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

Mohamed Khayet [1,2,*], Ersin Aytaç [1,3], Takeshi Matsuura[4]

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

[2]Madrid Institute for Advanced Studies of Water (IMDEA Water Institute), Calle Punto Net Nº 4, 28805, Alcalá de Henares, Madrid, Spain.

[3] Zonguldak Bülent Ecevit University, Department of Environmental Engineering, 67100, Zonguldak, Turkey

[4] Department of Chemical and Biological Engineering, University of Ottawa, Ottawa, ON K1N 6N5, Canada

* Corresponding author:

khayetm@fis.ucm.es

Tel. +34-91-3945185

Abstract

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

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

Highlights

- Prof. Srinivasa Sourirajan is known as the "*Father of Reverse Osmosis*".
- Prof. Sourirajan has been nominated three times for the Nobel Prize.
- His research studies were not only novel, but also very creative and even visionary.
- Prof. Sourirajan has filled 17 patents about reverse osmosis membranes and applications.
- Machine learning results reveal how influential scientist he was.

1. Introduction

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Articles Fractionized (AU_j) calculates each author's contribution to a set of published articles based on the idea that all co-authors contribute evenly to each article and can be calculated as follows:

$$AU_j = \sum_{j=1}^{n} \frac{1}{h} \quad (3)$$

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

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].

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

3. Results

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

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
h-index (Global – WoS Database)	32
h-index (Collection)	31
g-index (Collection)	57

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

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

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

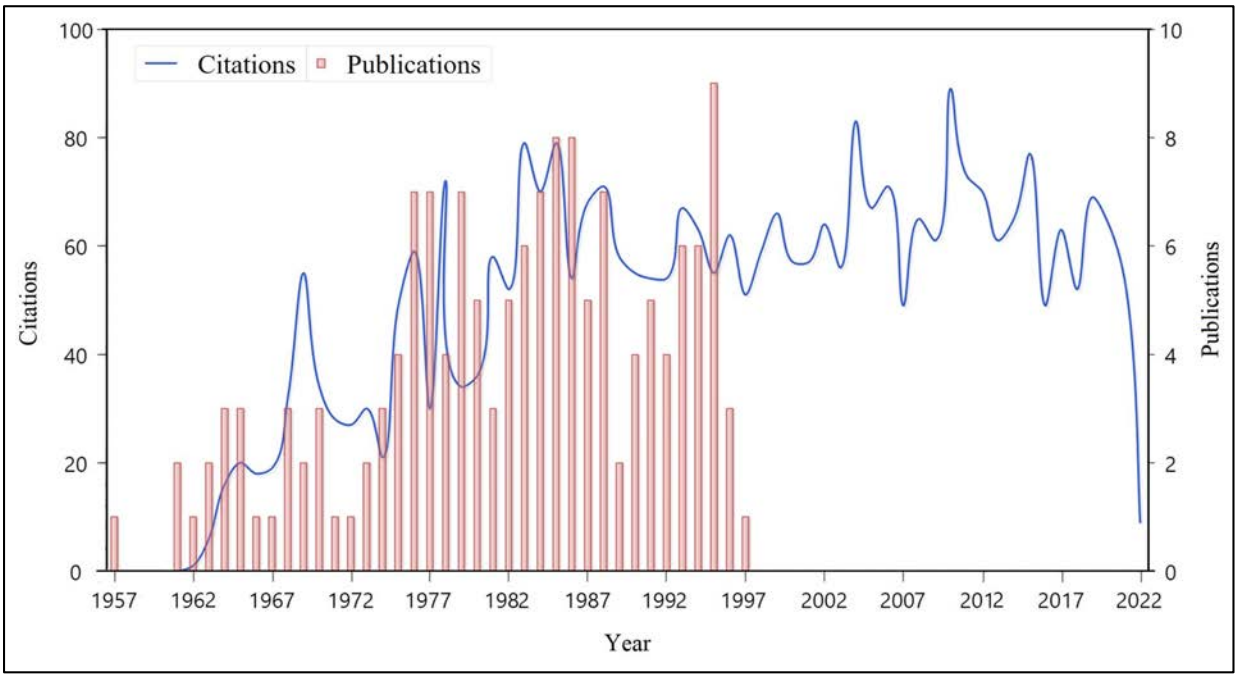


Figure 1. Yearly publications and citations of Prof. Sourirajan.

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

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

Figure 2. WoS categories of the collection.

