

ISSN 2096-157X

CN 10-1394/G2

JDIS

数据与情报科学学报

JOURNAL OF DATA AND INFORMATION SCIENCE (QUARTERLY)

| Volume 1 Number 4 2016

National Science Library,
Chinese Academy of Sciences



SUBMISSION GUIDELINES

AIMS & SCOPE

JDIS devotes itself to the study and application of the theories, methods, techniques, services, infrastructural facilities using big data to support knowledge discovery for decision & policy making. The basic emphasis is big data-based, analytics centered, knowledge discovery driven, and decision making supporting. The special effort is on the knowledge discovery to detect and predict structures, trends, behaviors, relations, evolutions and disruptions in research, innovation, business, politics, security, media and communications, and social development, where the big data may include metadata or full content data, text or non-textual data, structured or non-structured data, domain specific or cross-domain data, dynamic or interactive data.

REFEREEING PROCESS

Articles covering the topics or themes mentioned above will be refereed through a double-blind peer review process.

MANUSCRIPTS CATEGORIES

Research Articles represent original research work or a comprehensive and in-depth analysis of a topic. More than 3,000 words are considered as a proper length for such manuscripts, with a structured abstract ca. 200 words.

Application Articles cover the latest development and application in any segment of data analysis and information science field work. The length of the manuscript is preferred to be more than 3,000 words, with a structured abstract ca 200 words.

Review Articles summarize the status of knowledge and outline future directions of research within the scope of the journal.

Perspectives are forward-looking viewpoints that advocate important future directions in the theory or application of data and information science.

MANUSCRIPTS REQUIREMENTS

All papers are submitted in English with a double-line space. For the assurance that all the materials of the to-be submitted are included, please check the following:

Title page. The title, author list, affiliations, and author contribution statement should all be included on a title page as the first page of the manuscript file.

Structured abstract & Keywords. A structured abstract is required for all research papers in order for readers to acquire the key points of a manuscript.

References. Be sure all the references used should be cited properly in both in text and in bibliography. For the detailed information, please request a copy of Reference Citation Format.

COPYRIGHT

All submitted papers only normally should not have been previously published nor be currently under consideration for publication elsewhere. For all the materials translated or obtained from other published resources, they should be properly acknowledged. All copyright problems should be cleared without any legal entanglements prior to the publication.

NOTES FOR INTENDING SUBMISSIONS

A guide for authors and other relevant information, including submitting papers online, is available at <http://www.jdis.org>. For any questions, you can e-mail the Office at jdis@mail.las.ac.cn.

The JDIS Editorial Office

National Science Library, Chinese Academy of Sciences

No 33, Beisihuan Xilu, Zhongguancun, Haidian District, Beijing 100190, P.R. China

Website: <http://www.jdis.org>

Mapping Diversity of Publication Patterns in the Social Sciences and Humanities: An Approach Making Use of Fuzzy Cluster Analysis

Frederik T. Verleysen^{1†} & Arie Weeren²

¹Centre for Research & Development Monitoring (ECOOM), Faculty of Social Sciences, University of Antwerp, Middelheimlaan 1, 2020 Antwerp, Belgium

²Mathworks, Postbus 7921, 5605 SH Eindhoven, The Netherlands

Citation: Frederik T. Verleysen & Arie Weeren (2016). Mapping Diversity of Publication Patterns in the Social Sciences and Humanities: An Approach Making Use of Fuzzy Cluster Analysis.

Received: Aug. 2, 2016

Revised: Aug. 20, 2016

Accepted: Aug. 24, 2016

Abstract

Purpose: To present a method for systematically mapping diversity of publication patterns at the author level in the social sciences and humanities in terms of publication type, publication language and co-authorship.

Design/methodology/approach: In a follow-up to the hard partitioning clustering by Verleysen and Weeren in 2016, we now propose the complementary use of fuzzy cluster analysis, making use of a membership coefficient to study gradual differences between publication styles among authors within a scholarly discipline. The analysis of the probability density function of the membership coefficient allows to assess the distribution of publication styles within and between disciplines.

Findings: As an illustration we analyze 1,828 productive authors affiliated in Flanders, Belgium. Whereas a hard partitioning previously identified two broad publication styles, an international one vs. a domestic one, fuzzy analysis now shows gradual differences among authors. Internal diversity also varies across disciplines and can be explained by researchers' specialization and dissemination strategies.

Research limitations: The dataset used is limited to one country for the years 2000–2011; a cognitive classification of authors may yield a different result from the affiliation-based classification used here.

Practical implications: Our method is applicable to other bibliometric and research evaluation contexts, especially for the social sciences and humanities in non-Anglophone countries.

Originality/value: The method proposed is a novel application of cluster analysis to the field of bibliometrics. Applied to publication patterns at the author level in the social sciences and humanities, for the first time it systematically documents intra-disciplinary diversity.



JDIS
Journal of Data and
Information Science
Vol. 1 No. 4, 2016
pp 33–59

DOI: 10.20309/jdis.201624

<http://www.jdis.org>

[†] Corresponding author: Frederik T. Verleysen (E-mail: Frederik.Verleysen@uantwerpen.be).

Keywords Bibliometrics; Social sciences and humanities; Publication patterns; Dissemination; Cluster analysis

1 Introduction

It is a well-known fact that publication patterns in the social sciences and humanities (SSH) differ considerably from those observed in scientific, technical, and biomedical fields. SSH scholars publish their research in a much wider array of both international and domestic publication channels, including monographs and edited books; they frequently opt for other publication languages besides English; and their rate of research collaboration and ensuing co-authorship is considerably lower (Hicks, 2004; Nederhof, 2006). In recent years, bibliometric studies of the SSH have started to devote more attention to the topic of internal diversity. This has mostly been demonstrated by analyses at the disciplinary level, showing that an inter-disciplinary variety in terms of publication patterns exists across the spectrum of the SSH. One particular pattern described is that of a divide between most disciplines belonging to the social sciences and those classified as humanities. In the social sciences the use of international journals, English as a publication language and more frequent co-authorship are appearing to become predominant, while by contrast in the humanities books and chapters and the use of national or regional languages retain a central position, and co-authorship occurs less frequently (Engels, Ossenblok, & Spruyt, 2012; Ossenblok, 2016; Puuska, 2014; Sivertsen, 2009).

By contrast, the *intra*-disciplinary diversity of publication patterns, the variety within disciplines belonging to the SSH, has not received as much attention. A handful of recent studies have documented such diversity in terms of publication and citation patterns (Chi, 2015; Nederhof, 2011) or in that of cognitive structure, mostly of a single discipline (Lin & Kaid, 2000; Persson, 2015). In an effort to document intra-disciplinary diversity in a more systematic way, Verleysen and Weeren (2016) have performed a hard partitioning cluster analysis on the publication patterns of 1,828 individual authors belonging to 16 SSH disciplines and affiliated with the five universities in Flanders, Belgium. This analysis at the author level has demonstrated that intra-disciplinary diversity as regards publication patterns in Flanders is considerable, as well as that it is too simplistic to oppose the publication cultures of the social sciences to those in the humanities. For SSH scholars in Flanders two broad publication styles were identified: the first one is centered around co-authored English-language journal articles in high-profile outlets indexed by the Web of Science; the second one is far more reliant on single-authored articles in national journals and books in other languages than English, especially Dutch, the dominant language in Flanders (Verleysen & Weeren, 2016).



In the present article, we refine our previous results and propose additional steps for a method for the study of diversity of publication patterns in the social sciences and humanities.

2 Data and Method

This paper builds upon the data, method, and results of the cluster analysis by Verleysen and Weeren (2016) of the 1,828 most productive scholarly authors (ten or more weighted peer reviewed outputs during 2000–2011) registered in the Flemish Bibliographic Database for the Social Sciences and Humanities (or VABB-SHW).

The VABB-SHW is a comprehensive regional bibliographic database (i.e. not a citation index) used for calculating a share of the research funding provided by the government to the five Flemish universities. In this capacity the VABB-SHW registers five publication types: journal articles, monographs, edited books, book chapters, and proceedings papers. For inclusion in the Flemish funding model, a weight is attributed to each publication type: journal articles, edited books and book chapters all receive a weight of 1, whereas monographs have a weight of 4 and proceedings papers one of 0.5. Two parts comprise the VABB-SHW. The first, VABB-WoS, consists of records of publications (journal articles and proceedings papers) which are also indexed in a journal and/or proceedings index of the Web of Science (WoS). VABB-WoS consists of ca. 95% of English language publications, and concentrates most of the high-profile international journals in the SSH. The second part, VABB-GP, consists of records of publications which have additionally been identified as peer reviewed by the Authoritative Panel (Gezaghebbend Panel or GP), an independent scientific board of university professors, from the whole of the five universities' non-WoS publications. VABB-GP consists for ca. 70% of publications in other languages than English, especially Dutch (Engels, Ossenblok & Spruyt, 2012).

