

Rough Sets: a Bibliometric Analysis from 2014 to 2018

Ruben Heradio
Dept. of Computer Systems
and Software Engineering,
Universidad Nacional de
Educacion a Distancia,
Madrid, Spain
rheradio@issi.uned.es

David Fernandez-Amoros
Dept. of Computer Systems
and Software Engineering,
Universidad Nacional de
Educacion a Distancia,
Madrid, Spain
david@issi.uned.es

Jose A. Moral-Muñoz
Dept. of Nursing and
Physiotherapy,
University of Cadiz,
Cadiz, Spain
joseantonio.moral@uca.es

Manuel J. Cobo
Dept. of Computer Science
and Engineering,
University of Cadiz,
Cadiz, Spain
manueljesus.cobo@uca.es

Abstract

Along almost forty years, considerable research has been undertaken on rough set theory to deal with vague information. Rough sets have proven to be extremely helpful for a diversity of computer-science problems (e.g., knowledge discovery, computational logic, machine learning, etc.), and numerous application domains (e.g., business economics, telecommunications, neurosciences, etc.). Accordingly, the literature on rough sets has grown without ceasing, and nowadays it is immense. This paper provides a comprehensive overview of the research published for the last five years. To do so, it analyzes 4,038 records retrieved from the Clarivate Web of Science database, identifying (i) the most prolific authors and their collaboration networks, (ii) the countries and organizations that are leading research on rough sets, (iii) the journals that are publishing most papers, (iv) the topics that are being most researched, and (v) the principal application domains.

1. Introduction

In the early 1980s, Pawlak [1, 2, 3] created an innovative approach to deal with inconsistent and imprecise data, named **Rough Set Theory (RST)**. In contrast to other approaches that require enriching the data with additional assumptions or information, such as data probability distributions for *statistical inference* or membership functions for *fuzzy set theory*, RST does not require any extra information: the knowledge contained in the data is the only input RST needs.

According to RST, data are fundamentally structured as *information tables* whose rows represent *data objects*, and whose columns denote *attributes* that describe particular object features and *decisions* made in sight of the attribute values.

From its very inception, RST has faced two critical problems:

- **Reducing data redundancy.** There are different techniques for polishing information tables by reducing both rows and columns without losing information. Objects (rows) are grouped using a type of equivalence relation named *indiscernibility*; attributes (columns) are pruned to find *reducts* that preserve the indiscernibility relation while including a minimum amount of attributes.
- **Providing decision making support from the available data.** Different methods have been proposed to derive *decision rules* from information tables. A rule expresses the decision to be made when a combination of attribute values happens; it has the form $\alpha \Rightarrow d$, where α is a conjunction of attribute values, and d is the corresponding decision. Data may be inconsistent in the sense that distinct decisions can be derived from a single attribute value combination (i.e., $\alpha \Rightarrow d_1 \vee d_2 \vee \dots \vee d_n$). RST deals with inconsistency by managing two approximations of the decision set: (i) a lower approximation of *certainly valid rules*, which is a rule subset where unique decisions are guaranteed for specific attribute value combinations, and (ii) an upper approximation of *possibly valid rules*, where each rule expresses the range of all possible decisions that can be derived from a conflicting attribute value combination.

Interestingly, both problems are intertwined, as the synthesis of minimal decision rules (i.e., with the least terms in α) can be easily derived from minimal reducts [4]. However, finding a minimal reduct is NP-hard [5], and thus much research has been undertaken to solve this problem heuristically [6, 7]. Moreover, the use of RST has spread without ceasing on a variety of computer-science topics (knowledge acquisition, machine learning, decision analysis, knowledge discovery, computational logic,

etc.) and application domains (engineering, business economics, robotics, medicine, etc.).

Clarivate Web of Science (WoS) database includes approximately a total of 12,800 research publications on RST (since 1982 until now). This paper focuses on the last five years (from 2014 to 2018). For this period, WoS includes 4,038 papers. Although a research field could be analyzed manually, when the corpus comprises a large amount of information, it becomes a daunting task. For these situations, intelligence artificial techniques are available to extract useful insights and knowledge. Particularly, *the science of science* [8] is delved to analyze scientific documents to detect trends and patterns, employing bibliometric and science mapping analysis [9, 10, 8]. Thus, the conceptual, intellectual, and social aspects of a research field are automatically detected, analyzed, and visualized. Bibliometric and science mapping techniques have been successfully used to discover the international collaboration [11, 12], to predict future trends in science [13], or generally, to map research fields and journals[14, 15, 16].

In that sense, the main objective of this paper is to provide a complete bibliometric analysis of the RST research field during the last five years (2014-2018). As a result, the following *Research Questions (RQs)* are answered:

- RQ1: How numerous is the literature on RST nowadays? Is the publication trend increasing?
- RQ2: What countries and organizations are leading research on RST?
- RQ3: What journals are publishing most papers?
- RQ4: What are the foremost application domains?
- RQ5: What RST topics are being more researched at present?
- RQ6: Who are the most prolific authors? How do they collaborate?

The remainder of this paper is organized as follows: Section 2 describes the methodology used to perform our bibliometric analysis; Section 3 reports the analysis results and provides some discussion regarding the research questions above; finally, Section 4 gives some concluding remarks.

