, 1980; A special function held by the Indian
104 Membrane Society to felicitate Dr. Sourirajan for “His Pioneering Work in Membrane Science
105 and Technology”, 1990; A special volume of Desalination (volume 90-numbers 1-3, 1993) “In
106 honor of S. Sourirajan for His Lifelong Contribution to the International Community of Membrane
107 Science and Technology”; Honorary Doctorate Degree awarded by the University of Ottawa,
108 1994, in recognition of his “Immense Contributions to Society through His Pioneering Research
109 in the Areas of Reverse Osmosis and Ultrafiltration”; UDCT Golden Jubilee Visiting Fellow at the
110 Department of Chemical Technology, University of Bombay, 1995; Two special symposia
111 organized by the North American Membrane Society in honor of Dr. Sourirajan for “His
112 Significant Contributions to the Science and Technology of Membranes and His Pioneering Work
113 in the Area of Reverse Osmosis and Gas Separation, 1996”. American Membrane Technology
114 Association, Hall of Fame Award, 2016. Drs. Sourirajan and Loeb were also three-time nominees
115 for the Nobel Prize. Even if they did not win the Nobel Prize, Dr. Menachem Elimelech 's
116 comment about Prof. Sourirajan “*he should have received two Nobel Prizes- one for science and*
117 *one for peace*” affirms the huge impact he has made on our world [3].

118 The current state of water, food, and sanitation owes a lot to Prof. Sourirajan's revolutionary
119 discoveries [3]. In the late 1940s, potable water shortage in some desert areas of the world incited
120 researchers to investigate ways to recover freshwater from saline water. Prof. Sourirajan and his
121 colleagues were young and passionate scientists who also wanted to solve this problem [4]. The
122 discovery of RO technology by S. Sourirajan and S. Loeb in 1960 was the most important
123 contribution to seawater desalination. Their method involved physically pressing a solution against
124 a flat asymmetric membrane comprising a very thin submicron polymeric skin supported by a
125 porous substrate layer. This anisotropic and semipermeable membrane was based on cellulose
126 acetate (CA). Due to its greatly enhanced permeate flux, which was ten times that of other known
127 membrane materials, this membrane gave RO commercial feasibility [5-7]. Since then membrane
128 science and technology research has grown significantly, and numerous new sophisticated
129 materials have been discovered and used in membrane engineering [8]. It was also reported that
130 Loeb and Sourirajan discovered the process of precipitation by immersion in a water bath (i.e.
131 phase inversion), which is one of the most used techniques in membrane formation [9]. The non-
132 solvent induced phase separation (NIPS) technique, which is nowadays the most widely used
133 method in membrane engineering for the preparation of integrally-skinned asymmetric membranes
134 of all varieties, was also given as a gift to membranologists by him in 1962. His work with Prof.
135 Sydney Loeb led to the establishment of the world's first commercial RO desalination plant in
136 Coalinga, California. His RO studies inspired worldwide corporations to create commercial and
137 industrial applications [3]. Since DuPont announced its entry into the RO business in 1967,
138 membrane technologies and membrane production have grown tremendously around the world to
139 become the dominant desalination process. Dean Spatz, founded Osmonics, Inc. in 1969 in
140 Minnetonka, Minnesota with Prof. Sourirajan. His work about modeling of synthetic membranes

141 extended far beyond water desalination, focusing on process development and energy efficiency
142 of liquid and gas separations, including those in industrial, medical, healthcare, and environmental
143 applications. In addition, he built a membrane research laboratory at Singapore's National
144 University [2]. Prof. Sourirajan's vision was, however, well beyond desalination and water
145 treatment using membranes, as evidenced by his two papers published in Nature soon after the
146 development of the RO membrane [10, 11].

147 Although he is best known by the membrane community as the inventor of RO, there are two
148 other areas where Prof. Sourirajan has had a major impact. He developed novel research methods
149 and approaches that produced ground-breaking geophysical data, while his revolutionary
150 breakthrough in catalysis made it possible to produce unleaded gasoline to help with the smog
151 issue. In addition to his scientific contributions, Prof. Sourirajan was a person with an exceptional
152 humanitarian attribute. He played a significant role in establishing Indian community in Ottawa
153 (Canada) [12].

154 Author of numerous patents, books, book chapters, articles, proceeding papers, Prof.
155 Sourirajan passed away in February 2022. He is a scientist who will be remembered with gratitude
156 and appreciation by all researchers in membrane science and related technologies.

157 In this paper we apply machine learning (ML) (text mining and sentiment analysis) approaches
158 and bibliometric methods to analyze the scientific contribution of Prof. Sourirajan. Bibliometrics
159 (Scientometrics) is a statistical tool for assessing and describing research patterns in dedicated
160 areas [13, 14]. The word "bibliometrics" was first mentioned by Otlet in 1934 in his early
161 researches and was first applied by Pritchard in 1969. The development of worldwide citation
162 databases and automatic citation indexing systems, notably at the end of the twentieth century, is
163 correlated to the widespread use of bibliometric analysis. Before electronic databases, researchers

164 had to manually collect publication data in order to use bibliometric techniques. In every branch
165 of science, bibliometric studies on reliable databases can reveal valuable information in the
166 scientific literature with various quantitative and visualization techniques. Co-author analysis,
167 bibliographic coupling, citation analysis, co-citation analysis, word co-occurrence, research
168 achievements of institutions, researchers, and countries are some of the approaches utilized in
169 bibliometric studies. Being used on a variety of research fields, it is becoming more popular day
170 by day [15-19].

171 ML is a way of extracting patterns from big datasets using mathematical functions with
172 multiple parameters that map features to one or more outputs [20-22]. ML models excel at
173 modeling nonlinear systems and exposing underlying complex mechanisms [23]. It is used in a
174 variety of applications such as finding the similarity of catchments [24], predicting of shear
175 capacity of steel channels [25], text classification [26], predicting the pyrolytic kinetics of
176 feedstocks, optimization hydrothermal liquefaction of biomass [27], forecasting hazelnut export
177 quantities [28], etc. Among ML methods, an interesting branch called text mining (TM) is used to
178 examine textual data. TM is a process that employs ML and natural language processing (NLP) to
179 extract high-quality information from text data that is unstructured and vast amounts [29, 30]. TM
180 has many techniques for data cleaning, pre-processing, and processing steps to analyze textual
181 data. Tokenization, lemmatization, feature extraction, topic modelling, sentiment analysis and
182 word cloud are prominent among these techniques [31-34]. Word cloud is a graphic representation
183 of the frequency of words in a written corpus. The larger the term appears in the produced graphic,
184 the more frequently it appears in the article being examined [35]. A word cloud is a quick approach
185 to summarize the corpus, and this type of visualization can help researchers with exploratory
186 textual analysis [29]. A user's good, negative or neutral opinion, emotion, sensation, or thinking is

187 referred to as "sentiment." In general, sentiment analysis seeks to extract these sentiments from
188 textual data [36]. Due to its amazing potential in product evaluative analysis, social opinion
189 analysis, and content-based recommendation, sentiment analysis has become a popular research
190 topic [37].

191 The scientific research patterns of Prof. Sourirajan are revealed in this paper from bibliometric
192 and text mining perspectives. R-Biblioshiny, Exploratory Tool, and Orange Data Mining Tool was
193 used.

194

195 **2. Data and Methods**

196 **2.1. Data**

197 The collection required for the study was downloaded from the Web of Science (WoS)
198 database on 27.03.2022. This database contains 166 different documents types (articles, notes,
199 meeting abstracts, letters, etc.) of Prof. Sourirajan. In our study, analyzes were carried out on only
200 the scientific papers published in refereed journals (152 documents). It has been observed that
201 there are missing data in the data set, especially in the abstract column, which will be used in text
202 mining analysis. Therefore, missing data is filled in manually as much as possible.