As input for the hard partitioning cluster analysis (Verleysen & Weeren, 2016), a dataset was compiled listing the 1,828 author names, their main disciplinary affiliation, as well as 11 variables mapping author output during 2000–2011. These variables belong to three groups of attributes which are known to differentiate SSH publication patterns at the disciplinary level: publication type, publication language, and the share of co-authored publications. For the three VABB-SHW book publication types, combined with two publication language groups (English *vs.* other languages), this resulted in a subtotal of six variables, for each of which the fractional contribution to individual authors' total 12-year weighted output was calculated. For journal articles and proceedings papers, fractions were calculated based on the distinction between VABB-WoS and VABB-GP, resulting in a subtotal of four additional variables. The 11th variable is the fraction of weighted co-authored publications.



Mahalanobis distance or ‘generalized squared interpoint distance’ (Mahalanobis, 1936; Wicklin, 2015) was used to calculate dissimilarities between all possible pairs of the 1,828 authors. By means of *silhouettes* and the *k*-medoids clustering algorithm (Kaufman & Rousseeuw, 1990), we arrived at a ‘hard’ or ‘crisp’ clustering result of two distinct clusters of publication patterns within the Flemish publication data. As *k*-medoids identifies the most representative object in every cluster, i.e. the object for which the distance to every other object within its cluster is minimal, a simple retrieval of the original input data for both medoids (i.e. the most representative authors) allowed to label the two clusters—that is: to describe in qualitative terms their underlying publication pattern. We were thus able to identify two distinct styles for publishing research results in Flemish SSH, both of which were found to be present in all 16 disciplines, thereby cutting across the distinction between the social sciences and the humanities (cf. Sections 1, 3). For further elaboration on the method used for hard partitioning we refer to Verleysen and Weeren (2016) and Kaufman and Rousseeuw (1990).

For the present paper we take the analysis a step further by performing a fuzzy cluster analysis on the prior two-cluster result. Whereas the initial hard partitioning attributes all cases to just one of the (here: two) clusters, fuzzy clustering allows for some ambiguity in the data by calculating for each case a membership coefficient, or the degree of belonging of individual authors to each of both clusters (Kaufman & Rousseeuw, 1990). By including this additional information, binary decisions on cluster membership are avoided and the resulting picture of scholarly publication patterns should be more nuanced than the initial result. The fuzzy clustering algorithm used is *Fanny* by Kaufman and Rousseeuw (1990), implemented in MATLAB®R2016a. After applying the fuzzy principle to the two-cluster result by Verleysen and Weeren (2016), we present cluster plots for each individual discipline to illustrate diversity in publication patterns within and across disciplines. In a final step, histograms are used to illustrate the distribution of publication styles for each discipline, by showing the probability density function (Johnson & Wichern, 1992), i.e. a scaling of the original author frequencies denoting the relative probability of a cluster membership coefficient value for a random author in a given discipline. As the disciplines in our analysis are different in size, recalculating absolute author frequencies into the probability density function allows for easier comparison between disciplines.

3 Results

Figure 1 presents the fuzzy principle applied to the two-cluster result on the 1,828 authors of Verleysen and Weeren (2016) (one author = one dot). The two cluster



cores identified by the previous hard partitioning are still easily visible in the new result: *Cluster One* (bright red) and *Cluster Two* (bright green). By identification of the respective most representative authors (medoids) for both clusters, we previously determined (Verleysen & Weeren, 2016) that the core of Cluster One ‘groups those authors who mainly target an international audience of specialized academia, through the collaborative publication of mainly English language articles in high-profile journals indexed by the WoS’. Inversely, the core of Cluster Two contains ‘those authors who are more strongly oriented towards national journals and also book publications, make frequent use of other languages than English (i.e. mainly Dutch), and are much less inclined to co-author publications (Verleysen & Weeren, 2016).

In Figure 1, the result of the fuzzy algorithm Fanny, the belonging of individual authors to the two clusters is now visualized by various shades of red and green. This degree of fuzziness of the result is also expressed by the normalized version of Dunn’s partition coefficient (Dunn, 1976), which on a 0-1 scale gives an indication of how hard or fuzzy the clustering result is. A value of 0 would denote that each object (author) has equal membership in each cluster, or that the result is entirely fuzzy; a value of 1 would mean that each object has a membership of 1 in one cluster and a membership of 0 in the other cluster, or that the result is entirely hard. For the clustering of the 1,828 authors by means of the Fanny algorithm, Dunn’s normalized partition coefficient has a value of 0.2390, demonstrating the appropriateness of the fuzzy approach.

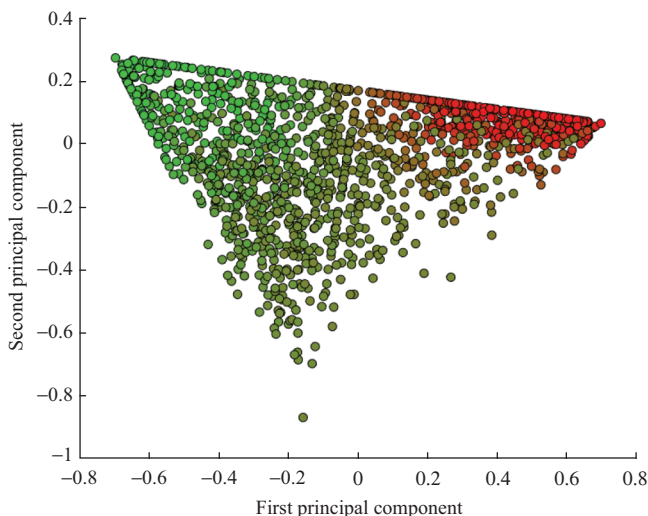


Figure 1. Fuzzy clustering of 1,828 productive authors in the SSH ($n = 1,828$).



The second part of this Section presents the clustering plots for authors belonging to two examples of individual SSH disciplines, Sociology (social sciences) and Linguistics (humanities). Plots and histograms for all other disciplines can be found in the Appendix. The histograms show the probability density function (y -axis) for the cluster membership coefficient of all authors affiliated with a discipline on a 0-1 scale (x -axis), whereby a value of 0 of the coefficient denotes a 100% membership of Cluster Two (green) and a value of 1 a 100% membership of Cluster One (red). For the comparison between disciplines of the histograms we note the existence of a scale variation on the y -axis (probability density function). This results from the considerable variety between disciplines regarding the concentration of cluster membership coefficient values, as the more strongly concentrated within certain ranges of the coefficient the outcomes for individual authors are, the higher the probability of a random author's value occurring within this same range. As in this paper we focus primarily on intra-disciplinary diversity, for which maximum legibility of the individual histograms is required, we opted not to standardize the scale of the probability density function. On the x -axis (membership coefficient), the bin width of the histograms is the software default calculated by the varying number of cases in each discipline, mentioned in the figure legends.

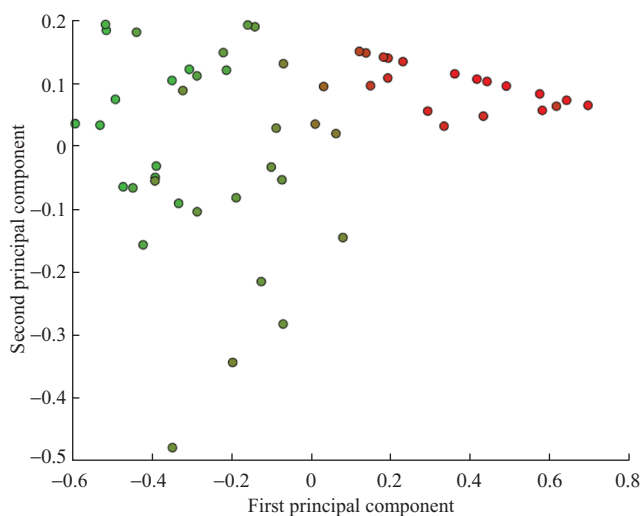


Figure 2. Fuzzy clustering of productive authors in Sociology ($n = 57$).