2. Research method

This section explains the systematic method followed to undertake our bibliometric analysis. In particular, the workflow proposed in [9] has been used, which is composed of three steps: data retrieval, preprocessing, and analysis.

2.1. Data retrieval

The first step to answer this work research questions was obtaining a representative publication sample. It is worth noting that, as other authors have pointed out [17, 18, 19], finding all the articles relevant for a literature review is unrealistic. Accordingly, we targeted the more humble goal of getting a sound publication sample that represents the population adequately.

A sample of 4,038 records was gathered from the *Clarivate WoS* database using the following query:

```
1 TOPIC: ("rough set" or "rough sets")
2 Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH,
3 BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC
4 Timespan=2014-2018
```

Line 1 sets the topic. Lines 2-3 constraint the query scope to the WoS *Core Collection*. Finally, Line 4 sets the time period our analysis comprises.

WoS was selected because it has the highest data quality reputation of all bibliographic databases [20], as it applies a rather selective inclusion procedure in contrast to other databases, such as Google Scholar, which tend to include everything that resembles scholarly work.

2.2. Data preprocessing

The records got from WoS were polished to standardize:

- **Authors' names.** To identify properly the authors' collaboration networks, it was needed to inspect and correct multiple mistaken authors' names. As an example of this kind of errors, in the WoS record of [21], one of the authors, Hamido Fujita, appears with the wrong name Harnido Fujita.
- **Paper keywords.** As it will be explained in Section 2.3, we have used a technique called co-word analysis to recognize the thematic subfields of the RST literature. This technique requires the keywords to be standardized following a de-duplication process. It is worth noting that standardizing keywords is a non-trivial laborious task that needs domain knowledge. For instance, in RST different words, such as *reduct* and *covering*, have the same meaning in most contexts, and thus they usually need be considered as synonyms in the co-word analysis. On the other hand, some words have an excessively broad meaning, and thus they have to be taken out from the analysis; e.g., *algorithm*, *model*, or *rough set theory*.

2.3. Data analysis

To identify the main RST research topics and their inter-relationship, a bibliometric method named

co-word analysis [22] was applied. Co-word analysis measures the association strength between keywords by examining how often every pair of keywords occurs in the articles. To do that, the software tool **SciMAT** [23] was employed, which combines *performance analysis* and *science mapping analysis* for examining a whole research field by representing its conceptual subdomains. SciMAT is based on a methodology [24] that divides the analysis into four stages. Nevertheless, as the last stage is concerned about the conceptual evolution analysis over a set of consecutive time periods and this paper analysis is focused on just one period, the fourth stage was omitted. In summary, the three stages we followed were:

1. *Detection of the research themes.* The corresponding research themes, also called thematic networks, are identified by applying a co-word analysis [25] to the published documents in the RST research field, followed by a clustering of keywords to topics/themes [26], which locates keyword networks that are strongly linked to each other and that correspond to centres of interest or to research problems that are subject of significant interest among researchers. The co-occurrence values in the whole co-word network must be normalized using a similarity measure. From the variety of measures available to determine keyword association strength, the **equivalence index** [27] was chosen because it is the most appropriate one for normalizing co-occurrence frequencies according to the comparative analysis in [28]. The equivalence index $e_{A,B}$ of two keywords A and B is defined as:

$e_{A,B} = \frac{c_{A,B}^2}{c_A \cdot c_B}$; where $c_{A,B}$ stands for the number of papers where A and B appear together; c_A and c_B are the total number of articles that include A and B , respectively. It is worth noting that $e_{A,B}$ ranges from zero to one: (i) when there is no paper where A and B appear together then $c_{A,B} = 0$, and thus $e_{A,B} = 0$; (ii) when A and B always show up together, $c_A = c_B = c_{A,B}$, and thus $e_{A,B} = 1$.

2. *Visualizing research themes and thematic networks.* In this phase, the detected themes are represented using two different visualization instruments: *strategic diagrams* [29] and *thematic networks* [24]. Each theme can be characterized by two measures named *centrality* and *density* [30]. Centrality measures the degree of interaction of a theme with other ones, and it is calculated as:

$$\text{centrality}_{\text{network}} = 10 \cdot \sum_{A \in \text{network}, B \notin \text{network}} e_{A,B}$$

Where A and B are keywords inside and outside the

thematic network, respectively.

On the other hand, density accounts for the network internal coherence by measuring the strength of the links that tie together its keywords as:

$$\text{density}_{\text{network}} = \frac{100}{\#\text{network}} \cdot \sum_{A,B \in \text{network}} e_{A,B}$$

Where A and B are keywords belonging to the network, and $\#\text{network}$ is the number of keywords of the network.

Given both measures, a research field can be visualized as a set of research themes, mapped in a two-dimensional strategic diagram (Figure 1) and classified into four groups:

- Themes in the upper-right quadrant are both well developed and important for the structure of the research field. They are known as the *motor-themes* of the specialty, given that they present strong centrality and high density.
- Themes in the upper-left quadrant have well-developed internal ties but unimportant external ties and so, they are of only marginal importance for the field. These themes are very *specialized and peripheral*.
- Themes in the lower-left quadrant are both weakly developed and marginal. The themes in this quadrant have low density and low centrality, and mainly represent either *emerging or disappearing* themes.
- Themes in the lower-right quadrant are important for the research field but have not been developed enough yet. This quadrant contains *transverse and general/basic* themes.