203

204 **2.2. Methods**

205 The analyzes and visualization were conducted using R-Biblioshiny, Exploratory Tool, and
206 Orange Data Mining Tool [20, 38, 39]. The analysis conducted in this work provides general
207 information about the collection such as, yearly scientific production and citations, research
208 categories of papers, three fields plot, most relevant sources and their impacts, Bradford's Law,

209 author impact (*h*-index, *g*-index), authors production over time, most globally impactful
210 documents, collaboration network, word cloud, sentiment analysis and timeline, and collaboration
211 of patents. Predictive statistics like these pinpoint certain features of bibliometric data. These
212 analyses include the following calculations and assumptions:

213 Co-authors per document is a metric that shows the average number of co-authors per study
214 and can be calculated by Eq. (1) as follows:

$$215 \text{ Co-Authors per Document} = \frac{\text{Authors Appearances}}{\text{Document}} \quad (1)$$

216 The Collaboration Index is a Co-Authors per article score that is derived solely on the basis
217 of multi-authored articles and calculated by the following equation:

$$218 \text{ Collaboration Index} = \frac{\text{Authors of Multi-Authored Articles}}{\text{Multi-Authored Articles}} \quad (2)$$

219 The number of research papers (*h*) published by a journal (or author), each of which has been
220 cited at least *h* times in other articles, is known as the Hirsch index (*h* – index). The *g* – index is a
221 unique integer indicating that the top *g* articles received at least *g*² citations. The Bibliometrix
222 package calculates these indexes in the collection domain.

223 Bradford's Law illustrates how to work in a subject area, which can be divided into zones
224 (mainly three), with the core zone being the first. These zones represent the level of significance
225 of the sources in the domain [40].

226 Articles Fractionized (*AAU_{jj}*) calculates each author's contribution to a set of published articles
227 based on the idea that all co-authors contribute evenly to each article and can be calculated as
228 follows:

229
$$m = 1 \tag{3}$$

230
$$AAW_{jj} = \sum_{jj=1}^h$$

231 where jj is the article and h is the number of co-authors.

232 Global citations are the total amount of citations an article has gotten from all papers in the database (in this study it is Web of Science database) [41, 42].

233 The word cloud approach was applied to the abstracts of published studies, and the data was
234 pre-processed for this purpose. The pre-processing stage included transforming to lowercase,
235 tokenization, and removing stop words and numbers. Tokenization is the segmentation and
236 classification of separate elements of a corpus [43]. Besides, the emotions of the abstracts were
237 presented with sentiment analysis. For this kind of evaluation Valence Aware Dictionary for
238 Sentiment Reasoning (VADER) algorithm was used. In 2014 Gilbert and Hutto created VADER,
239 which is an advanced lexicon-based sentiment analysis technique. To assess feelings, VADER
240 employs a lexicon tailored to microblog-like settings. This algorithm classifies sentences as
241 negative, positive, neutral, and then calculates the compound value. The compound score, which
242 is the normalized sum of negative, positive, and neutral scores is between -1 to 1 , where -1 is a
243 very bad sentiment, and 1 is a very positive sentiment [44].

244

245 **3. Results**

246 As this study includes only Prof. Sourirajan's articles published in journals, there may be
247 discrepancies between the WoS data (all documents of Prof. Sourirajan) and the results presented
248 here. To eliminate this confusion, the results obtained from the WoS database by April 27th, 2022,
249 which is the day the data set was fetched, are given in Table 1.

250

251 **Table 1.** Prof. Sourirajan’s analysis results according to WoS and our Collection.

Data	Value
Documents (WoS Database)	166
Documents (Collection)	152
Sum of the Global Citations (WoS Database)	4253
Sum of the Global Citations (Collection)	3187
<i>h</i> -index (Global – WoS Database)	32
<i>h</i> -index (Collection)	31
<i>g</i> -index (Collection)	57

252

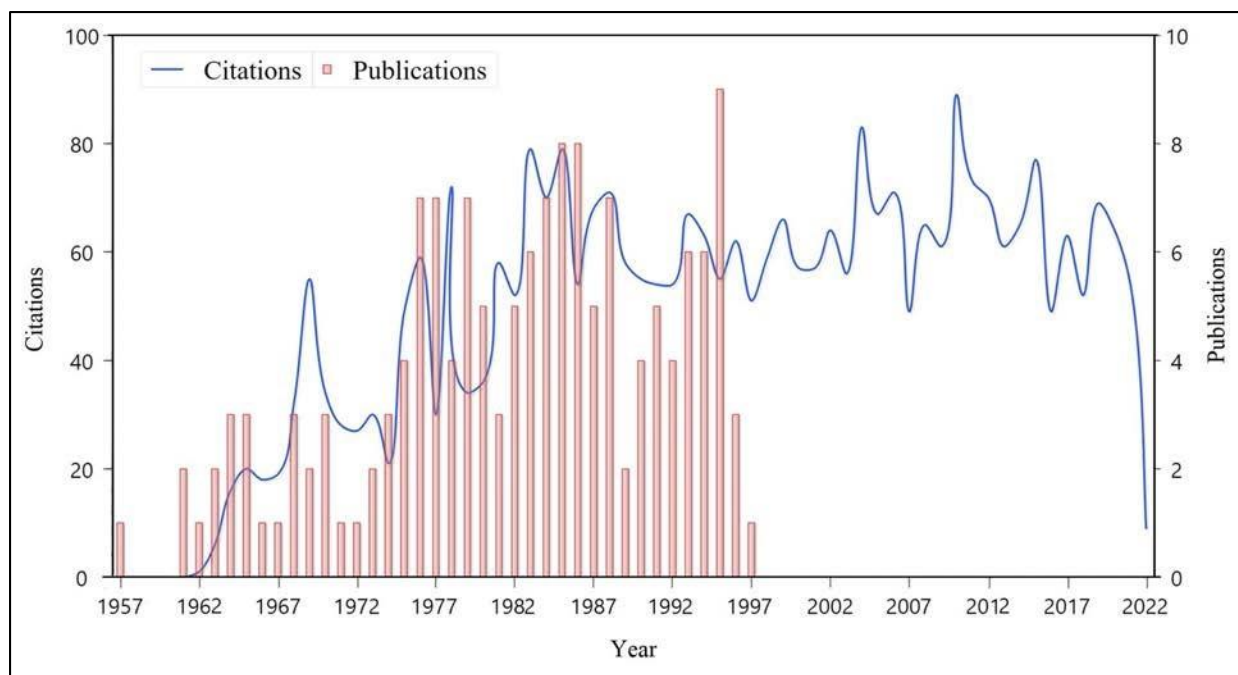
253 Prof. Sourirajan is an outstanding scientist who has received 4253 citations based on WoS
254 database with his 166 documents, including letters, meeting abstracts, editorial materials, etc.
255 According to WoS data, his *h*-index is 32. The main information of the dataset used in this study
256 can be seen in Table 2.

257

Table 2. Main Information about the domain.

Information	Result
Timespan	1957:1997
Sources (Journals, Books, etc.)	35
Documents	152
Average years from publication	40.6
Average citations per document	27.78
Average citations per year per document	0.6368
References	1438
Articles	138
Articles; proceedings papers	1
Letters	1
Meeting abstracts	2
Notes	5
Reviews	5
Keywords Plus	52
Author's Keywords	44
Authors	95
Author Appearances	475
Authors of single-authored documents	1
Authors of multi-authored documents	94
Single-authored documents	9
Documents per Author	1.6
Authors per Document	0.625
Co-Authors per Documents	3.12
Collaboration Index	0.657

261 The dataset in the study reveals that Prof. Sourirajan published papers in 35 different sources
 262 during his 40-year (1957-1997) career and reached 27.78 citations per document. This is a high
 263 value considering the document types in the collection. Fig. 1 shows his annual publications and
 264 received citations. He collaborated with 95 different authors and used 44 distinct keywords. In
 265 addition, Prof. Sourirajan published 9 single-authored documents. The document per author value
 266 was calculated as 1.6, and author per document value was calculated as 0.625. By considering
 267 these values with the collaboration index value (0.657), it is understood that Prof. Sourirajan
 268 worked with a limited research group in his career. The value of co-authors per document indicates
 269 that he published articles mainly with his three colleagues (Takeshi Matsuura, Kam Chan,
 270 Ramamurti Rangarajan).