For both Sociology and Linguistics, the cluster plots and histograms (Figures 2–5) show that intra-disciplinary diversity of publication patterns occurs across a wide spectrum. While linguists show by far the strongest presence on the fringes of

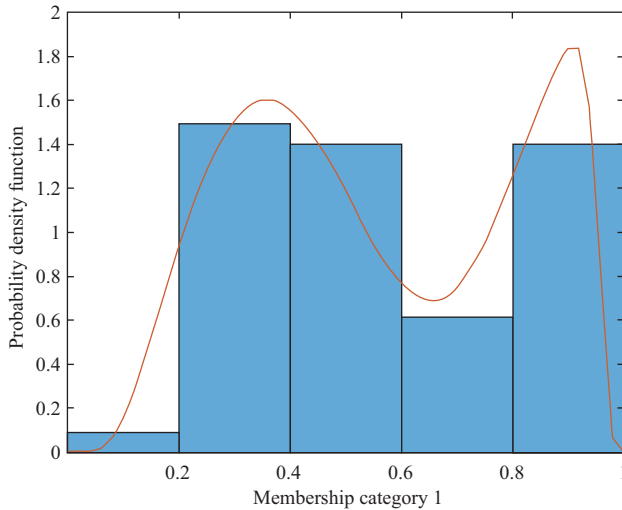


Figure 3. Probability density function of cluster membership coefficients for Sociology ($n = 57$).

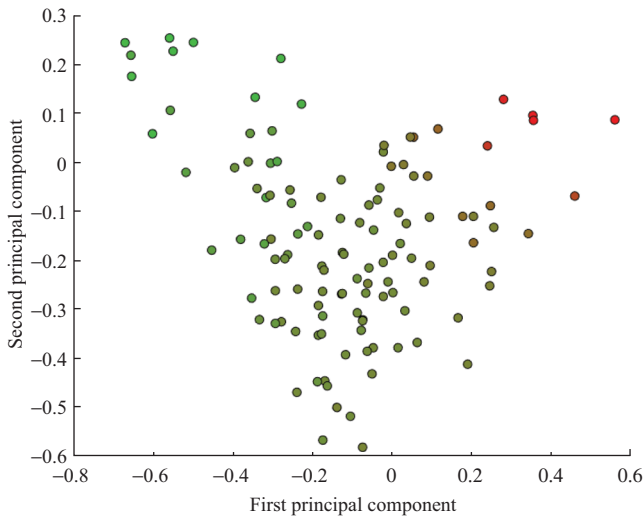


Figure 4. Fuzzy clustering of productive authors in Linguistics ($n = 121$).

Cluster Two (dark green), a limited number of them clearly belong to Cluster One (red), with an equally modest number of authors (brown) occupying the middle ground of Cluster One. In the histogram this predominance of Cluster Two is confirmed by the value of the probability density function for the membership coefficient range of 0.4–0.5, which, though near the center between both clusters,



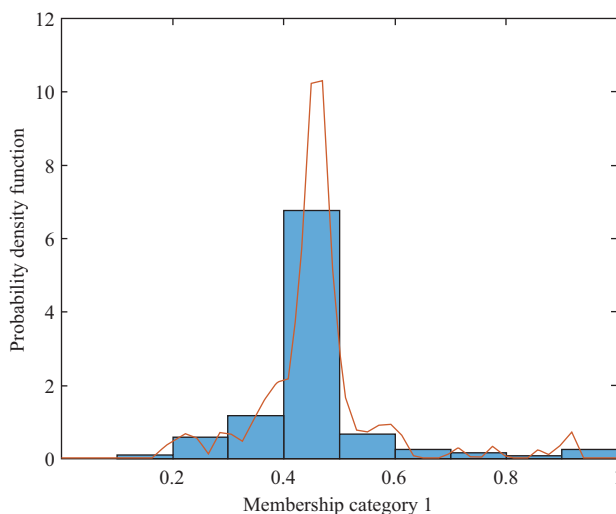


Figure 5. Probability density function of cluster membership coefficients for Linguistics ($n = 121$).

is still closer to that of Cluster Two. For sociologists the divide between publication patterns within the discipline is more profound. Bright green and bright red dots (authors) are dominant, with relatively fewer authors occupying the middle ground (dark green and brown). The probability density function confirms this outspoken divide between publication styles within the discipline.

4 Discussion

Cluster analysis based on bibliographic data for individual researchers reveals how publication patterns can differ widely between authors affiliated with the same discipline. It also demonstrates how publication patterns of social scientists cannot simplistically be opposed to those of humanities scholars. At the same time, there remain considerable differences between the Flemish SSH disciplines used as an example here. Several of the humanities such as Art History, History, Law, Literature, and Theology show a concentration of researchers who publish most often in national journals and books, make use of other languages besides English, and who frequently publish on their own. Other humanities such as Archeology, Communication Studies, Linguistics, and Philosophy show a more dispersed pattern, with a number of their researchers clearly adhering to the other publication model reliant on international journals and English as publication language. In the social sciences, the international journal model is dominant in Psychology and Social Health Sciences, whereas Economics, Educational Sciences, and Sociology show a



dispersed pattern across a broad spectrum of publication styles. Both Criminology and Political Sciences appear to be mostly similar to the humanities with a concentration of authors working in the national-journals-and-books model.

In general, any explanation of inter- and intra-disciplinary heterogeneity of publication patterns in the SSH should point to the intrinsic diversity of many aspects of scholarly research and information dissemination. Most humanities and social sciences are deeply fragmented with regard to intellectual interest and approach, conceptions of standards, as well as target audience (Hicks, 2004; Whitley, 2000). Specialization also relates to methodological differences, and these as well have an impact on the way in which scholarly work is published. In strongly quantitative fields of research, collaboration and ensuing co-authorship for the publication of journal articles is more easily achieved than in fields where qualitative methods are the norm (Kyvik, 2003; Moody, 2004).

Flemish sociologists, one of the cases documented in Section 3, can serve as a telling example of the way in which specialization can divide the researchers belonging to a single discipline. When clustered by means of the hard partitioning, the publication practices of sociologists show a distinctive pattern, with 47.7% belonging to Cluster One (international journals and English) and 52.3% to Cluster Two (national journals and books) (Verleysen & Weeren, 2016). Topical specialization does indeed explain this division to a considerable extent. A study from 2010 on publication patterns in Flemish Sociology has found that some communities of Flemish sociologists in more recent years have initiated an active participation in international communication networks (Vanderstraeten, 2010), which at the disciplinary level is attested to by growing shares of WoS-indexed journal articles (Engels et al., 2012) and English-language books published by prestigious international academic publishing houses (Verleysen & Engels, 2014). In stark contrast, other research groups in Sociology in Flanders have retained a focus on studies at the national or regional level, with articles mainly in three Dutch-language journals published in Flanders or the Netherlands, which retain a strong national profile and hardly attract an international authorship or readership (Vanderstraeten, 2010).

Returning to the methodological point of view, the results of the fuzzy analysis presented in this paper are somewhat different from those of the hard partitioning previously conducted by Verleysen and Weeren (2016). Not only does the fuzzy result display additional information, it also avoids binary decisions for individual authors on cluster membership. This makes the result more complicated and slightly ambiguous. We note that especially for several humanities disciplines (Art History, History, Humanities General, Law, Literature and Theology) and two social sciences (Criminology and Political Sciences) the fuzzy result is indicative of gradual



differences between authors from the same discipline, a majority of which now lean towards the publication model in which national journals and books are the dominant publication types. This appears largely congruent with the traditional picture of research practices and information dissemination by humanities scholars. However, fuzzy cluster analysis of publication patterns at the author level does not result in equally less sharp internal divisions for all SSH disciplines. The cluster plots and probability density functions for four of the social sciences (Economics, Educational Sciences, Social Sciences General, and Sociology) point to a bifurcation of publication styles among researchers affiliated with the same discipline.

5 Conclusion

Cluster analysis has shown a valuable tool for the analysis of intra-disciplinary diversity of publication patterns in the social sciences and humanities. A fuzzy cluster analysis based on a prior hard partitioning results in a maximum of information: the partitioning based on the k -medoids algorithm allows for straightforward identification of the publication patterns underlying the clustering result, while the fuzzy principle shows for every individual author the degree of belonging to each cluster. In a final step, the shape of the probability density function shows for each discipline how publication styles are distributed over its authors and how heterogeneous scholarly fields of research really are.

All in all, this method for analyzing publication patterns seems well applicable to other bibliometric or research evaluation contexts, provided that the attributes of the cases to be clustered are derived from the actual scholarly research environment one wishes to analyze. The variables used for the Flemish case, or very similar ones, are probably also applicable to other non-Anglophone countries or regions (Verleysen & Weeren, 2016).

Acknowledgements

The authors thank Tryuken Ossenblok for making available the dataset used in this and a previous article, and its adaptation at our request. We also thank Tim Engels, Raf Guns, and Ronald Rousseau for their comments on a first manuscript of this paper. ECOOM is funded by the Flemish Government.