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

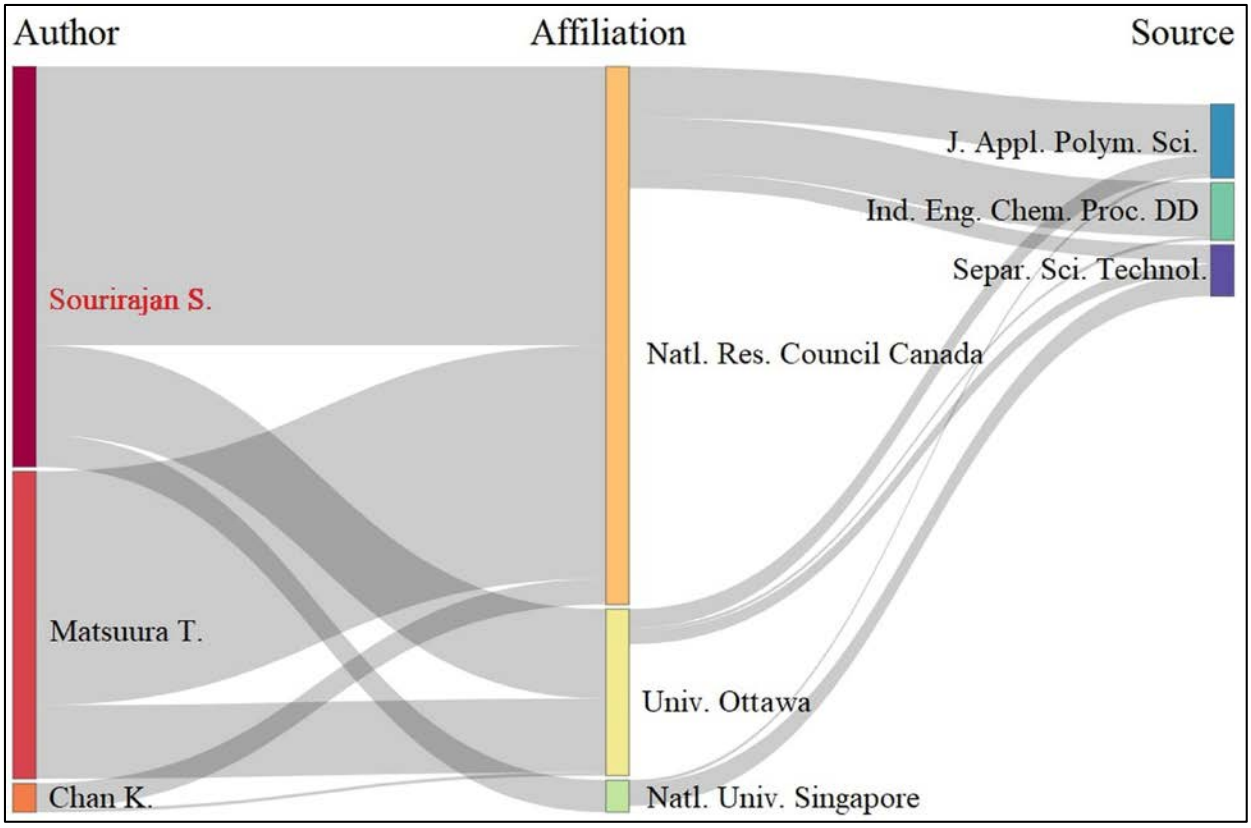


Figure 3. Three-field plot.

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

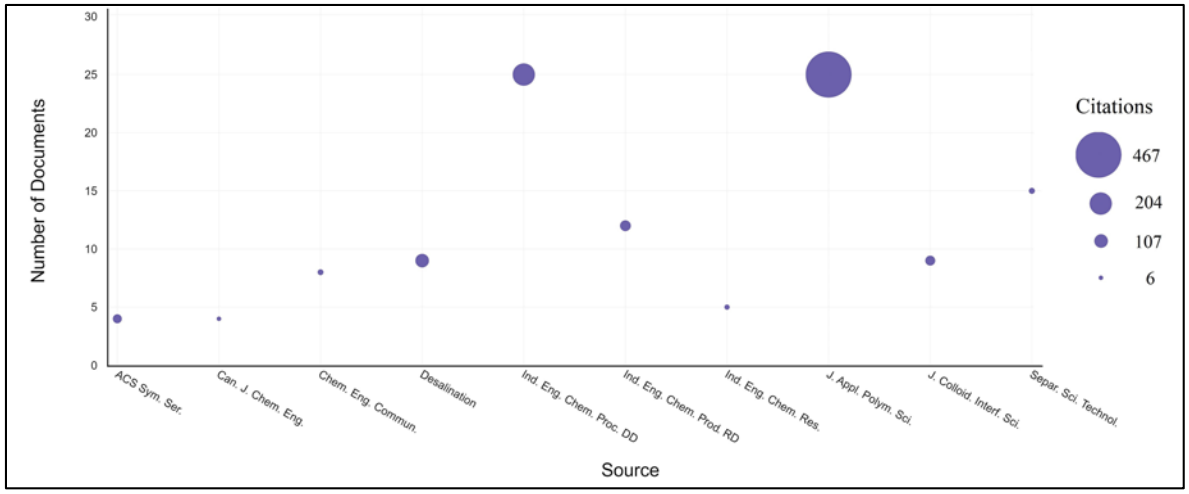


Figure 4. Top 10 most relevant sources with citation values.

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.