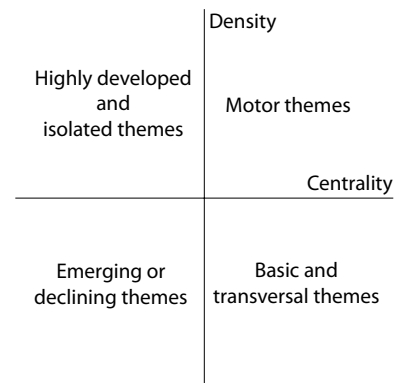


Figure 1. Theme distribution in a strategic diagram.

3. *Performance analysis.* In this phase, each theme relative contribution the whole research field is

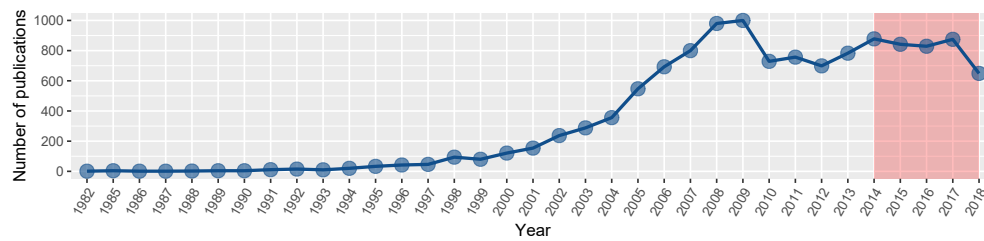


Figure 2. Number of publications per year.

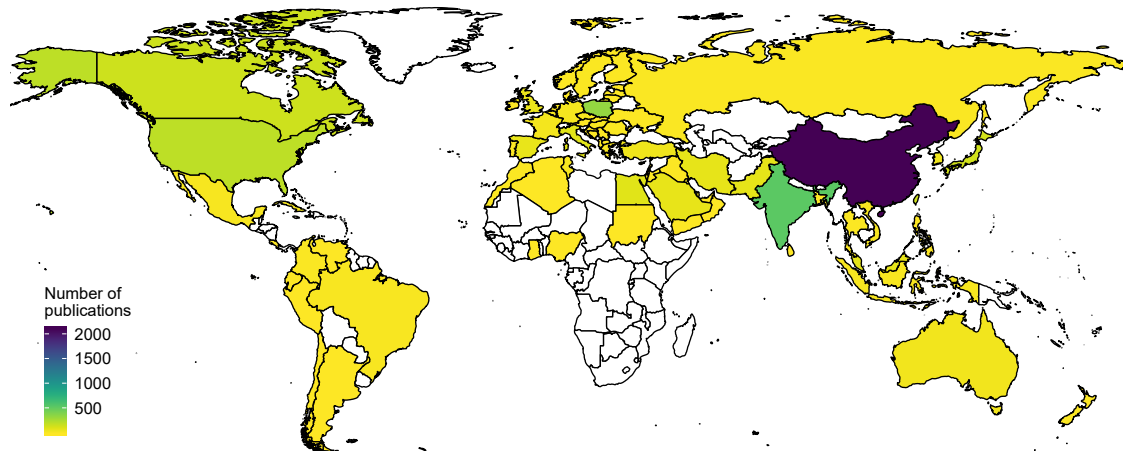


Figure 3. Number of publications per country.

measured (quantitatively and qualitatively), and used to establish the most prominent, most productive, and highest-impact subfields. Some of the bibliometric indicators to use are the number of published documents, the number of citations, and different types of *h-index* [31, 32]. For each theme, the performance measure is computed taking into account the documents associated with it. Thus, for instance, the h-index is computed using the citations of the theme documents.

The social relationships among the RST researchers are also studied in this paper. To do that, we applied the *Leiden algorithm* [33] to detect clusters or communities of researchers strongly linked between them [34]. This algorithm was selected because it has proven to identify the structure of large and complex networks, yielding communities that are guaranteed to be connected, and converging to a partition in which all subsets of every community are locally optimally assigned when the algorithm is applied iteratively.

3. Results and discussion

This section presents and discusses the prime results of our analysis. It is organized according to the research

questions this paper faces.

3.1. RQ1: Article publication trend

Figure 2 shows the article publication progression over time; the red area bounds the years this paper covers. Three periods may be distinguished:

1. An initial phase from 1982 to 1999, where the seminal ideas were proposed and developed by a rather reduced RST community (in eighteen years, there were published only 340 articles).
2. A growth stage from 2000 to 2009, where the publication rate increased one order of magnitude (from one hundred papers in 2000 to one thousand in 2008, approximately).
3. A stabilization period from 2010 to nowadays, where about eight hundred papers are published per year.