Author Contributions

The research questions were proposed by F.T. Verleysen (Frederik.Verleysen@uantwerpen.be, corresponding author). The data were collected by F.T. Verleysen (see also acknowledgements). The research framework was conceived by both authors. The programming in MATLAB®R2016a was carried out by A. Weeren (Arie.Weeren@mathworks.nl). F.T. Verleysen wrote and revised the final manuscript.



References

- Chi, P.S. (2015). Changing publication and citation patterns in political science in Germany. *Scientometrics*, 105, 1833–1848.
- Dunn, J.C. (1976). Indices of partition fuzziness and the detection of clusters in large data sets. In M. Gupta (Ed.), *Fuzzy Automata and Decision Processes*. New York: Elsevier.
- Engels, T.C.E., Ossenblok, T.L.B., & Spruyt, E.H.J. (2012). Changing publication patterns in the social sciences and humanities, 2000–2009. *Scientometrics*, 93(2), 373–390.
- Hicks, D. (2004). The four literatures of social science. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems* (pp. 473–496). Dordrecht: Kluwer Academic.
- Johnson, R.A., & Wichern, D.W. (1992). *Applied multivariate statistical analysis* (Third ed.). New Jersey: Prentice Hall.
- Kaufman, L., & Rousseeuw, P.J. (1990). *Finding groups in data: An introduction to cluster analysis*. New York: Wiley.
- Kyvik, S. (2003). Changing trends in publishing behaviour among university faculty, 1980–2000. *Scientometrics*, 58(1), 35–48.
- Lin, Y., & Kaid, L.L. (2000). Fragmentation of the intellectual structure of political communication study: Some empirical evidence. *Scientometrics*, 47(1), 143–164.
- Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963–1999. *American Sociological Review*, 69(2), 213–239.
- Nederhof, A.J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, 66(1), 81–100.
- Nederhof, A.J. (2011). A bibliometric study of productivity and impact of modern language and literature research. *Research in Science Education*, 20(2), 117–129.
- Ossenblok, T.L.B. (2016). *Scientific communication in the social sciences and humanities. Analysis of publication and collaboration in Flanders*. Antwerp. (University of Antwerp Ph.D. dissertation)
- Persson, O. (2015). Bibliometric analysis of two subdomains in philosophy: Free will and sorites. *Scientometrics*, 103(1), 47–73.
- Puuska, H.M. (2014). *Scholarly publishing patterns in Finland—a comparison of disciplinary groups*. (doctoral dissertation), University of Tampere, Tampere, Finland.
- Sivertsen, G. (2009). Publication patterns in all fields. In F. Aström, R. Danell, B. Larsen, & J. W. Schneider (Eds.), *Celebrating Scholarly Communication Studies: A Festschrift for Olle Persson at His 60th Birthday* (pp. 55–60). Leuven, Belgium: International Society for Scientometrics and Informetrics.
- Vanderstraeten, R. (2010). Scientific communication: Sociology journals and publication practices. *Sociology*, 44, 559–576.
- Verleysen, F.T., & Weeren, A. (2016). Clustering by publication patterns of senior authors in the social sciences and humanities. *Journal of Informetrics*, 10, 254–272.
- Whitley, R. (2000). *The intellectual and social organization of the sciences*. Oxford: Oxford University Press.



This is an open access article licensed under the Creative Commons—Attribution-Non Commercial-NoDerivs License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



Appendix[Ⓞ]

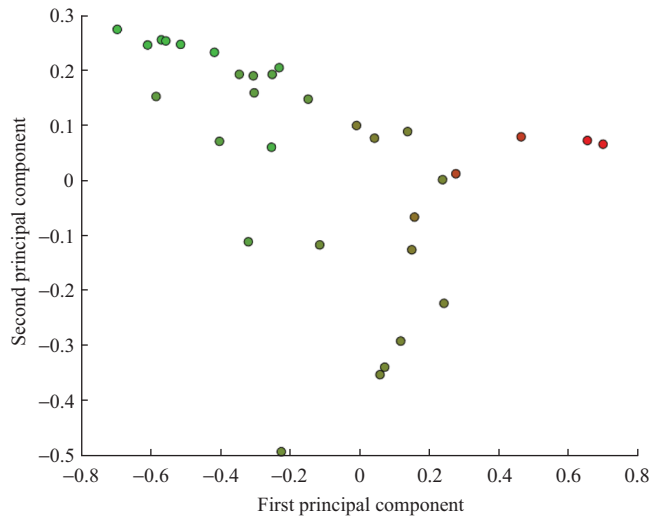


Figure 1. Fuzzy clustering of productive authors in Archeology ($n = 32$).

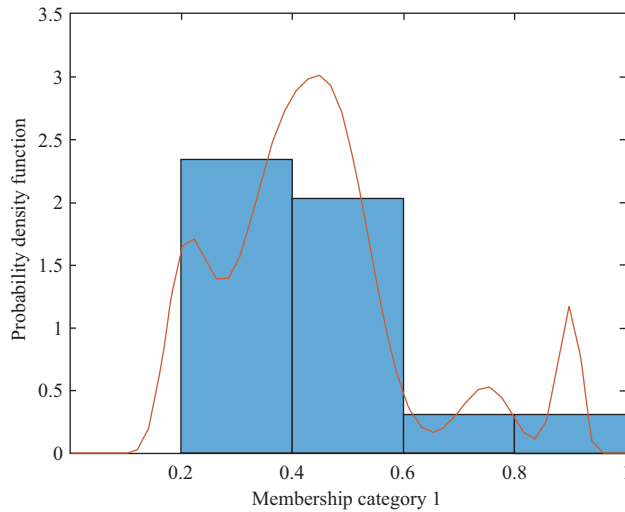


Figure 2. Probability density function of cluster membership coefficients for Archeology ($n = 32$).



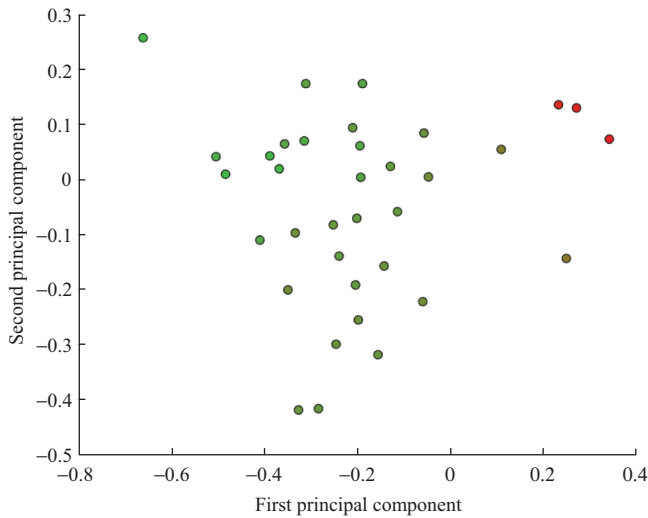


Figure 3. Fuzzy clustering of productive authors in Communication Studies ($n = 35$).

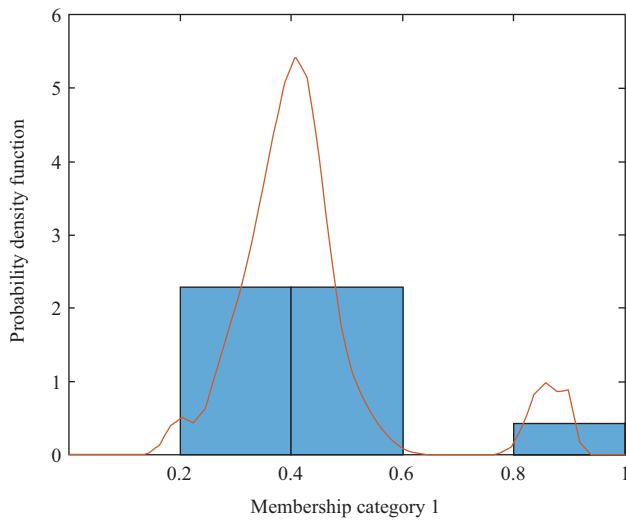


Figure 4. Probability density function of cluster membership coefficients for Communication Studies ($n = 35$).



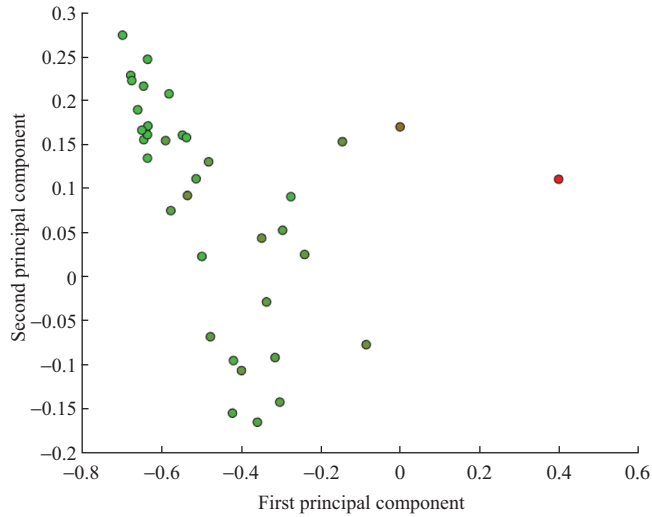


Figure 5. Fuzzy clustering of productive authors in Criminology ($n = 39$).