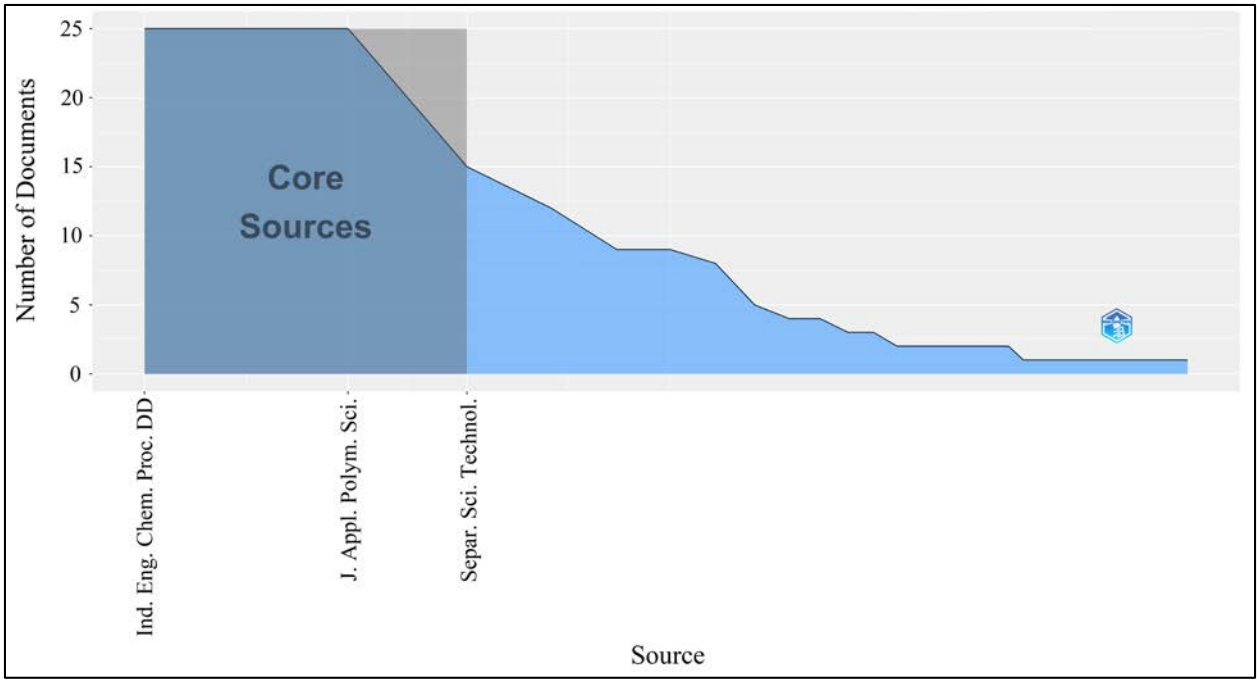


Figure 5. Bradford's Law plot.

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

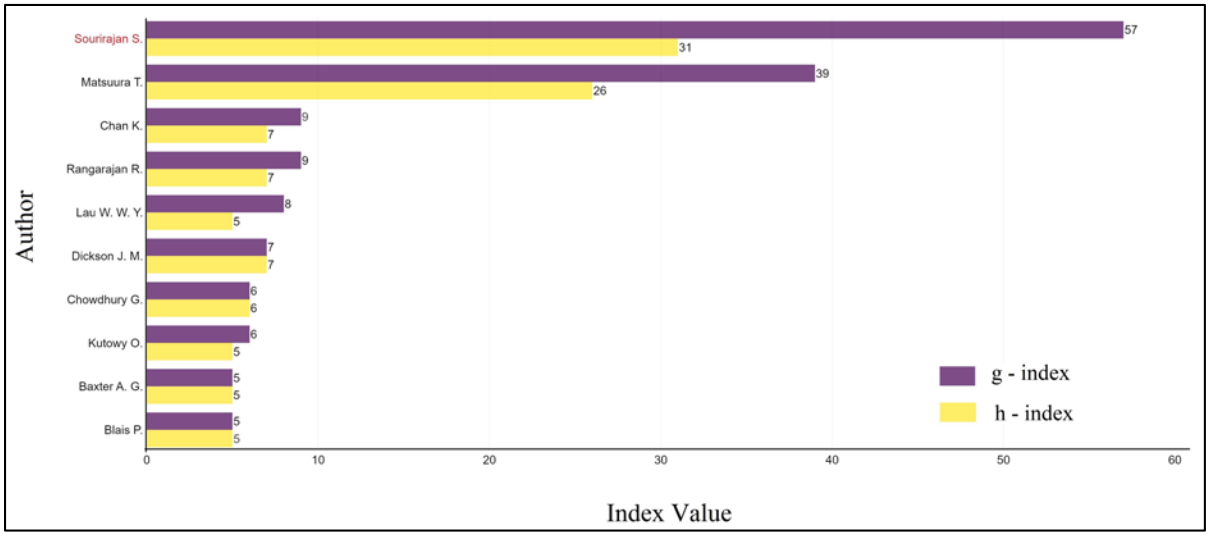


Figure 6. *h*-index and *g*-index values of authors.