3.2. RQ2: Principal countries and organizations

Figure 3 depicts the number of papers published from 2014 to 2018 per country. In particular, the countries whose researchers have published most

literature are China (52.08% of the papers), India (13.17%), Poland (8.32%), United States of America (5.15%), Canada (3.91%), Taiwan (3.79%), and Japan (3.49%). Figure 4 summarizes the fifteen most prolific organizations. According to the results, it is noticeable that most of the research is being performed at Chinese universities.

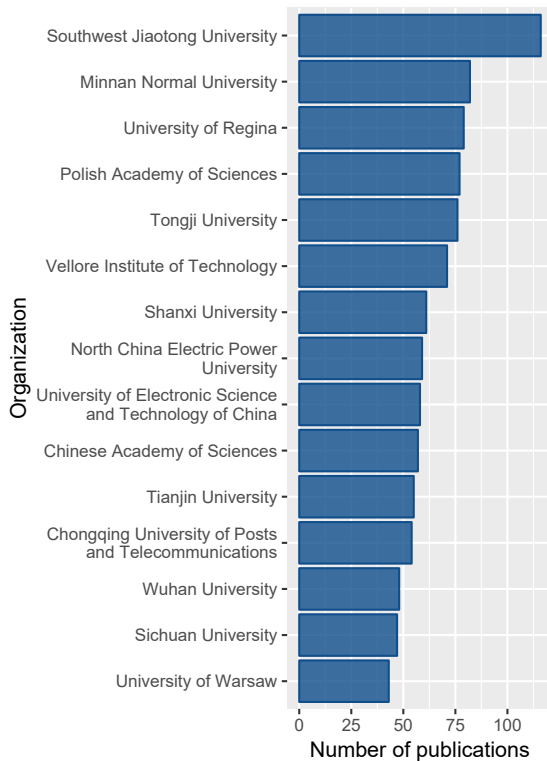


Figure 4. Most prolific research centers.

3.3. RQ3: Most important journals

Figure 5 shows the fifteen journals that have published most articles. Most of them are computer-science journals, encompassing a variety of topics: soft computing, fuzzy systems, intelligent systems, expert systems, neural networks, etc. Also, there are journals devoted to other application domains, such as engineering (e.g., *Mathematical Problems in Engineering*) or life sciences (e.g., *The Scientific World Journal*).

3.4. RQ4: Main application domains

WoS uses a categorization scheme called *Research Areas* that classifies publications according to their subjects into 252 areas. 70.24% of the RST articles fall into the *Computer Science* area; in

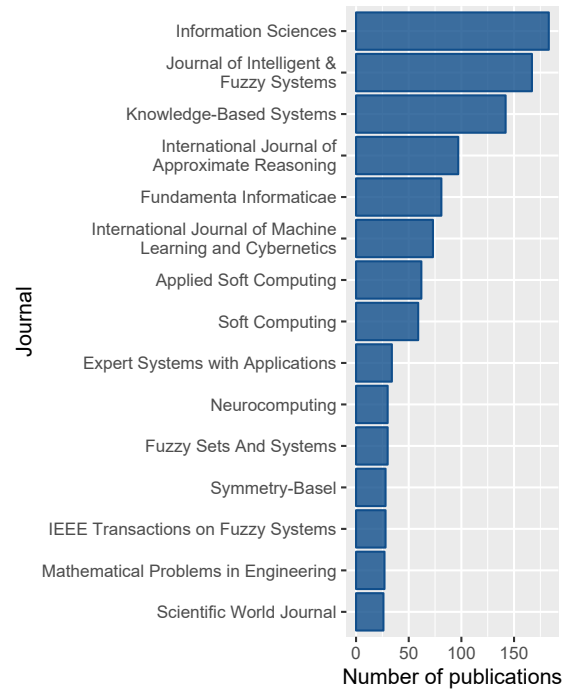


Figure 5. Most prolific journals.

addition, RST literature spreads over other application domains: *Engineering* (24.63%), *Mathematics* (9.23%), *Control Systems* (5.43%), *Operations Research* (4%), *Telecommunications* (2.97%), *Robotics* (2.95%), etc. Figure 6 is a *word cloud* that summarizes most of these other application domains, where the size of the words is proportional to the number of papers classified in the corresponding areas.

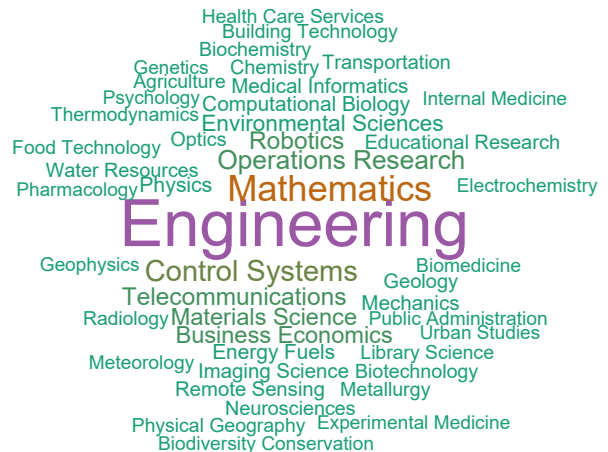


Figure 6. Main application domains.