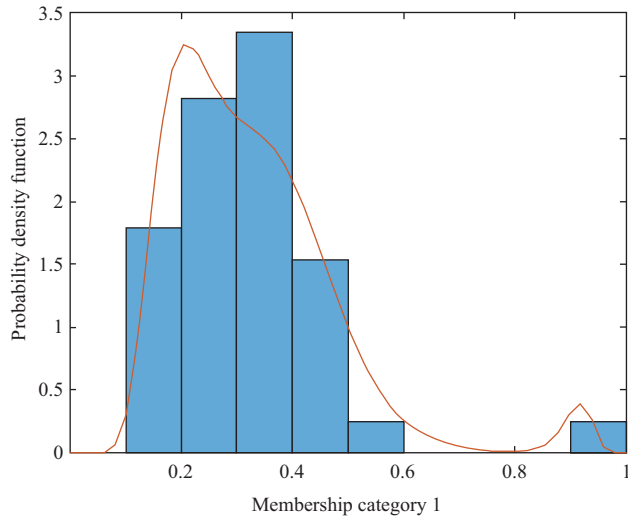


Figure 6. Probability density function of cluster membership coefficients for Criminology ($n = 39$).



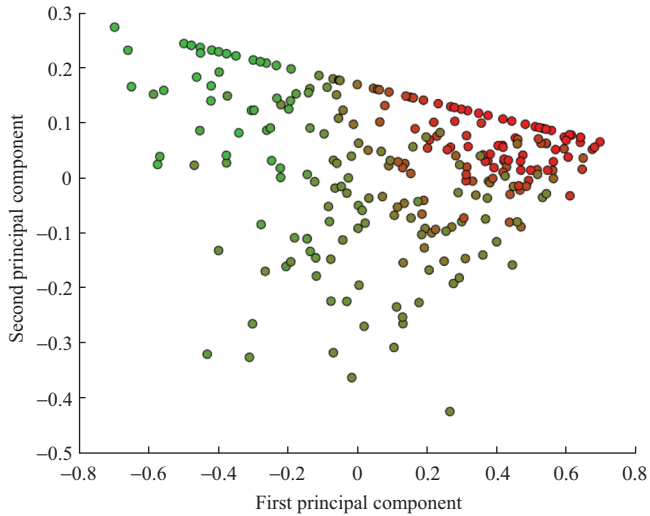


Figure 7. Fuzzy clustering of productive authors in Economics ($n = 271$).

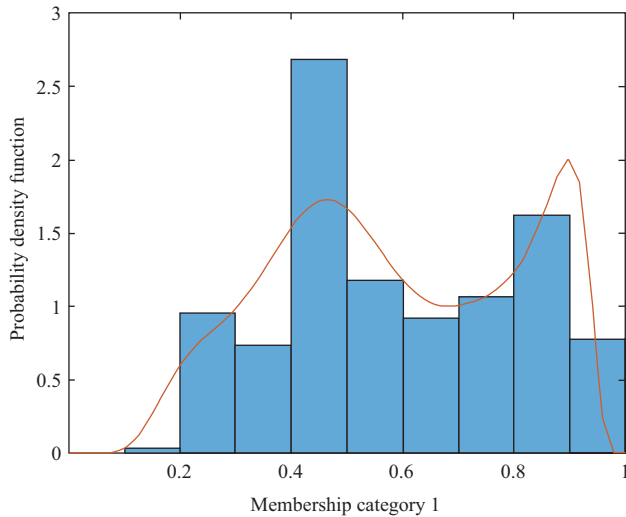


Figure 8. Probability density function of cluster membership coefficients for Economics ($n = 271$).



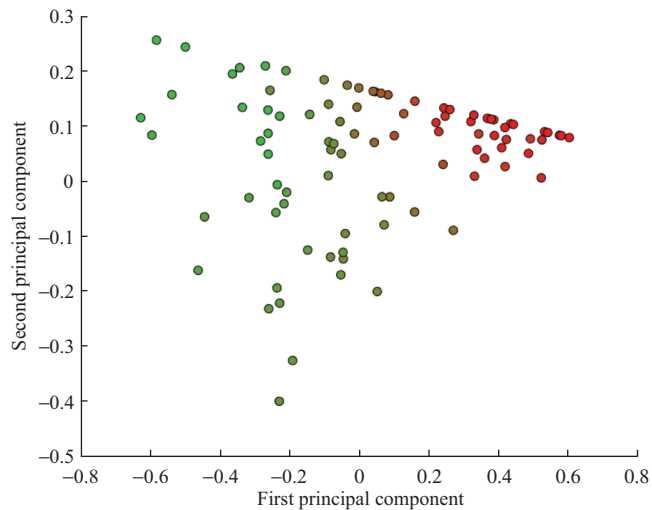


Figure 9. Fuzzy clustering of productive authors in Educational Sciences ($n = 96$).

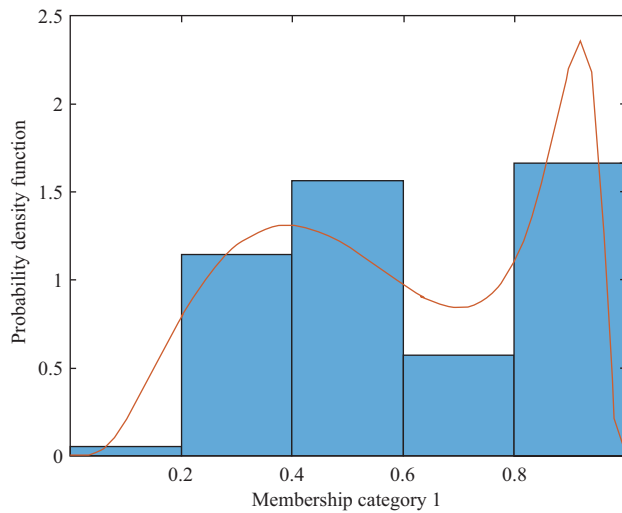


Figure 10. Probability density function of cluster membership coefficients for Educational Sciences ($n = 96$).



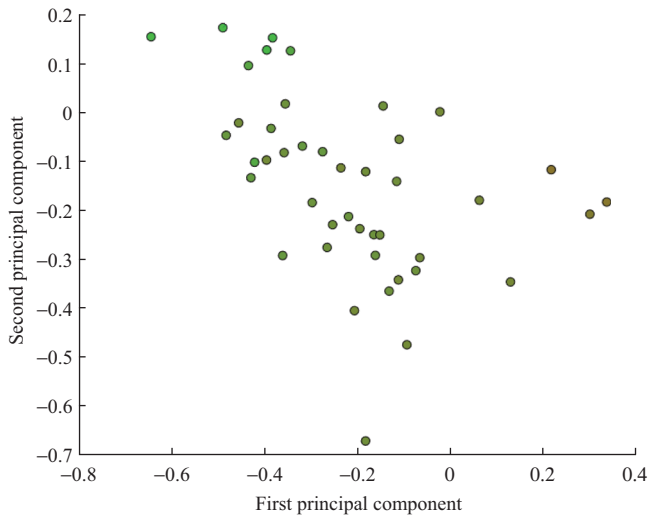


Figure 11. Fuzzy clustering of productive authors in Humanities General ($n = 43$).

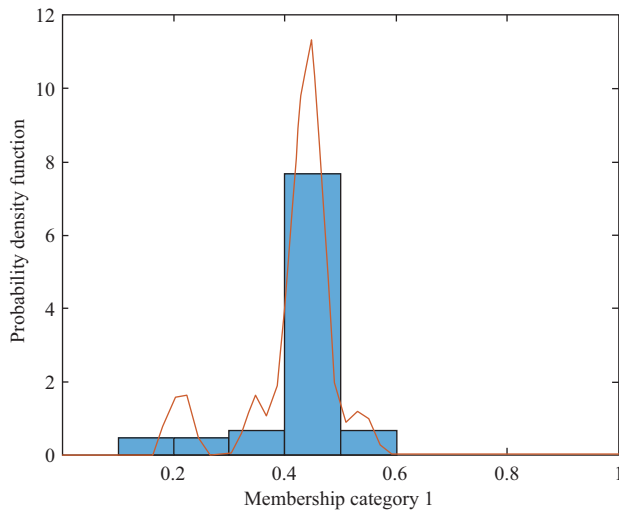


Figure 12. Probability density function of cluster membership coefficients for Humanities General ($n = 43$).