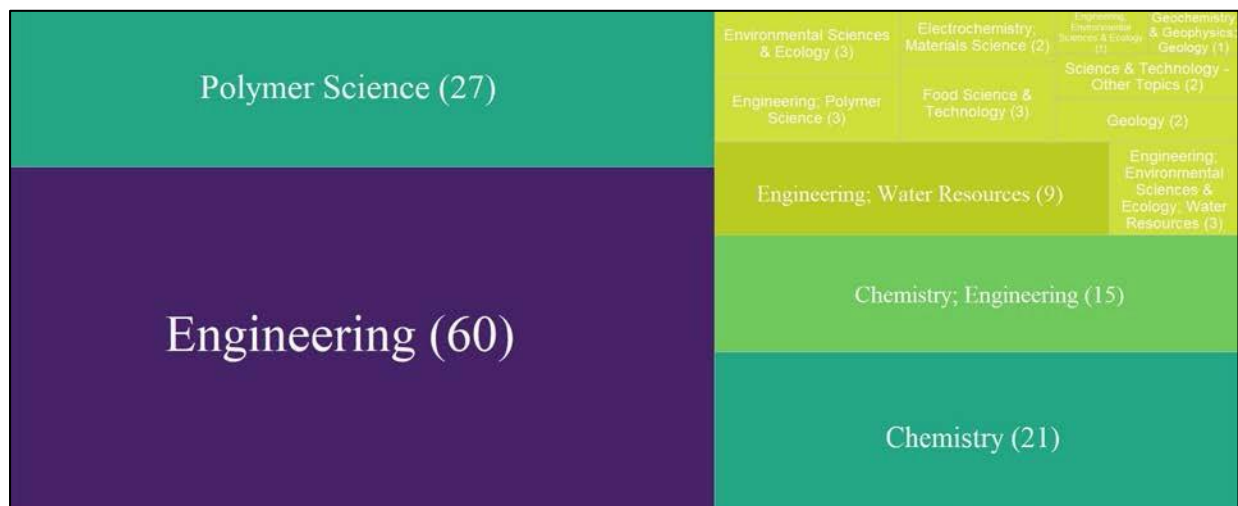


271

272 **Figure 1.** Yearly publications and citations of Prof. Sourirajan.

273 Since his first publication in 1957, Prof. Sourirajan maintained his scientific productivity with
 274 increasing momentum until he finished his stay at the National University of Singapore (NUS) in

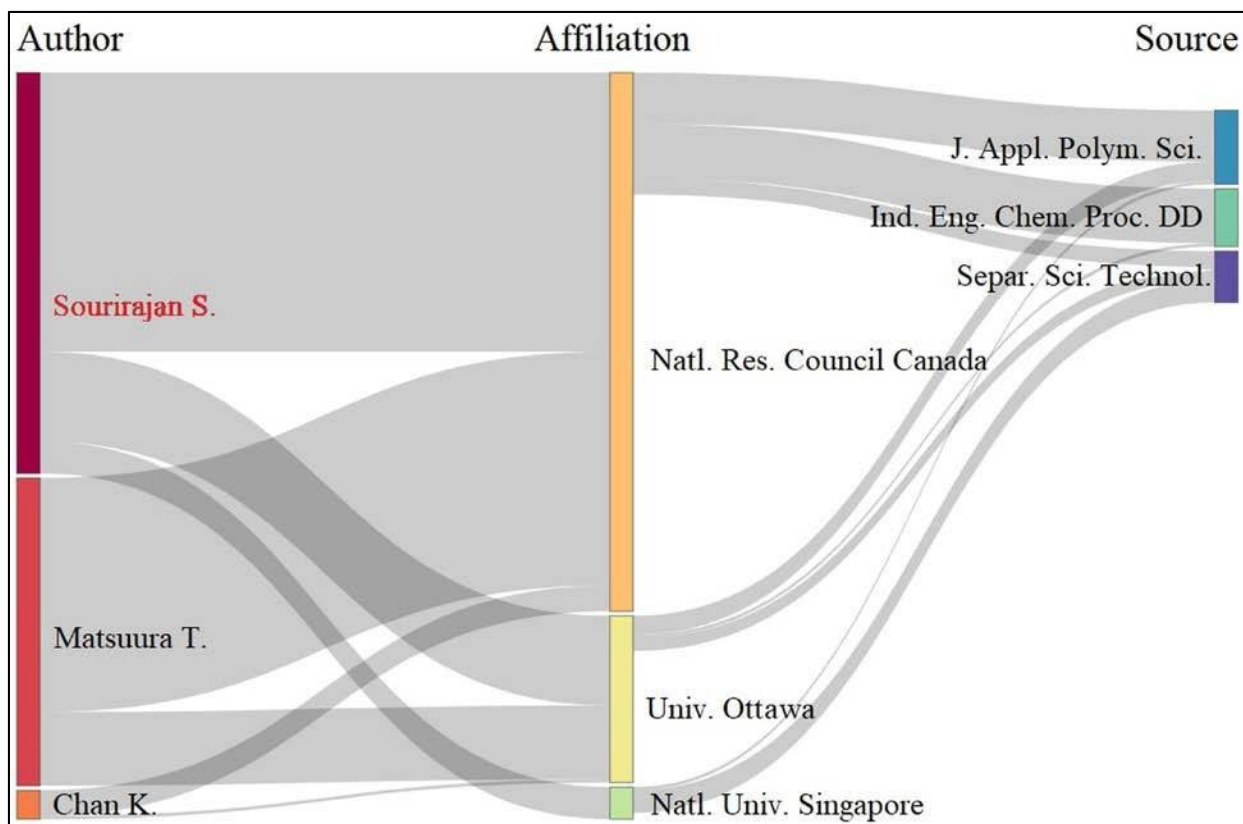
275 1995, when he published the largest number of articles (9). His articles received many citations by
 276 other researchers, and this statistic reached a peak value in 2010 with 89 citations. The WoS
 277 categories of Prof. Sourirajan's work are given in Fig. 2.



278

279 **Figure 2.** WoS categories of the collection.

280 Prof. Sourirajan published documents that contributed significantly to engineering and polymer
 281 science (See Fig. 2). His research studies were mainly published in engineering journals, with a
 282 number of 60. The next largest clusters were polymer science and chemistry journals.
 283 Collaborative actions are frequently the source of scientific publications. As a result, looking into
 284 author or organization collaboration is an useful method of bibliometric analysis [45]. The
 285 relationships between authors, affiliations, and sources (top 3) are depicted in Fig. 3.

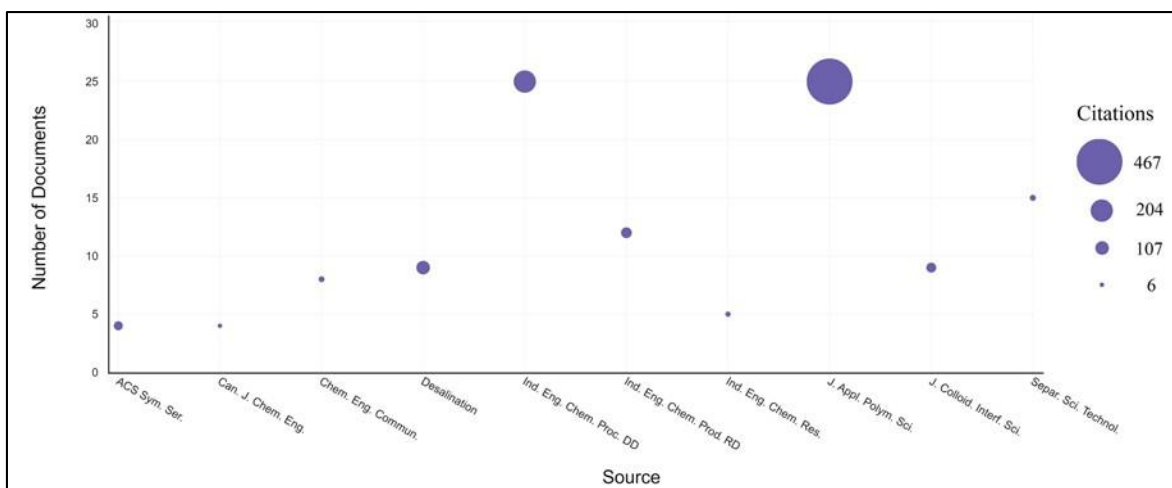


286

287

Figure 3. Three-field plot.