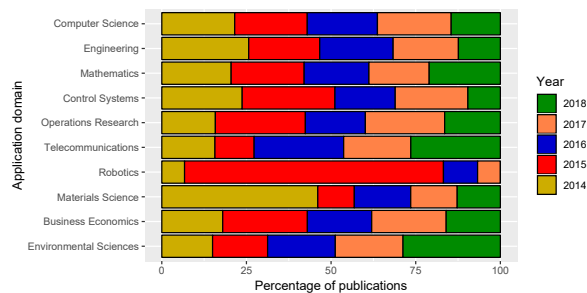


Figure 7. Evolution of the percentage of publications per application domain over time.

Figure 7 shows the proportion of publications per year for the ten most prolific application domains. Whereas in some areas the publication rate is stable (e.g., *Computer Science* or *Mathematics*), in other domains it varies significantly:

- *Robotics*. In 2015, the number of publications on robotics increased from 8 to 91. Then, this number decreased to 12 in 2016 and, finally, no paper was published about robotics in 2018.
- *Materials Science*. In 2014 there were published 47 articles on materials science. Then, paper production decreased to 11. For the last three years, the publication rate has remained stable with a mean of 14.66 articles per year.
- *Environmental Sciences*. The number of papers on environmental sciences has increased continuously from 12 articles in 2014 to 23 in 2018.

3.5. RQ5: Main topics

The strategic diagram in Figure 8 depicts the most important themes of the RST research field for the whole period (2014–2018). These themes have been automatically detected processing the keywords with the techniques described in Section 2.3. The sphere volumes are proportional to the number of published documents associated with each research theme.

According to Figure 8, the RST research field was devoted to 23 themes. Among them, 12 themes could be highlighted due to their position in the strategic diagram (motor and basic-transverse): *neural network* (Figure 10), *attribute reduction* (Figure 9), *updating approximations*, *ideals*, *decision making* (Figure 11), *concept lattices*, *three-way decision*, *uncertainty measure*, *approximation operators*, *granular computing*, *optimization*, and *machine learning*.

As it was discussed in Section 2.3, the detected

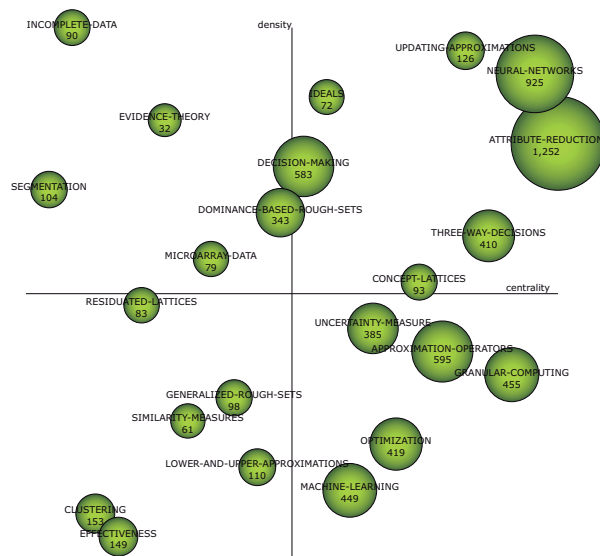


Figure 8. Main research themes.



Figure 9. Research theme “attribute reduction”.

Table 1. Theme performance for the whole period

Name	Number of documents	Number of citations	h-index
ATTRIBUTE-REDUCTION	1,252	9,580	44
NEURAL-NETWORKS	925	5,243	30
APPROXIMATION-OPERATORS	595	3,983	31
DECISION-MAKING	583	4,072	32
GRANULAR-COMPUTING	455	4,147	33
MACHINE-LEARNING	449	1,847	19
OPTIMIZATION	419	2,553	25
THREE-WAY-DECISIONS	410	4,538	36
UNCERTAINTY-MEASURE	385	2,839	29
DOMINANCE-BASED-ROUGH-SETS	343	2,006	23
CLUSTERING	153	682	14
EFFECTIVENESS	149	912	17
UPDATING-APPROXIMATIONS	126	1,848	23
LOWER-AND-UPPER-APPROXIMATIONS	110	545	13
SEGMENTATION	104	488	11
GENERALIZED-ROUGH-SETS	98	458	12
CONCEPT-LATTICES	93	749	15
INCOMPLETE-DATA	90	520	11
RESIDUATED-LATTICES	83	615	14
MICROARRAY-DATA	79	436	11
IDEALS	72	300	8
SIMILARITY-MEASURES	61	521	11
EVIDENCE-THEORY	32	326	9

themes are classified into four different categories in the strategic diagram, according to their density and centrality. Thus, themes located at the top-right quadrant (*attribute-reduction, neural-networks, decision making, etc.*) are motor topics because they show high centrality and density. Accordingly, these themes have significant internal development (wide connection among the keywords that synthesize the themes) and great external cohesion (exceptional connection with other themes).