Research Paper

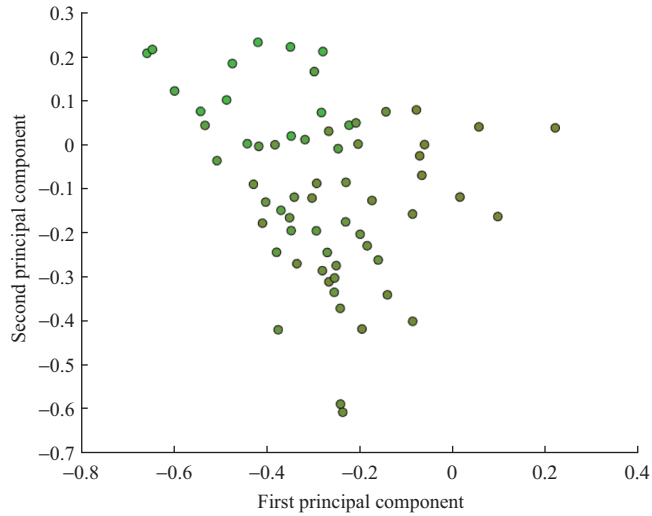


Figure 13. Fuzzy clustering of productive authors in History ($n = 66$).

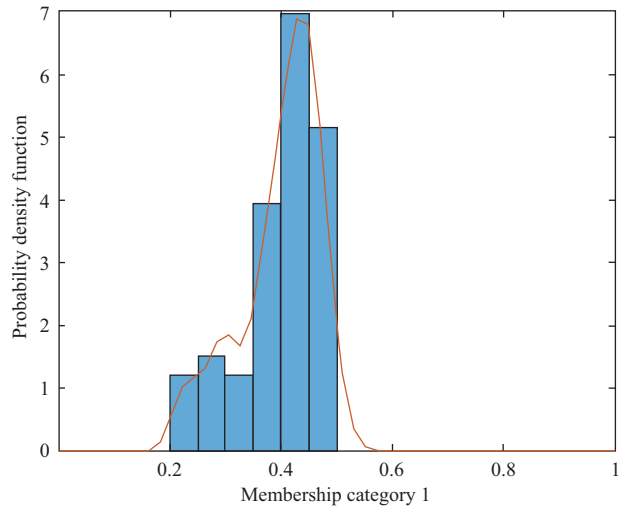


Figure 14. Probability density function of cluster membership coefficients for History ($n = 66$).



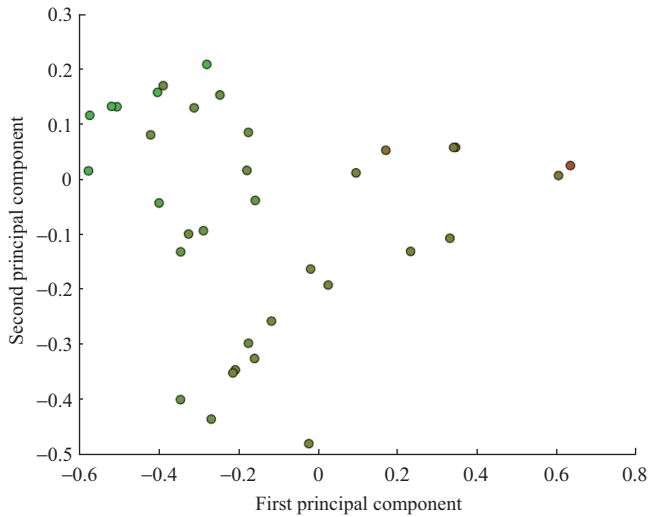


Figure 15. Fuzzy clustering of productive authors in History of Arts ($n = 35$).

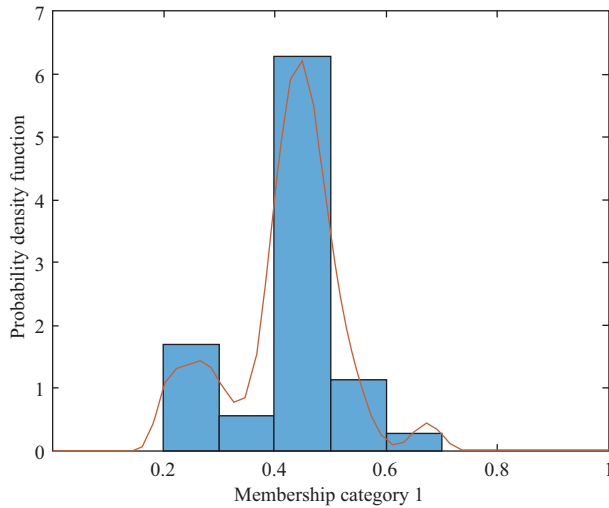


Figure 16. Probability density function of cluster membership coefficients for History of Arts ($n = 35$).



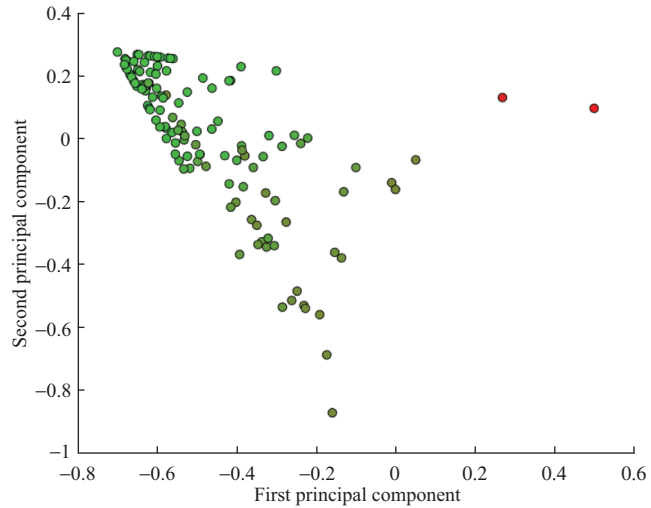


Figure 17. Fuzzy clustering of productive authors in Law ($n = 206$).

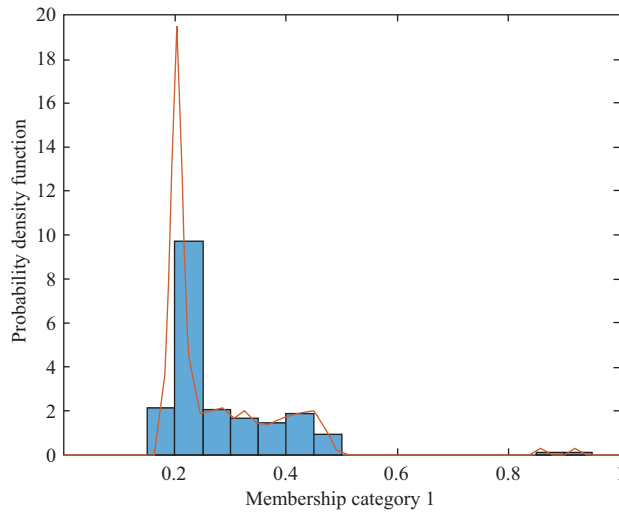


Figure 18. Probability density function of cluster membership coefficients for Law ($n = 206$).



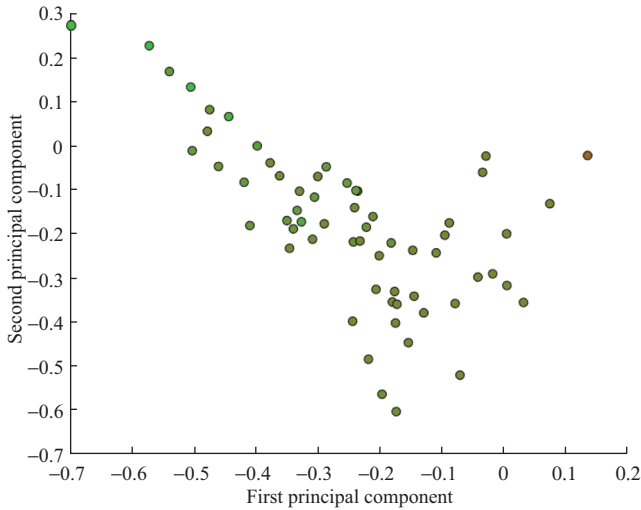


Figure 19. Fuzzy clustering of productive authors in Literature ($n = 62$).