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.

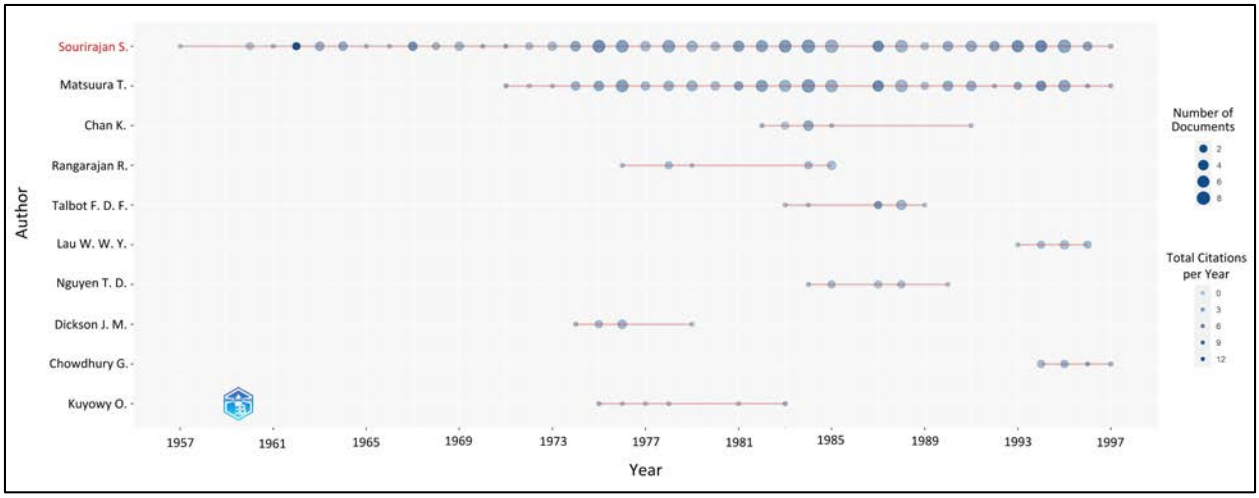


Figure 7. Author's scientific production over time.

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

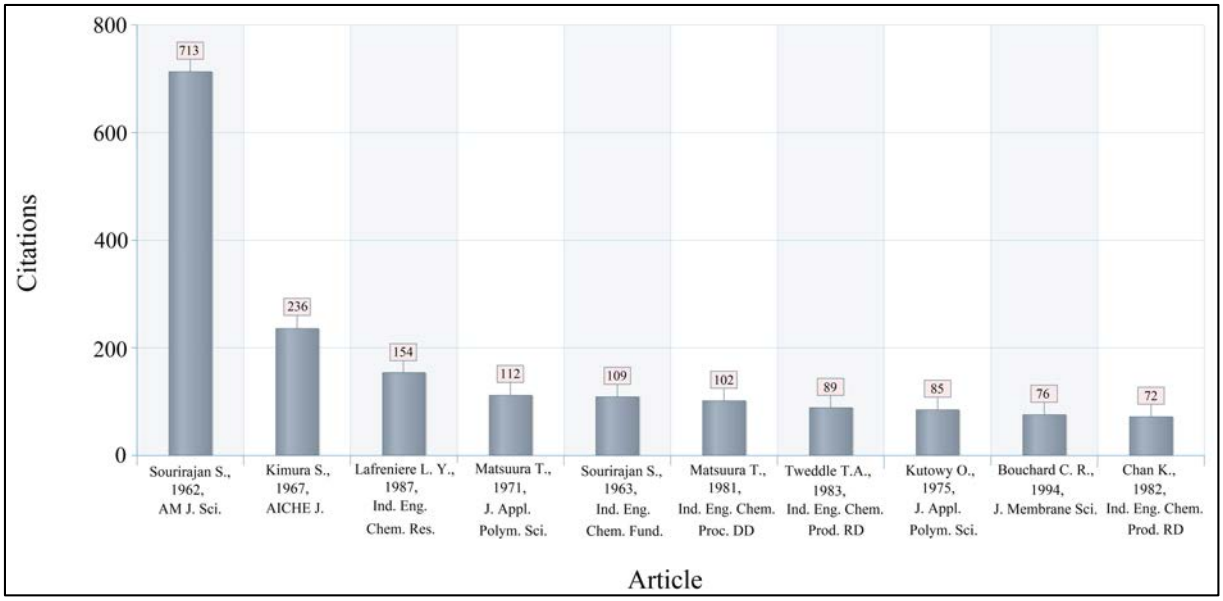


Figure 8. Most global cited documents.

The paper entitled "*System H_2O-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 H_2O. This research was conducted at pressures up to 1240 bar and temperatures ranging from 250 to 700 ^{0}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].