288 The three-field plot shows that Prof. Sourirajan was most productive while he was at NRC
 289 (Canada) and his publications in this center were mostly published in Journal of Applied Polymer
 290 Science, Industrial Engineering Chemistry Process Design and Development, and Separation
 291 Science and Technology. Prof. Sourirajan's other affiliations were University of Ottawa and
 292 National University of Singapore (NUS). The second most important author in this collection is T.
 293 Matsuura. Prof. Sourirajan has 99 joint publications with T. Matsuura in this collection. Fig. 4
 294 indicates the top 10 most important sources in the dataset with cited by values.



295

296

Figure 4. Top 10 most relevant sources with citation values.

297

298

299

300

301

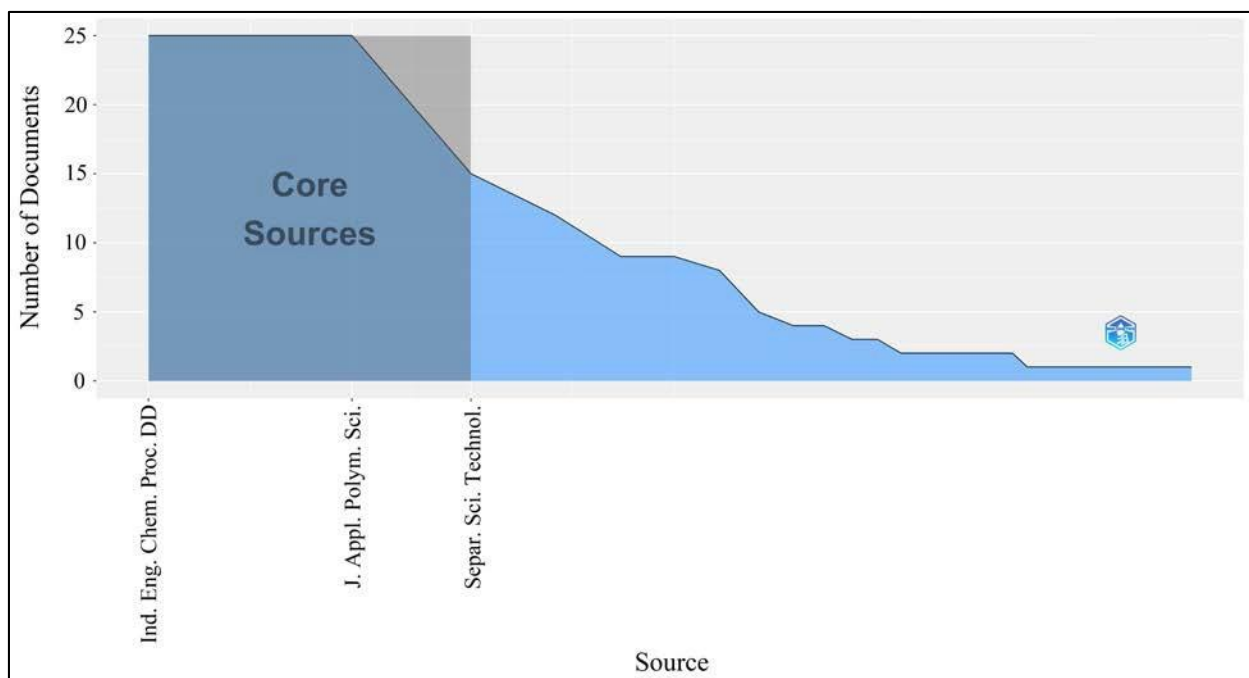
302

303

304

305

Fig. 4, which confirms Fig. 3, the first three journals in the top 10 of Sourirajan's most published journals are Journal of Applied Polymer Science, Industrial Engineering Chemistry Process Design and Development, and Separation Science and Technology, with 25, 25, and 15 publications, respectively. When the citations to ground studies of Prof. Sourirajan in these journals are examined, it is understood that the articles in the Journal of Applied Polymer Science are the most cited studies, with 467 citations. The interesting point in Fig. 4 is that although the number of documents in the Desalination journal is low (9 papers), the number of citations is high (107 cited by value). The Bradford's Law, shown in the following figure (Fig. 5), is a great approach for looking at journal production.

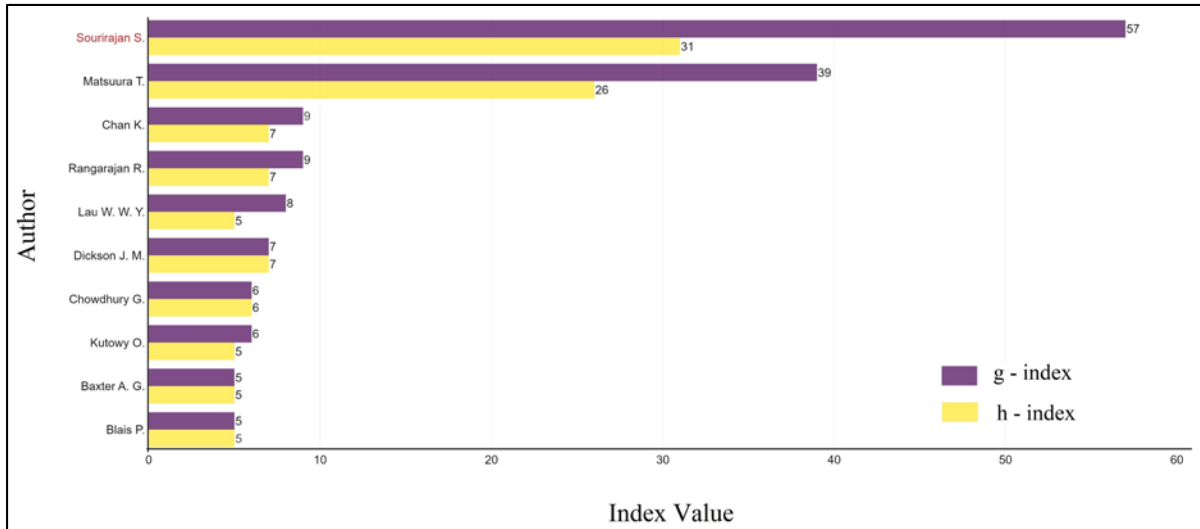


306

307

Figure 5. Bradford's Law plot.

308 According to both raw data and Bradford's Law, three journals stand out in Prof. Sourirajan's
 309 collection. Industrial Engineering Chemistry Process Design and Development, Journal of Applied
 310 Polymer Science, and Separation Science and Technology are the core sources and accounts for
 311 43 % of all his productivity. Besides identifying which journals publish the greatest number of
 312 articles, we can also look for the author analytics. The *h*-index and *g*-index values of the scientists
 313 in the data set are given in the figure below (Fig. 6).



314

315

Figure 6. *h*-index and *g*-index values of authors.