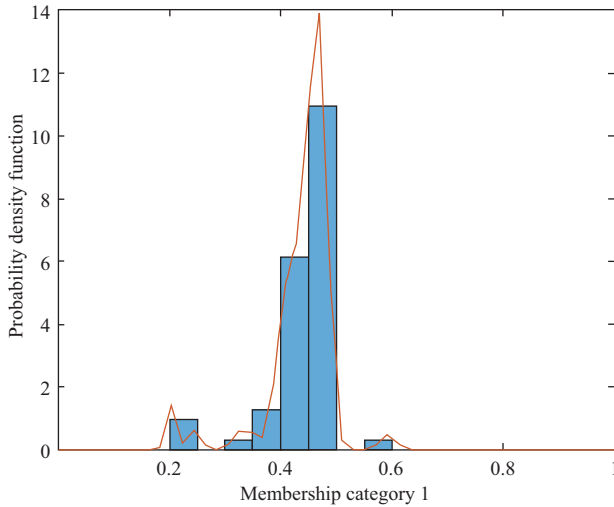


Figure 20. Probability density function of cluster membership coefficients for Literature ($n = 62$).



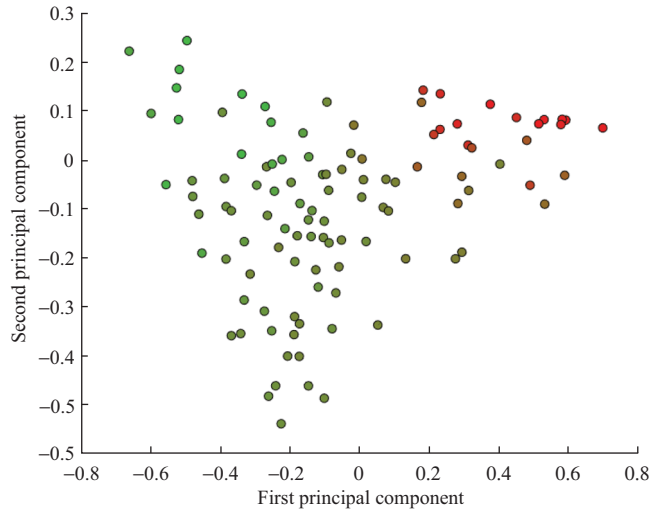


Figure 21. Fuzzy clustering of productive authors in Philosophy ($n=108$).

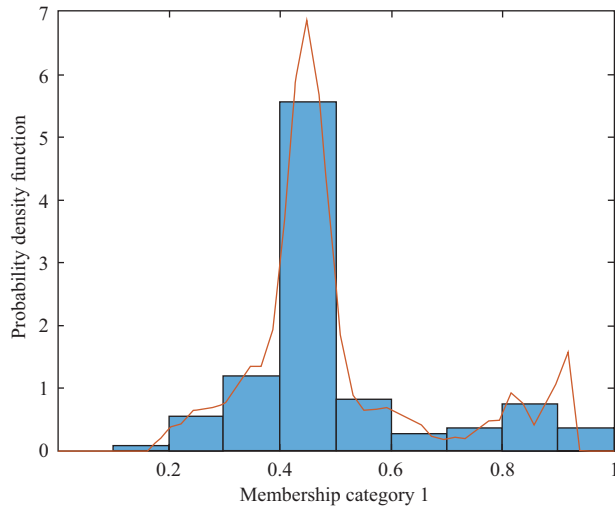


Figure 22. Probability density function of cluster membership coefficients for Philosophy ($n=108$).



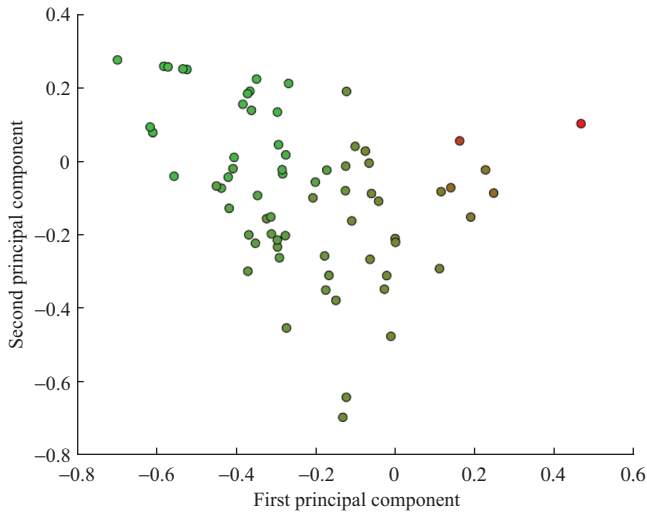


Figure 23. Fuzzy clustering of productive authors in Political Science ($n = 74$).

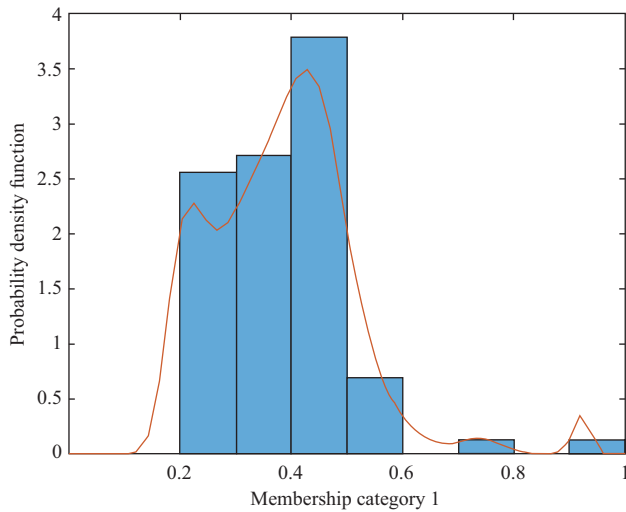


Figure 24. Probability density function of cluster membership coefficients for Political Science ($n = 74$).



Research Paper

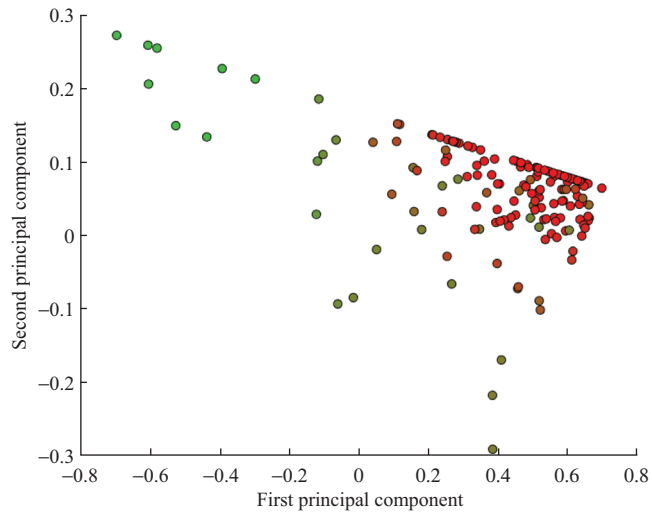


Figure 25. Fuzzy clustering of productive authors in Psychology ($n = 199$).

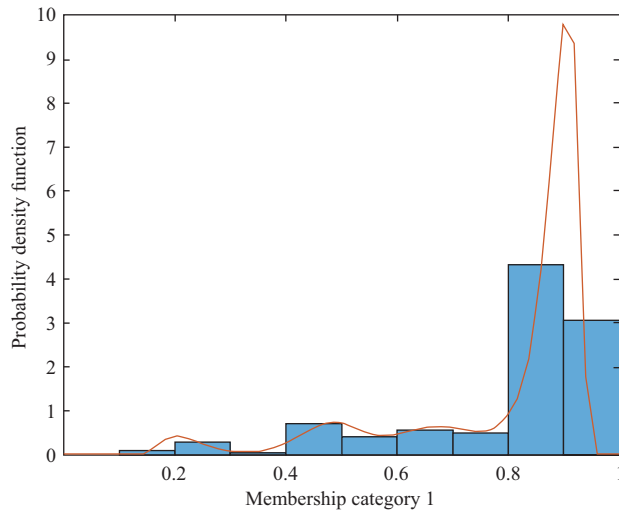


Figure 26. Probability density function of cluster membership coefficients for Psychology ($n = 199$).