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

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

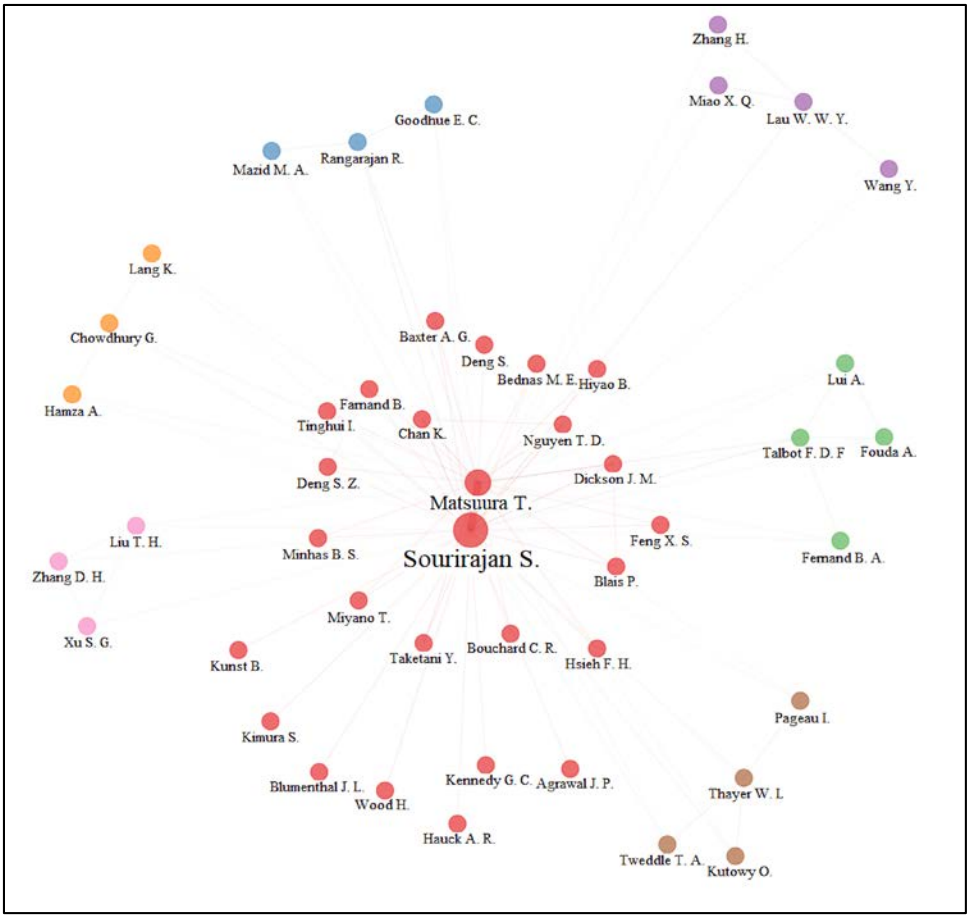


Figure 9. Collaboration network of authors.

It was stated in Table 2 that there were 95 different authors in the dataset. The linkage of 47 researchers out of these 95 scientists who collaborated the most is clearly seen in Fig. 9. Different colors represent different network clusters. Seven different workgroups have been identified in the domain. The main connection of the scientists in these working groups is with Prof. Sourirajan, but there are also other connections between other names. As expected, Prof. Sourirajan is at the center of the collaboration network. Moreover, he is not only the center of the spider web but also the most important element in the largest work group indicated in red and located in this figure. T. Matsuura, the scientist who has collaborated with Prof. Sourirajan the most, has the thickest connection line with him and appears as one of the main elements of the whole network and the

red workgroup. Fig. 9 indicates how many collaborators Prof. Sourirajan was able to work with from all over the world.

The word cloud method was used to visually illustrate the words in the paper's abstracts based on Prof. Sourirajan's portrait (Note that 17 documents did not have abstracts, so this analysis was conducted on 135 documents). The word cloud displays the most frequently occurring words in larger characters, while the smaller the size of the word, the less essential it is. The results can be seen in Fig. 10.