316

317

318

319

320

321

322

323

324

325

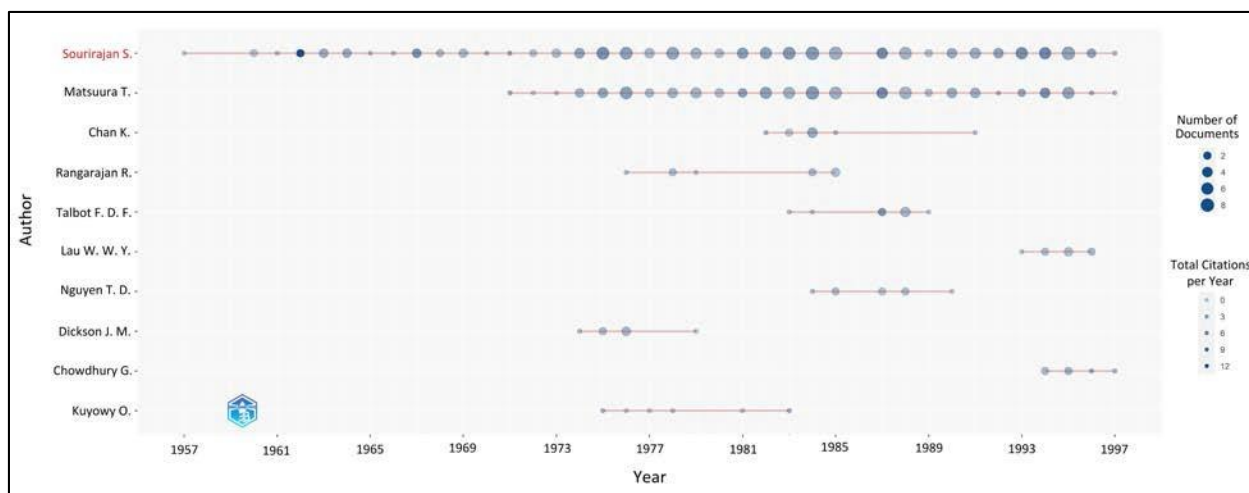
326

327

328

329

Since this dataset was fetched from WoS based on the author search criteria “Sourirajan Srinivasa” inherently, the metrics of Prof. Sourirajan have the highest values. His *g*-index value was calculated as 57, and his *h*-index was calculated as 31 based on our collection. As we mentioned while interpreting Table 1, there is a difference between the WoS database and our collection due to the adopted filtering process. For example, while the WoS database shows the *g*-index of Prof. Sourirajan as 32, our data set shows this value as 31, but the difference is very small. This situation also proves that Prof. Sourirajan was cited mostly for his works published in journals. T. Matsuura achieved the second-highest metrics with *g*-index of 39 and *h*-index of 26 (considered only Prof. Sourirajan’s data collection not T. Matsuura’s data collection). When assessing an author's relevance in a certain domain, two factors should be considered: influence and productivity. The number of papers published by an author in a certain period was used to assess productivity. In contrast, the influence was measured by the number of citations obtained each year. In Fig. 7, both of these criteria are thought to offer a snapshot of the top 10 authors in our collection.

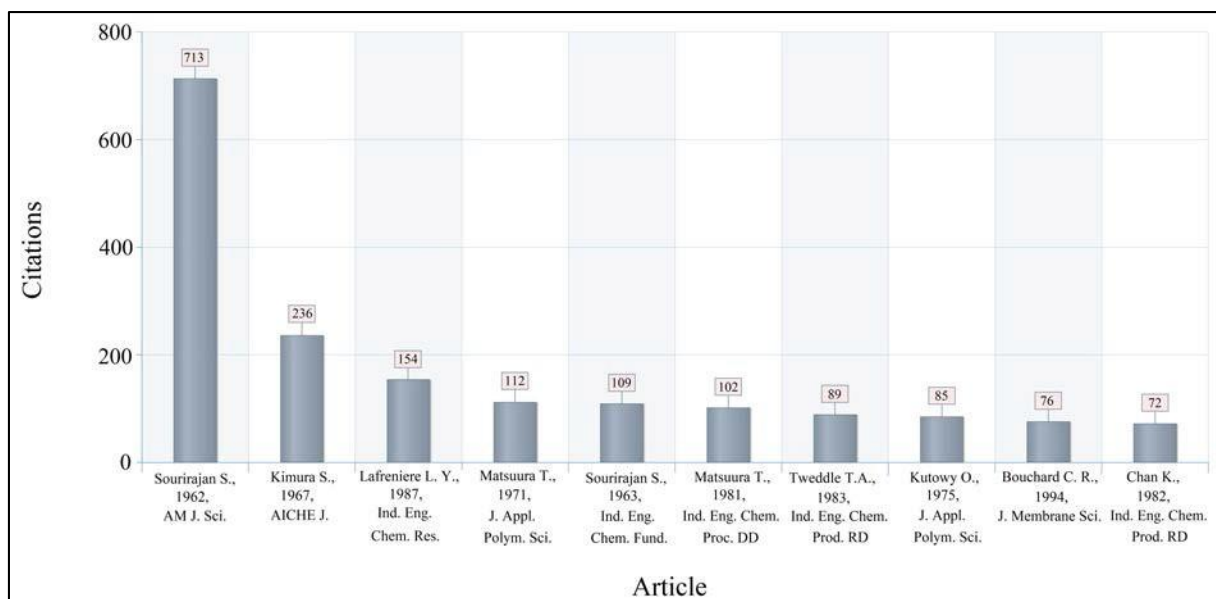


330

331

Figure 7. Author’s scientific production over time.

332 Fig. 7 shows the continuity of Prof. Sourirajan's scientific career. Since his first publication in
 333 1957, he has published scientific papers almost every year. In addition, the cited-by counts of his
 334 publications reveal how much of an influencer scientist he was. Fig. 7 can also give an idea about
 335 the researchers who have worked with him the longest. T. Matsuura is the longest-time
 336 collaborator with Sourirajan for 28 years, from 1971 to 1997. He was followed by K. Chan, R.
 337 Rangarajan, and O. Kutowy, respectively. It is critical to uncover new facts by building on past
 338 queries while doing scientific research. Looking at the most quoted sources may be relevant in a
 339 descriptive assessment of this bibliometric study. Fig. 8 shows the most cited documents in the
 340 collection.



341

342

Figure 8. Most global cited documents.