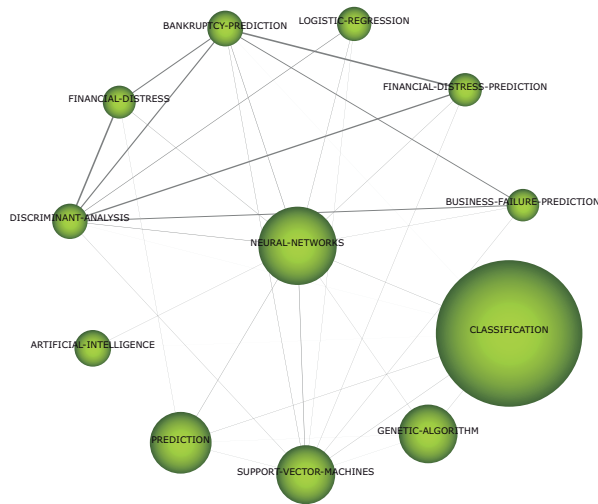


Figure 10. Research theme “neural networks”.



Figure 11. Research theme “decision making”.

On the other hand, themes placed in the bottom-right quadrant have high centrality, but low density, which means that they are considered *auxiliar* for the RST community. In other words, they are applied to solve RST-related problems, but they are not of interest for themselves. That is the case, for example, of *machine learning* and *optimization*.

Finally, themes located on the left have low centrality, which means that they are isolated and not very used in the whole RST field. For example,

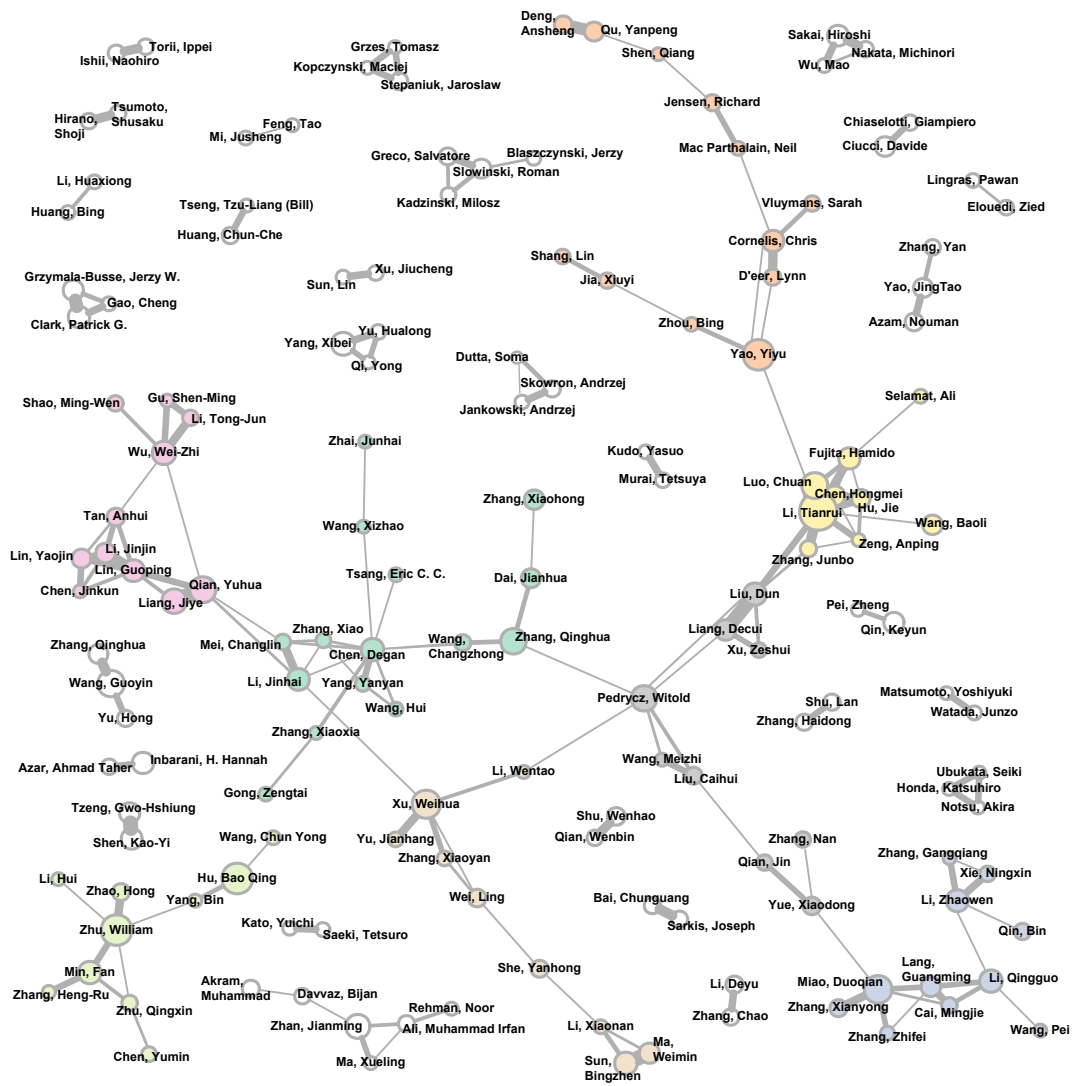


Figure 12. Authors' collaboration networks.

the theme *incomplete-data* has low centrality and high density, so, although it has a tremendous internal development, it is peripheral.