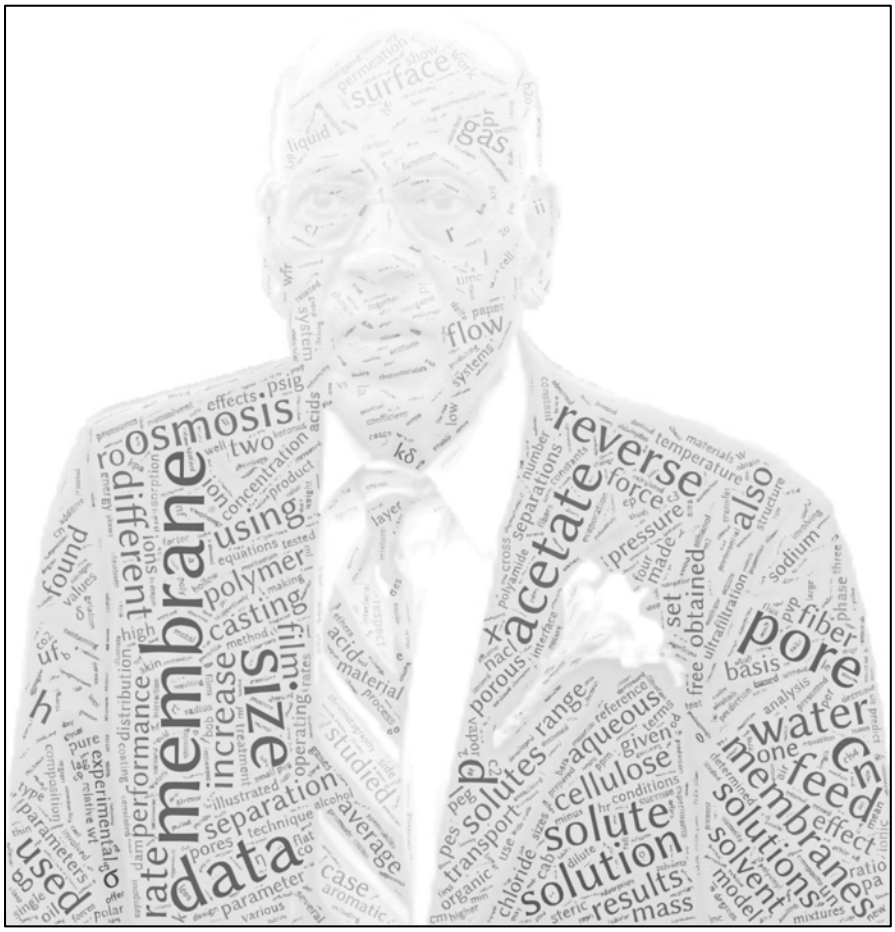


Figure 10. Word cloud of papers abstracts based on Prof. Sourirajan's portrait.

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

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

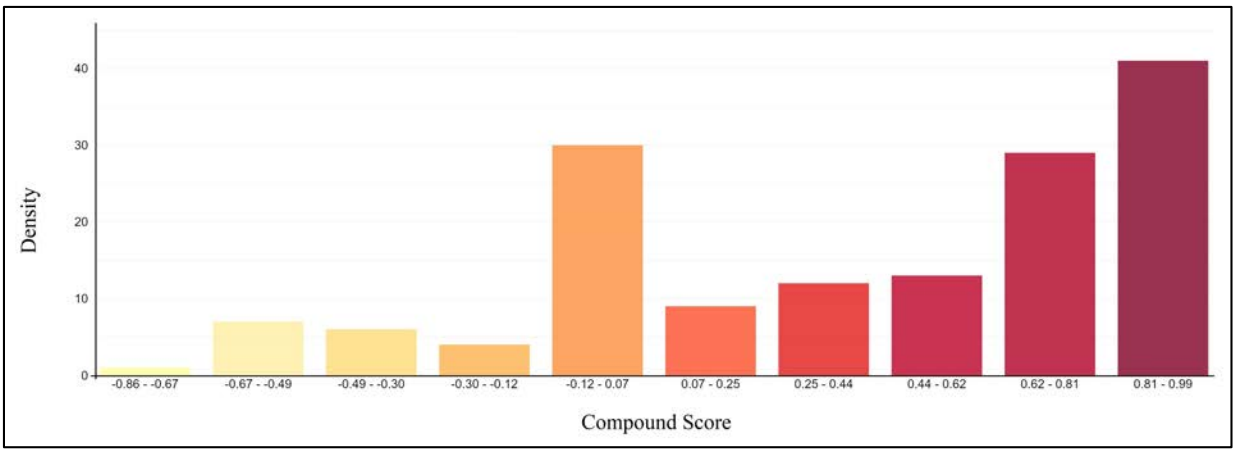


Figure 11. Sentiment analysis results.