343

344

345

346

347

348

349

350

351

352

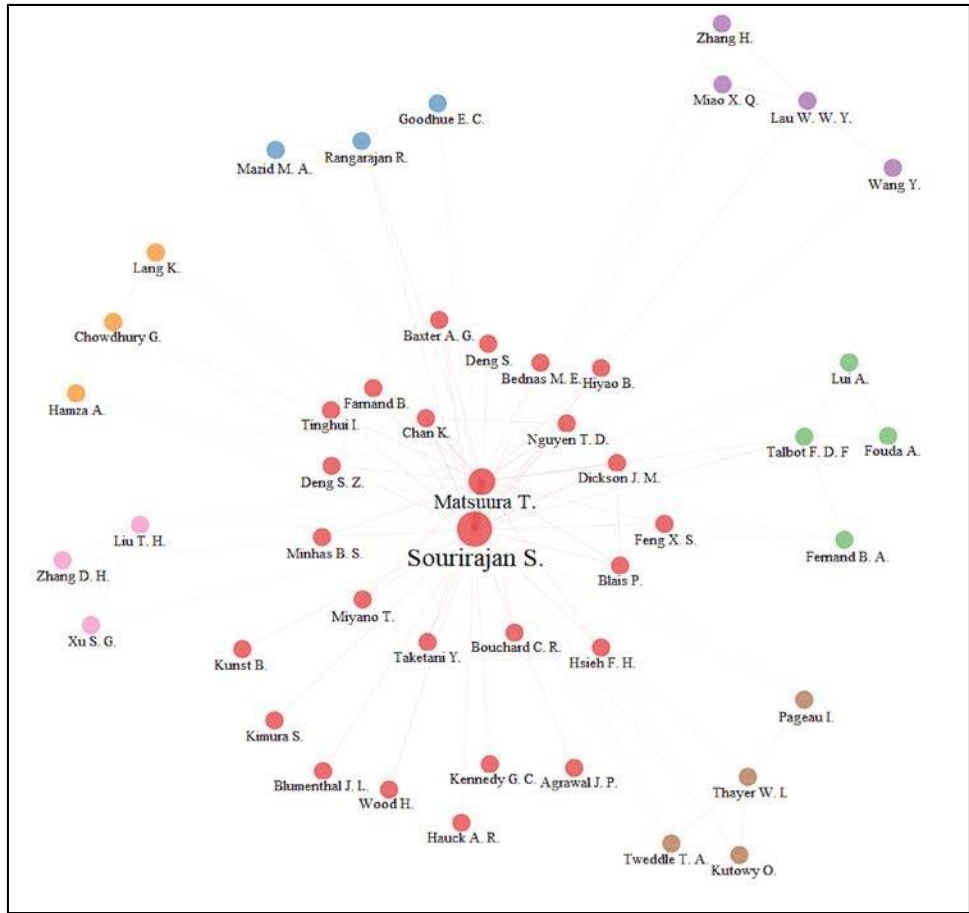
353

354

The paper entitled “*System H₂O-NaCl at Elevated Temperatures and Pressures*” by Sourirajan and Kennedy is the most globally cited document with 713 citations until April 27th 2022. In this paper it was claimed the NaCl content in saline hot springs didn’t contradict the hypothesis that the NaCl was transported from the magma source in gaseous H₂O. This research was conducted at pressures up to 1240 bar and temperatures ranging from 250 to 700 °C. For the first time a temperature-pressure curve was illustrated for the gas-solid-liquid boundary of geothermal water. Sourirajan and Kennedy measured the solubility of solid NaCl up to saturation pressures in steam. Various isotherms were reported for critical pressure and composition [46]. The next paper following this study taking into account the number of citations is entitled “*Analysis of Data in Reverse Osmosis with Porous Cellulose Acetate Membranes Used*” with 236 citations. In this publication, Kimura and Sourirajan conducted some experiments in a reverse osmosis system with porous cellulose acetate membrane to evaluate the diffusivity of some inorganic salts [47].

355 Prof. Sourirajan's metrics such as *h*-index, *g*-index, citations, collaboration network, average
356 citations per document, average citations per year per document, and collaboration index, may
357 seem low when compared to the metrics of today's scientists. However, it is unfair to compare the
358 metrics of a researcher who published the first manuscript late fifties, like Prof. Sourirajan, with
359 today's researchers. Decades ago, the publication of a single manuscript was time consuming.
360 Manuscripts used to be submitted by post and it took months to get a reply from journals editors.
361 Many of the technologies we now take for granted such as the internet, software applications,
362 electronic journals were not available at the time. The rising usage of internet sources for academic
363 and research purposes is a sign that web-based information is playing a significant role in scholarly
364 communication, research collaboration, manuscripts publications and citations, etc.

365 It is a common method for analyzing bibliometric data to produce a list of the collaboration
366 network in a specific field and examine the details. These kinds of approaches illustrate the
367 relationships between authors, institutions or countries. For this purpose, the illustration of the
368 collaboration network of authors in the selected domain is given in Fig. 9.



369

370

Figure 9. Collaboration network of authors.

371

It was stated in Table 2 that there were 95 different authors in the dataset. The linkage of 47

372

researchers out of these 95 scientists who collaborated the most is clearly seen in Fig. 9. Different

373

colors represent different network clusters. Seven different workgroups have been identified in the

374

domain. The main connection of the scientists in these working groups is with Prof. Sourirajan,

375

but there are also other connections between other names. As expected, Prof. Sourirajan is at the

376

center of the collaboration network. Moreover, he is not only the center of the spider web but also

377

the most important element in the largest work group indicated in red and located in this figure. T.

378

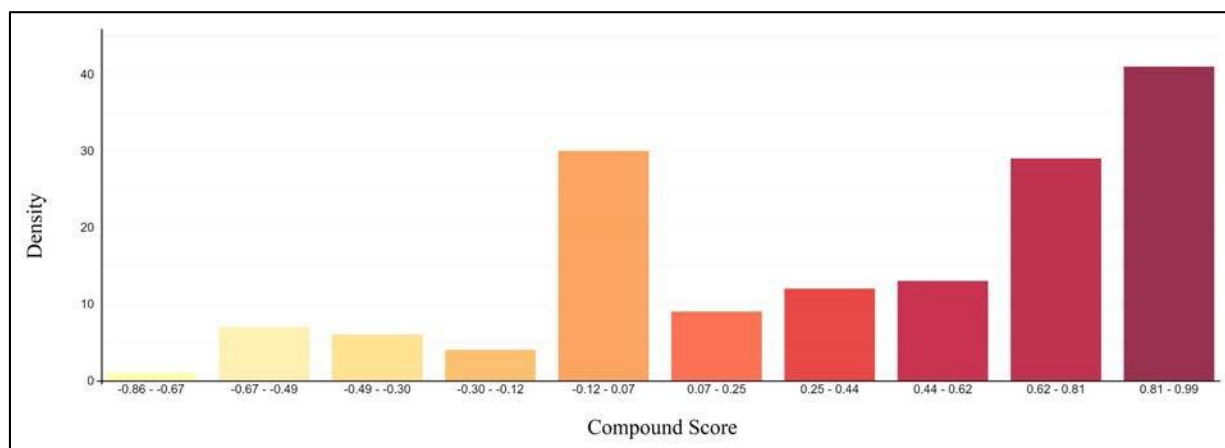
Matsuura, the scientist who has collaborated with Prof. Sourirajan the most, has the thickest

379

connection line with him and appears as one of the main elements of the whole network and the

389 Fig. 10 appears as a self-explanatory illustration. Since Prof. Sourirajan is known as the
390 “*father of reverse osmosis*”, “co-creator of the *cellulose acetate membrane*”, “a scientist
391 dedicating most of his time to *membrane separation*”, “a *data* analyst dealing with *parameters*”,
392 “a membrane engineer working with many *polymers, solutions, solvents and solutes*”, “an
393 individual who has served humanity with the *results* he has *achieved*”, “a co-inventor of *casting*
394 and *phase inversion technique*”, we can easily understand his life purpose by just looking to the
395 sketch above.

396 The next stage of our work is the sentiment analysis of the abstracts of Prof. Sourirajan's
397 papers. In Fig. 11, the analysis results are given in bar charts.



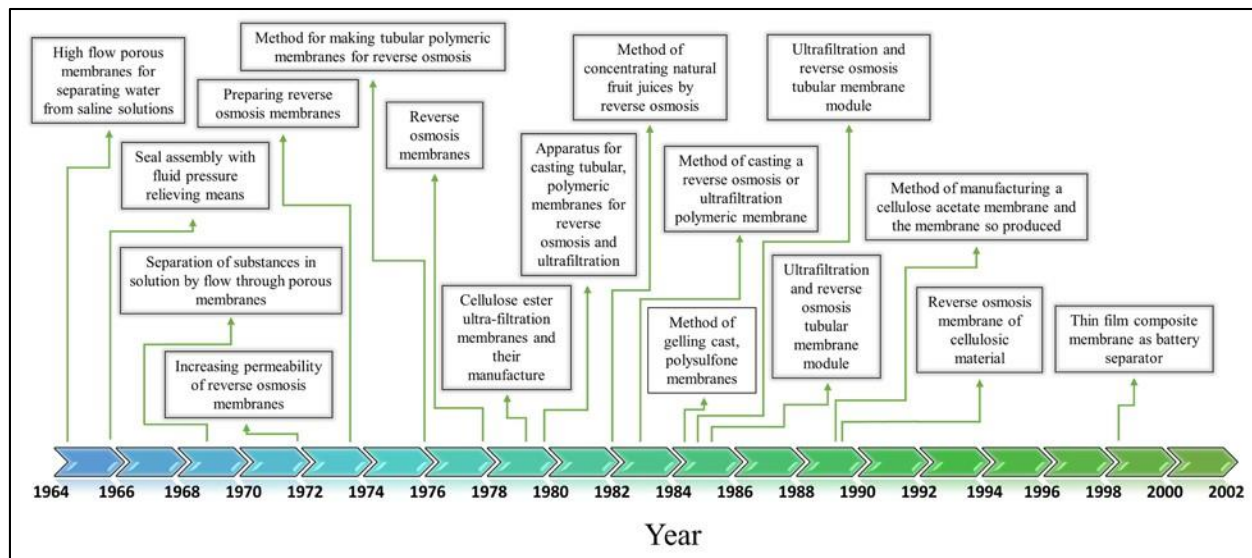
398

399 **Figure 11.** Sentiment analysis results.