As each theme is associated with a set of documents, it can be enriched with performance and impact measures. In this sense, the performance measures of the detected themes are given in Table 1, showing the number of documents, number of citations, and h-index per theme. Accordingly to these performance measures, the theme attribute *reduction* stands out due to the high citations achieved (9,580), which almost doubles the second one. Moreover, we should remark other themes that achieved more than 1,000 citations, such as, *neural network*, *three-way decisions*, *granular computing*, *decision making*, *approximation operators*, *uncertainty measure*, *optimization*, *dominance-based*

rough sets, *updating approximations*, and *machine learning*.

3.6. RQ6: Principal researchers

A total of 6,564 researches have co-authored the 4,038 articles this paper analyzes. Two kinds of contributors can be distinguished: (i) a vast majority of casual authors (the *median* number of papers per author for the whole period 2014-2018 is only 1); and (ii) a reduced group of authors (only 4.89% of them) whose research is much more focused on RST and thus they have published at least five papers. The graph in Figure 12 summarizes this latter core of authors. Each node represents a researcher, being its size proportional to the number of articles he/she has authored. Nodes are

connected by an edge when the corresponding authors have written together at least two papers; edges become thicker as the number of co-authored papers grows. The Leiden Algorithm was used to detect authors' collaboration networks. In the figure, each identified network has a different color.

4. Concluding remarks

After nearly forty years of intense research, RST has developed a formidable maturity level, being successfully applied not only to specific computer-science problems, such as machine learning or computational logic, but to a vast range of application domains. According to WoS, RST has been used in about one hundred areas, including telecommunications, mathematics, neurosciences, robotics, materials science, environmental sciences, remote sensing, radiology, chemistry, business economics, etc.

Especially since the change of the century, the production of scientific papers on RST has increased dramatically. As reported by WoS, in just eight years, the production passed from one hundred articles in 2000 to one thousand in 2008, approximately. As a result, more than twelve thousand papers have been published since 2000. This paper has put into perspective such an enormous amount of publications by fulfilling its primary objective, i.e., to answer the following six research questions: (RQ1) The publication rate for the last five years is stable, after a growth period from 1997 to 2008; (RQ2) Chinese research centers are the most prolific ones, having published half the papers; (RQ3) Most papers have been published in four journals (Information Sciences, Journal of Intelligent & Fuzzy Systems, Knowledge-Based Systems, and International Journal of Approximate Reasoning); (RQ5) At the moment, the most researched themes are the *attribute reduction* problem stated in this paper introduction, and the usage of *neural networks*; and (RQ6) The foremost authors' collaboration networks have been identified.

Moreover, our work identifies opportunities for future research. First, themes in the lower-right quadrant in Figure 8 are important but insufficiently developed: *machine learning*, *uncertainty measure*, *approximation operators*, *optimization*, and *granular-computing*. Second, RST has begun to be applied to a variety of new domains: *genetics*, *food technology*, *geology*, etc. (see Figure 6). These domains might open promising research directions.

Acknowledgements

This work has been supported by the Spanish Ministry of Science, Innovation and Universities under Project DPI2016-77677-P, FEDER funds under grant TIN2016-75850-R, and the Community of Madrid under Grant RoboCity2030-DIH-CM P2018/NMT-4331.

References

- [1] Z. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.
- [2] Z. Pawlak, "Rough classification," *International Journal of Man-Machine Studies*, vol. 20, no. 5, pp. 469 – 483, 1984.
- [3] W. Marek and Z. Pawlak, "Rough sets and information systems," *Fundamenta Informaticae*, vol. 7, no. 1, pp. 105–115, 1984.
- [4] J. Bazan, *Rough Sets in Knowledge Discovery 1: Methodology and Applications*, ch. a Comparison of Dynamic and non-Dynamic Rough Set Methods for Extracting Laws from Decision Tables, pp. 321–365. Physica-Verlag Heidelberg, 1998.
- [5] A. Skowron and C. R. C., *Intelligent Decision Support. Theory and Decision Library (Series D: System Theory, Knowledge Engineering and Problem Solving)*, vol. 11, ch. the Discernibility Matrices and Functions in Information Systems, pp. 331–362. Springer, 1992.
- [6] L. Yong, H. Wenliang, J. Yunliang, and Z. Zhiyong, "Quick attribute reduct algorithm for neighborhood rough set model," *Information Sciences*, vol. 271, pp. 65 – 81, 2014.
- [7] X. Jia, L. Shang, B. Zhou, and Y. Yao, "Generalized attribute reduct in rough set theory," *Knowledge-Based Systems*, vol. 91, pp. 204 – 218, 2016.
- [8] S. Fortunato, C. T. Bergstrom, K. Börner, J. A. Evans, D. Helbing, S. Milojević, A. M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, A. Vespignani, L. Waltman, D. Wang, and A.-L. Barabási, "Science of science," *Science*, vol. 359, p. eaao0185, 2018.
- [9] M. J. Cobo, A. Lopez-Herrera, E. Herrera-Viedma, and F. Herrera, "Science mapping software tools: Review, analysis, and cooperative study among tools," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 7, pp. 1382–1402, 2011.
- [10] R. Heradio, H. Perez-Morago, D. Fernandez-Amoros, F. J. Cabrerizo, and E. Herrera-Viedma, "A bibliometric analysis of 20 years of research on software product lines," *Information and Software Technology*, vol. 72, pp. 1 – 15, 2016.
- [11] C. R. Sugimoto, N. Robinson-Garcia, D. S. Murray, A. Yegros-Yegros, R. Costas, and V. Larivière, "Scientists have most impact when they're free to move," *Nature*, vol. 550, no. 7674, pp. 29–31, 2017.
- [12] Z. Chinchilla-Rodríguez, C. R. Sugimoto, and V. Larivière, "Follow the leader: On the relationship between leadership and scholarly impact in international collaborations," *PLOS ONE*, vol. 14, no. 6, p. e0218309, 2019.