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

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

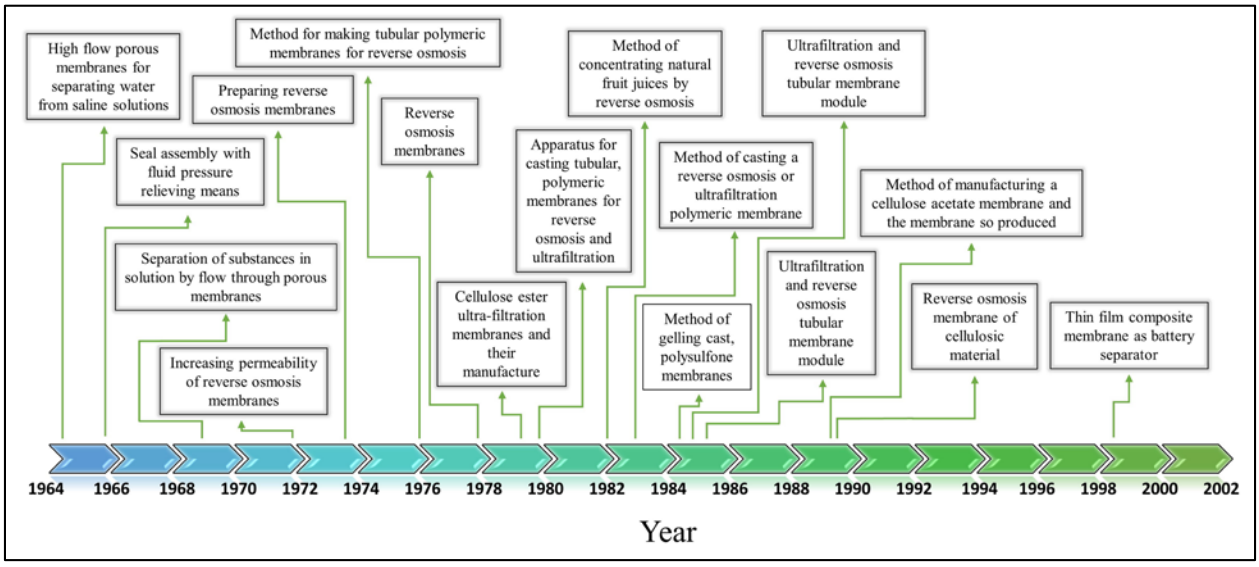
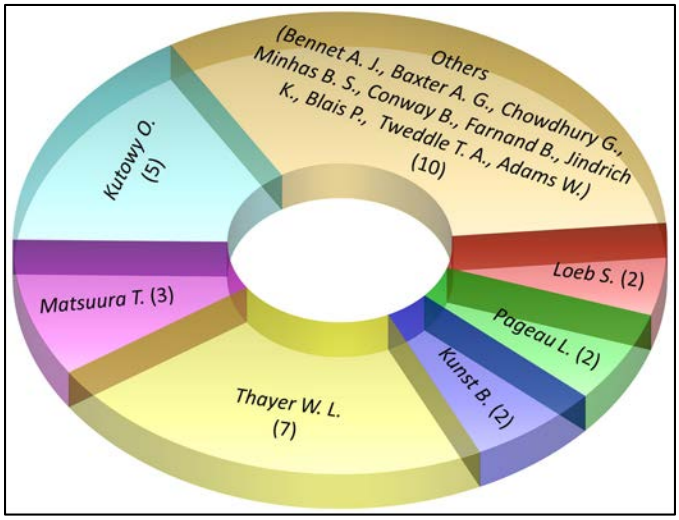


Figure 12. Timeline of Prof. Sourirajan's patents.



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. Kutowy 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.

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

4. Conclusions

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

Acknowledgements

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

References

[1] S. Sourirajan, Reverse Osmosis, Academic Press, 1970.

[2] Water-Desalination-Report, [A Tribute to Srinivasa Sourirajan], Houston, Texas, 2022.

[3] Y. Orooji, Father of reverse osmosis who made a huge impact on our world: Srinivasa Sourirajan (October 16, 1923–February 20, 2022), npj Clean Water, 5 (2022) 20.

[4] A. Husain, A. El-Nashar, A.A. Radif, Membrane-Based Desalination Processes, in: D.M.K. Al-Gobaisi (Ed.) Physical, Chemical and Biological Aspects of Water, UNESCO-EOLSS, 2010, pp. 160.

[5] M.M. Denn, Chemical Engineering: An Introduction, 1st ed., Cambridge University Press, 2011.

[6] J. Kucera, Reverse Osmosis: Industrial Processes and Applications, 2nd ed., John Wiley & Sons Inc, 2015.

[7] V.H. Edwards, S. Shelley, Careers in Chemical and Biomolecular Engineering, 1st ed., CRC Press, 2018.

[8] H.-H. Tseng, W.J. Lau, M.A. Al-Ghouti, L. An, Preface, in: H.-H. Tseng, W.J. Lau, M.A. Al-Ghouti, L. An (Eds.) 60 Years of the Loeb-Sourirajan Membrane, Elsevier, 2022, pp. xxvii-xxviii.

[9] R.W. Baker, Membranes and Modules, in: Membrane Technology and Applications, John Wiley & Sons, Ltd, 2004, pp. 89-160.

[10] S. Sourirajan, Separation of Gases by Flow under Pressure through Porous Membranes, Nature, 199 (1963) 590-591.

[11] S. Sourirajan, Separation of Hydrocarbon Liquids by Flow Under Pressure Through Porous Membranes, Nature, 203 (1964) 1348-1349.

[12] S. Sourirajan, Ottawa Citizen, (2022). https://ottawacitizen.remembering.ca/obituary/srinivasa-sourirajan-1084507224 (accessed August 8, 2022)

[13] Q. Zhan, C. Wu, H. Ding, Y. Huang, Z. Jiang, N. Liao, K. Wang, Y. Li, Emerging trends in photodynamic therapy for head and neck cancer: A 10-year bibliometric analysis based on CiteSpace, Photodiagnosis and Photodynamic Therapy, 38 (2022) 102860.

[14] T. Pang, J. Shen, Visualizing the landscape and evolution of capacitive deionization by scientometric analysis, Desalination, 527 (2022) 115562.