400 Sentiment analysis results are in a spectrum that -1 is a pessimistic score, 0 is a neutral score,
401 and 1 is an optimistic score. In Prof. Sourirajan's studies, it was seen that the intensity of the
402 emotions is either neutral and positive. He and his colleagues achieved a compound score of 0.07
403 or higher in 104 studies. This rate corresponds to $\sim 68\%$ of documents in the domain. These results
404 showed that Prof. Sourirajan heralded more positive results to the readers in his published papers.
405 Maybe this is the reason for his prolific and highly influential work aided in establishing scientific

406 routes that tackled some of the world's most important social concerns, such as the access to
 407 drinkable water. Prof. Sourirajan is an outstanding professional who has left important studies with
 408 positive emotions to the scientific community. Apart from his works published in journals, Prof.
 409 Sourirajan also published books, book chapters, and filled patents. Prof. Sourirajan is a scientist
 410 who has gone too far in this regard with 17 patents [48]. The timeline and co-inventors of his
 411 patents is shown in Figs. 12 and 13, respectively.

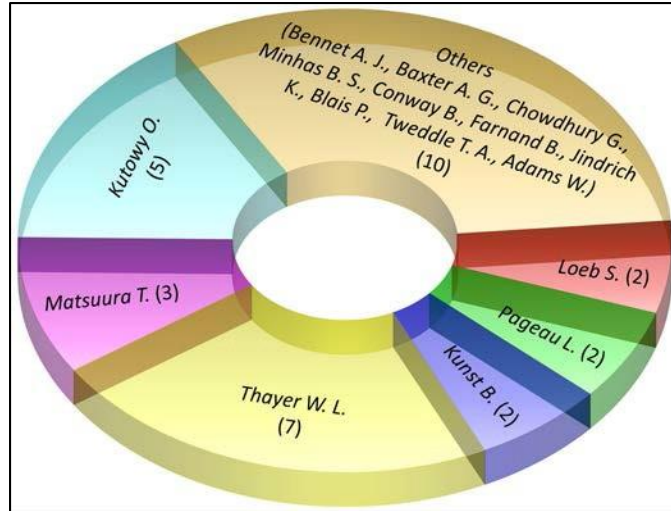
412



413

414

Figure 12. Timeline of Prof. Sourirajan’s patents.



415

416 **Figure 13.** Co-inventors of Prof. Sourirajan in his patents (the numbers in parenthesis show the
 417 number of joint patents).

418 Fig. 12 indicated that Prof. Sourirajan made inventions throughout all his scientific career.
 419 This proves how he was devoted to science and progress. In Fig. 13, it can be seen that Prof.
 420 Sourirajan filled patents with a total of 16 colleagues (Among them, W. L. Thayer contributed in
 421 7 patents and O. Kutow comes in the second place with 5 joint patents and then with T. Matsuura
 422 with 3 patents).

423 Prof. Sourirajan was also dedicated to academy. He published the following 3 comprehensive
 424 books on RO and UF separation processes including synthetic membranes and mathematical
 425 analysis.

- 426• Reverse Osmosis, 1970, published by Logos Press (Later Academic Press)
- 427• Reverse Osmosis and Synthetic Membranes-Theory, Technology and Engineering, Edited by S.
 428 Sourirajan, 1977, published by National Research Council of Canada
- 429• Reverse Osmosis/Ultrafiltration Process Principles, 1985, published by National Research Council
 430 of Canada.

431 Through these books, his main objective was to contribute to the development of the science,
432 technology, and engineering involved in these separation processes and in all their applications.

433

434 **4. Conclusions**

435 This study was carried out to keep the memory of Prof. Sourirajan alive by examining his
436 documents published in refereed journals with bibliometric and text mining analyses. According
437 to the findings of the results, it was understood that Prof. Sourirajan was an outstanding and a
438 productive researcher. The impact of his works on the scientific world has been revealed with
439 figures such as the number of citations he received, his patents, and his collaboration network. He
440 is a role model for all of us and for the future generations with interest in science, membrane
441 engineering, desalination and water treatment. His works paved the way for all membranologists
442 and researchers devoted to desalination and water treatment. On behalf of the scientific community
443 and researchers working on membrane processes, we would like to thank him for his outstanding
444 contributions to science and engineering.

445

446 **Acknowledgements**

447 Dr. Ersin Aytaç would like to express his acknowledgement for the postdoctoral grant received
448 from The Scientific and Technological Research Council of Turkey (TUBITAK) at the University
449 Complutense of Madrid (UCM) with the grant number of 1059B191900618.

450

451

452 **References**

- 453 [1] S. Sourirajan, Reverse Osmosis, Academic Press, 1970.
- 454 [2] Water-Desalination-Report, [A Tribute to Srinivasa Sourirajan], Houston, Texas, 2022.
- 455 [3] Y. Orooji, Father of reverse osmosis who made a huge impact on our world: Srinivasa
456 Sourirajan (October 16, 1923–February 20, 2022), npj Clean Water, 5 (2022) 20.
- 457 [4] A. Husain, A. El-Nashar, A.A. Radif, Membrane-Based Desalination Processes, in: D.M.K.
458 Al-Gobaisi (Ed.) Physical, Chemical and Biological Aspects of Water, UNESCO-EOLSS, 2010,
459 pp. 160.
- 460 [5] M.M. Denn, Chemical Engineering: An Introduction, 1st ed., Cambridge University Press,
461 2011.
- 462 [6] J. Kucera, Reverse Osmosis: Industrial Processes and Applications, 2nd ed., John Wiley &
463 Sons Inc, 2015.
- 464 [7] V.H. Edwards, S. Shelley, Careers in Chemical and Biomolecular Engineering, 1st ed., CRC
465 Press, 2018.
- 466 [8] H.-H. Tseng, W.J. Lau, M.A. Al-Ghouti, L. An, Preface, in: H.-H. Tseng, W.J. Lau, M.A. Al-
467 Ghouti, L. An (Eds.) 60 Years of the Loeb-Sourirajan Membrane, Elsevier, 2022, pp. xxvii-xxviii.
- 468 [9] R.W. Baker, Membranes and Modules, in: Membrane Technology and Applications, John
469 Wiley & Sons, Ltd, 2004, pp. 89-160.
- 470 [10] S. Sourirajan, Separation of Gases by Flow under Pressure through Porous Membranes,
471 Nature, 199 (1963) 590-591.
- 472 [11] S. Sourirajan, Separation of Hydrocarbon Liquids by Flow Under Pressure Through Porous
473 Membranes, Nature, 203 (1964) 1348-1349.

- 474 [12] S. Sourirajan, Ottawa Citizen, (2022).
475 <https://ottawacitizen.remembering.ca/obituary/srinivasa-sourirajan-1084507224> (accessed
476 August 8, 2022)
- 477 [13] Q. Zhan, C. Wu, H. Ding, Y. Huang, Z. Jiang, N. Liao, K. Wang, Y. Li, Emerging trends in
478 photodynamic therapy for head and neck cancer: A 10-year bibliometric analysis based on
479 CiteSpace, *Photodiagnosis and Photodynamic Therapy*, 38 (2022) 102860.
- 480 [14] T. Pang, J. Shen, Visualizing the landscape and evolution of capacitive deionization by
481 scientometric analysis, *Desalination*, 527 (2022) 115562.
- 482 [15] A.P. Mörschbacher, C.E. Granada, Mapping the worldwide knowledge of antimicrobial
483 substances produced by *Lactobacillus* spp.: A bibliometric analysis, *Biochemical Engineering*
484 *Journal*, 180 (2022) 108343.
- 485 [16] Y. Fomina, A. Glińska-Noweś, A. Ignasiak-Szulc, Community supported agriculture: Setting
486 the research agenda through a bibliometric analysis, *Journal of Rural Studies*, 92 (2022) 294-305.
- 487 [17] Z. Zheng, S. Zhu, M. Lv, Z. Gu, H. Hu, Harnessing nanotechnology for cardiovascular disease
488 applications - a comprehensive review based on bibliometric analysis, *Nano Today*, 44 (2022)
489 101453.
- 490 [18] Z.A. Polat, M. Alkan, J. Paulsson, J.M. Paasch, E. Kalogianni, Global scientific production
491 on LADM-based research: A bibliometric analysis from 2012 to 2020, *Land Use Policy*, 112
492 (2022) 105847.
- 493 [19] J.K. Tamala, E.I. Maramag, K.A. Simeon, J.J. Ignacio, A bibliometric analysis of sustainable
494 oil and gas production research using VOSviewer, *Cleaner Engineering and Technology*, 7 (2022)
495 100437.