- [13] Y. Ma and B. Uzzi, "Scientific prize network predicts who pushes the boundaries of science," *Proceedings of the National Academy of Sciences*, vol. 115, no. 50, pp. 12608–12615, 2018.
- [14] M. J. Cobo, M. A. Martínez, M. Gutiérrez-Salcedo, H. Fujita, and E. Herrera-Viedma, "25 years at Knowledge-Based Systems: A bibliometric analysis," *Knowledge-Based Systems*, vol. 80, pp. 3–13, 2015.
- [15] J. Moral-Muñoz, L. Carballo-Costa, E. Herrera-Viedma, and M. Cobo, "Production Trends, Collaboration, and Main Topics of the Integrative and Complementary Oncology Research Area: A Bibliometric Analysis," *Integrative Cancer Therapies*, vol. 18, 2019.
- [16] J. López-Robles, J. Otegi-Olaso, I. Porto Gómez, and M. Cobo, "30 years of intelligence models in management and business: A bibliometric review," *International Journal of Information Management*, vol. 48, 2019.
- [17] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," tech. rep., Keele University and University of Durham, UK, 2007.
- [18] C. Wohlin, P. Runeson, P. A. da Mota Silveira Neto, E. Engström, I. do Carmo Machado, and E. S. de Almeida, "On the reliability of mapping studies in software engineering," *Journal of Systems and Software*, vol. 86, no. 10, pp. 2594–2610, 2013.
- [19] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Information and Software Technology*, vol. 64, no. Supplement C, pp. 1–18, 2015.
- [20] C. R. Sugimoto and V. Lariviere, *Measuring Research: What Everyone Needs to Know*. Oxford University Press, 2018.
- [21] J. Hu, T. Li, C. Luo, H. Fujita, and Y. Yang, "Incremental fuzzy cluster ensemble learning based on rough set theory," *Knowledge-Based Systems*, vol. 132, pp. 144–155, 2017.
- [22] N. Coulter, I. Monarch, and S. Konda, "Software engineering as seen through its research literature: A study in co-word analysis," *Journal of the American Society for Information Science*, no. 13, pp. 1206–1223, 1998.
- [23] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "SciMAT: A new science mapping analysis software tool," *Journal of the American Society for Information Science and Technology*, vol. 63, no. 8, pp. 1609–1630, 2012.
- [24] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field," *Journal of Informetrics*, vol. 5, no. 1, pp. 146–166, 2011.
- [25] M. Callon, J. P. Courtial, W. A. Turner, and S. Bauin, "From translations to problematic networks: An introduction to co-word analysis," *Social Science Information*, vol. 22, no. 2, pp. 191–235, 1983.
- [26] N. Coulter, I. Monarch, and S. Konda, "Software engineering as seen through its research literature: A study in coword analysis," *Journal of the American Society for Information Science*, vol. 49, no. 13, pp. 1206–1223, 1998.
- [27] M. Callon, J. Courtial, and F. Laville, "Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry," *Scientometrics*, no. 1, pp. 155–205, 1991.
- [28] N. J. van Eck and L. L. Waltman, "How to normalize cooccurrence data? an analysis of some well-known similarity measures," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 8, pp. 1635–1651, 2009.
- [29] Q. He, "Knowledge Discovery Through Co-Word Analysis," *Library Trends*, vol. 48, no. 1, pp. 133–159, 1999.
- [30] M. Callon, J. P. Courtial, and F. Laville, "Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry," *Scientometrics*, vol. 22, no. 1, pp. 155–205, 1991.
- [31] S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, and F. Herrera, "h-Index: A review focused in its variants, computation and standardization for different scientific fields," *Journal of Informetrics*, vol. 3, no. 4, pp. 273–289, 2009.
- [32] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National Academy of Sciences*, vol. 102, no. 46, pp. 16569–16572, 2005.
- [33] V. A. Traag, L. Waltman, and N. J. van Eck, "From Louvain to Leiden: guaranteeing well-connected communities," *Nature. Scientific Reports*, vol. 9, no. 1, pp. 1–12, 2019.
- [34] V. Kandylas, S. P. Upham, and L. H. Ungar, "Analyzing knowledge communities using foreground and background clusters," *ACM Transactions on Knowledge Discovery from Data*, vol. 4, no. 2, pp. 7:1–7:35, 2010.