[15] A.P. Mörschbächer, C.E. Granada, Mapping the worldwide knowledge of antimicrobial substances produced by Lactobacillus spp.: A bibliometric analysis, Biochemical Engineering Journal, 180 (2022) 108343.

[16] Y. Fomina, A. Glińska-Neweś, A. Ignasiak-Szulc, Community supported agriculture: Setting the research agenda through a bibliometric analysis, Journal of Rural Studies, 92 (2022) 294-305.

[17] Z. Zheng, S. Zhu, M. Lv, Z. Gu, H. Hu, Harnessing nanotechnology for cardiovascular disease applications - a comprehensive review based on bibliometric analysis, Nano Today, 44 (2022) 101453.

[18] Z.A. Polat, M. Alkan, J. Paulsson, J.M. Paasch, E. Kalogianni, Global scientific production on LADM-based research: A bibliometric analysis from 2012 to 2020, Land Use Policy, 112 (2022) 105847.

[19] J.K. Tamala, E.I. Maramag, K.A. Simeon, J.J. Ignacio, A bibliometric analysis of sustainable oil and gas production research using VOSviewer, Cleaner Engineering and Technology, 7 (2022) 100437.

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

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

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

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

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

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

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

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

[28] E. Aytac, Forecasting Turkey's Hazelnut Export Quantities with Facebook's Prophet Algorithm and Box-Cox Transformation, ADCAIJ-Advances in Distributed Computing and Artificial Intelligence Journal, 10 (2021) 33-47.

[29] E. Aytaç, Exploring Electrocoagulation Through Data Analysis and Text Mining Perspectives, Environmental Engineering & Management Journal (EEMJ), 22 (2022) 671-685.

[30] B. Garner, C. Thornton, A. Luo Pawluk, R. Mora Cortez, W. Johnston, C. Ayala, Utilizing text-mining to explore consumer happiness within tourism destinations, Journal of Business Research, 139 (2022) 1366-1377.

[31] M.A. Sayed, X. Qin, R.J. Kate, D.M. Anisuzzaman, Z. Yu, Identification and analysis of misclassified work-zone crashes using text mining techniques, Accident Analysis & Prevention, 159 (2021) 106211.

[32] G. Liu, M. Boyd, M. Yu, S.Z. Halim, N. Quddus, Identifying causality and contributory factors of pipeline incidents by employing natural language processing and text mining techniques, Process Safety and Environmental Protection, 152 (2021) 37-46.

[33] T. Zhang, N.V. Sahinidis, C.P. Rosé, S. Amaran, B. Shuang, Forty years of Computers and Chemical Engineering: Analysis of the field via text mining techniques, Computers & Chemical Engineering, 129 (2019) 106511.

[34] B. Goodwin, E.S. Wilson, M.M. Oware, H.H. Kestian, T.J. Stigdon, E.A. Miller, Text mining and the examination of language used to report child maltreatment: How language impacts child welfare intake reports, Children and Youth Services Review, 133 (2022) 106344.

[35] R.L. Atenstaedt, Word cloud analysis of historical changes in the subject matter of public health practice in the United Kingdom, Public Health, 197 (2021) 39-41.

[36] M. Bibi, W.A. Abbasi, W. Aziz, S. Khalil, M. Uddin, C. Iwendi, T.R. Gadekallu, A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for twitter sentiment analysis, Pattern Recognition Letters, 158 (2022) 80-86.

[37] J. Liao, M. Wang, X. Chen, S. Wang, K. Zhang, Dynamic commonsense knowledge fused method for Chinese implicit sentiment analysis, Information Processing & Management, 59 (2022) 102934.

[38] M. Aria, C. Cuccurullo, bibliometrix: An R-tool for comprehensive science mapping analysis, Journal of Informetrics, 11 (2017a) 959-975.

[39] Exploratory-Inc., (2020). https://exploratory.io (accessed February 18, 2022)

[40] N. Desai, L. Veras, A. Gosain, Using Bradford's law of scattering to identify the core journals of pediatric surgery, Journal of Surgical Research, 229 (2018) 90-95.

[41] M. Aria, C. Cuccurullo, Biblioshiny : Bibliometrix for no Coders, (2017b). https://www.bibliometrix.org/biblioshiny/assets/player/KeynoteDHTMLPlayer.html#1 (accessed March 30, 2022)

[42] M. Aria, C. Cuccurullo, A brief introduction to bibliometrix, (2017c). https://www.bibliometrix.org/vignettes/Introduction_to_bibliometrix.html (accessed March 30, 2022)

[43] Y. Ding, J. Ma, X. Luo, Applications of natural language processing in construction, Automation in Construction, 136 (2022) 104169.

[44] S. Suardi, A.R. Rasel, B. Liu, On the predictive power of tweet sentiments and attention on bitcoin, International Review of Economics & Finance, 79 (2022) 289-301.

[45] O. Kilicoglu, H. Mehmetcik, Science mapping for radiation shielding research, Radiation 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