496 [20] E. Aytaç, Unsupervised learning approach in defining the similarity of catchments:
497 Hydrological response unit based k-means clustering, a demonstration on Western Black Sea
498 Region of Turkey, *International Soil and Water Conservation Research*, 8 (2020) 321-331.

499 [21] E.R. Okoroafor, C.M. Smith, K.I. Ochie, C.J. Nwosu, H. Gudmundsdottir, M. Aljubran,
500 Machine learning in subsurface geothermal energy: Two decades in review, *Geothermics*, 102
501 (2022) 102401.

502 [22] E. Aytaç, Modeling Future Impacts on Land Cover of Rapid Expansion of Hazelnut Orchards:
503 A Case Study on Samsun, Turkey, *European Journal of Sustainable Development Research*, 6
504 (2022) em0193.

505 [23] F. Saffarimiandoab, R. Mattesini, W. Fu, E.E. Kuruoglu, X. Zhang, Insights on features'
506 contribution to desalination dynamics and capacity of capacitive deionization through machine
507 learning study, *Desalination*, 515 (2021) 115197.

508 [24] E. Aytaç, Havzaların Benzerliklerini Tanımlamada Alternatif Bir Yaklaşım: Hiyerarşik
509 Kümeleme Yöntemi Uygulaması, *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri*
510 *Dergisi*, 21 (2021) 958 - 970.

511 [25] M. Dissanayake, H. Nguyen, K. Poologanathan, G. Perampalam, I. Upasiri, H. Rajanayagam,
512 T. Suntharalingam, Prediction of shear capacity of steel channel sections using machine learning
513 algorithms, *Thin-Walled Structures*, 175 (2022) 109152.

514 [26] A. Occhipinti, L. Rogers, C. Angione, A pipeline and comparative study of 12 machine
515 learning models for text classification, *Expert Systems with Applications*, 201 (2022) 117193.

516 [27] A. Shafizadeh, H. Shahbeig, M. Hossein Nadian, H. Mobli, M. Dowlati, V. Kumar Gupta, W.
517 Peng, S. Shiung Lam, M. Tabatabaei, M. Aghbashlo, Machine learning predicts and optimizes
518 hydrothermal liquefaction of biomass, *Chemical Engineering Journal*, (2022) 136579.

- 519 [28] E. Aytac, Forecasting Turkey's Hazelnut Export Quantities with Facebook's Prophet
520 Algorithm and Box-Cox Transformation, *ADCAIJ-Advances in Distributed Computing and*
521 *Artificial Intelligence Journal*, 10 (2021) 33-47.
- 522 [29] E. Aytac, Exploring Electrocoagulation Through Data Analysis and Text Mining
523 Perspectives, *Environmental Engineering & Management Journal (EEMJ)*, 22 (2022) 671-685.
- 524 [30] B. Garner, C. Thornton, A. Luo Pawluk, R. Mora Cortez, W. Johnston, C. Ayala, Utilizing
525 text-mining to explore consumer happiness within tourism destinations, *Journal of Business*
526 *Research*, 139 (2022) 1366-1377.
- 527 [31] M.A. Sayed, X. Qin, R.J. Kate, D.M. Anisuzzaman, Z. Yu, Identification and analysis of
528 misclassified work-zone crashes using text mining techniques, *Accident Analysis & Prevention*,
529 159 (2021) 106211.
- 530 [32] G. Liu, M. Boyd, M. Yu, S.Z. Halim, N. Quddus, Identifying causality and contributory
531 factors of pipeline incidents by employing natural language processing and text mining techniques,
532 *Process Safety and Environmental Protection*, 152 (2021) 37-46.
- 533 [33] T. Zhang, N.V. Sahinidis, C.P. Rosé, S. Amaran, B. Shuang, Forty years of Computers and
534 Chemical Engineering: Analysis of the field via text mining techniques, *Computers & Chemical*
535 *Engineering*, 129 (2019) 106511.
- 536 [34] B. Goodwin, E.S. Wilson, M.M. Oware, H.H. Kestian, T.J. Stigdon, E.A. Miller, Text mining
537 and the examination of language used to report child maltreatment: How language impacts child
538 welfare intake reports, *Children and Youth Services Review*, 133 (2022) 106344.
- 539 [35] R.L. Atenstaedt, Word cloud analysis of historical changes in the subject matter of public
540 health practice in the United Kingdom, *Public Health*, 197 (2021) 39-41.

- 541 [36] M. Bibi, W.A. Abbasi, W. Aziz, S. Khalil, M. Uddin, C. Iwendi, T.R. Gadekallu, A novel
542 unsupervised ensemble framework using concept-based linguistic methods and machine learning
543 for twitter sentiment analysis, Pattern Recognition Letters, 158 (2022) 80-86.
- 544 [37] J. Liao, M. Wang, X. Chen, S. Wang, K. Zhang, Dynamic commonsense knowledge fused
545 method for Chinese implicit sentiment analysis, Information Processing & Management, 59 (2022)
546 102934.
- 547 [38] M. Aria, C. Cuccurullo, bibliometrix: An R-tool for comprehensive science mapping analysis,
548 Journal of Informetrics, 11 (2017a) 959-975.
- 549 [39] Exploratory-Inc., (2020). <https://exploratory.io> (accessed February 18, 2022)
- 550 [40] N. Desai, L. Veras, A. Gosain, Using Bradford's law of scattering to identify the core journals
551 of pediatric surgery, Journal of Surgical Research, 229 (2018) 90-95.
- 552 [41] M. Aria, C. Cuccurullo, Biblioshiny : Bibliometrix for no Coders, (2017b).
553 <https://www.bibliometrix.org/biblioshiny/assets/player/KeynoteDHTMLPlayer.html#1> (accessed
554 March 30, 2022)
- 555 [42] M. Aria, C. Cuccurullo, A brief introduction to bibliometrix, (2017c).
556 https://www.bibliometrix.org/vignettes/Introduction_to_bibliometrix.html (accessed March 30,
557 2022)
- 558 [43] Y. Ding, J. Ma, X. Luo, Applications of natural language processing in construction,
559 Automation in Construction, 136 (2022) 104169.
- 560 [44] S. Suardi, A.R. Rasel, B. Liu, On the predictive power of tweet sentiments and attention on
561 bitcoin, International Review of Economics & Finance, 79 (2022) 289-301.
- 562 [45] O. Kilicoglu, H. Mehmetcik, Science mapping for radiation shielding research, Radiation
563 Physics and Chemistry, 189 (2021) 109721.

564 [46] S. Sourirajan, G.C. Kennedy, The system H₂O-NaCl at elevated temperatures and pressures,
565 American Journal of Science, 260 (1962) 115-141.

566 [47] S. Kimura, S. Sourirajan, Analysis of data in reverse osmosis with porous cellulose acetate
567 membranes used, Aiche Journal, 13 (1967) 497-503.

568 [48] Google-Patents, Patents of Sourirajan Srinivasa, (2022).

569 <https://patents.google.com/?inventor=sourirajan+srinivasa&oq=sourirajan+srinivasa> (accessed

570 April 28, 2022)

571