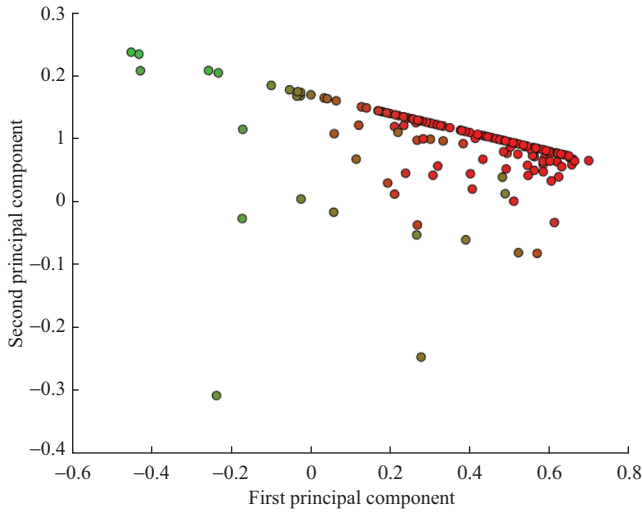


Figure 27. Fuzzy clustering of productive authors in Social Health Sciences ($n = 228$).

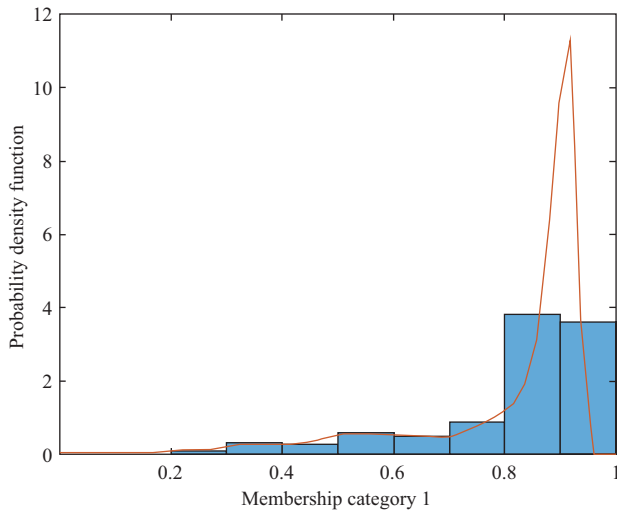


Figure 28. Probability density function of cluster membership coefficients for Social Health Sciences ($n = 228$).



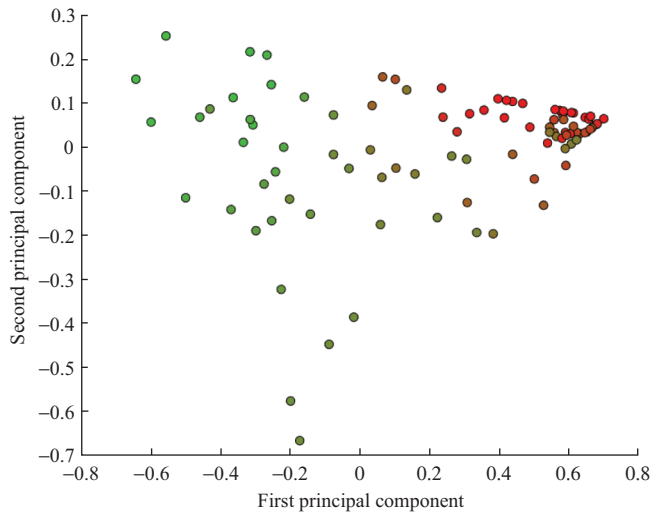


Figure 29. Fuzzy clustering of productive authors in Social Sciences General ($n = 102$).

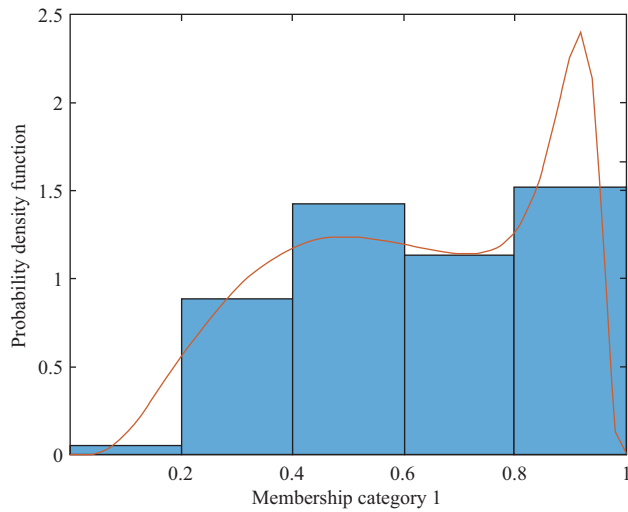


Figure 30. Probability density function of cluster membership coefficients for Social Sciences General ($n = 102$).



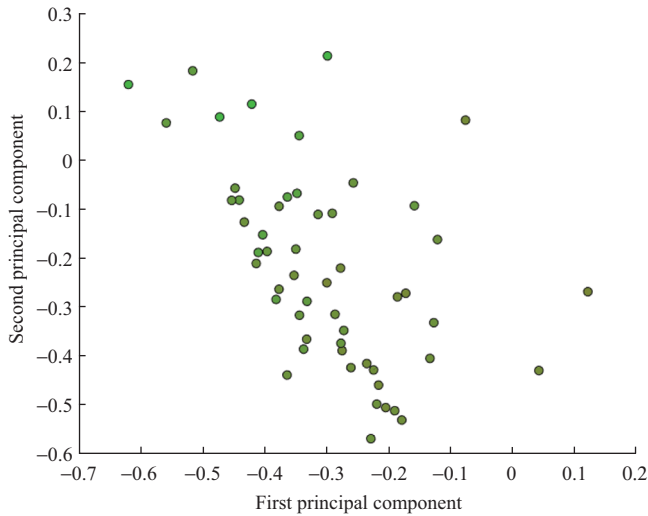


Figure 31. Fuzzy clustering of productive authors in Theology ($n = 54$).

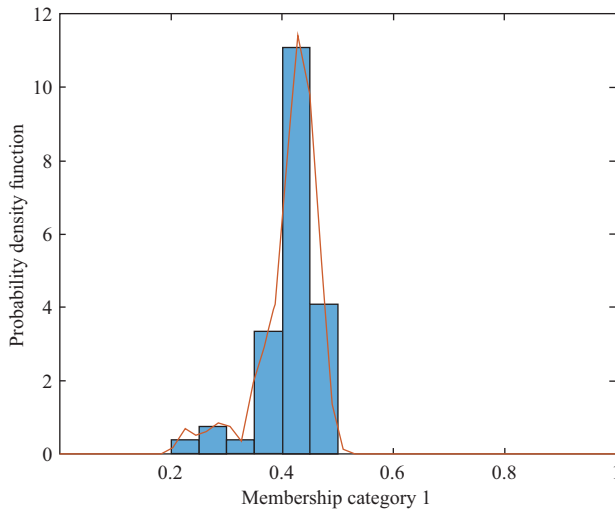


Figure 32. Probability density function of cluster membership coefficients for Theology ($n = 54$).



CO-EDITORS-IN-CHIEF

Xiaolin Zhang

National Science Library, Chinese Academy of Sciences, China

Ronald Rousseau

University of Leuven, KU Leuven, Belgium

Ying Ding

Indiana University, USA

EDITORIAL BOARD

Per Ahlgren

KTH Royal Institute of Technology, Sweden

Judit Bar-Ilan

Bar-Ilan University, Israel

Christine L. Borgman

University of California, Los Angeles, USA

Kevin Boyack

SciTech Strategies Inc., USA

Chaomei Chen

Drexel University, USA

Dar-Zen Chen

Taiwan University, Taiwan, China

Ling Chen

Peking University, China

Cinzia Daraio

Sapienza University of Rome, Italy

Haiyan Hou

Dalian University of Technology, China

Tao Jia

Southwest University, China

Renaud Lambiotte

University of Namur, Belgium

Guangjian Li

Peking University, China

Yuelin Li

Nankai University, China

Wei Liu

Library of Shanghai, China

Wei Lu

Wuhan University, China

Xiaobin Lu

Renmin University, China

Everard Noyons

University of Leiden, the Netherlands

José Miguel B. Nunes

Sun Yat-Sen University, China

Han Woo Park

Yeung Nam University, South Korea

Qing Qian

Institute of Medical Information/Medical Library, Chinese Academy of Medical Sciences, China

Min Song

Yonsei University, South Korea

Neil Smalheiser

University of Illinois at Chicago, USA

Xinning Su

Nanjing University, China

Tan Sun

Agriculture Information Institute, Chinese Academy of Agricultural Sciences, China

Jie Tang

Tsinghua University, China

Li Tang

Fudan University, China

Mike Thelwall

University of Wolverhampton, UK

Nees-Jan van Eck

Leiden University, the Netherlands

Yuefen Wang

Nanjing University of Science and Technology, China

Fang Wang

Nankai University, China

Jevin West

Washington University, USA

Dietmar Wolfram

University of Wisconsin-Milwaukee, USA

Dan Wu

Wuhan University, China

Yishan Wu

Chinese Academy of Science and Technology for Development, China

Feng Xia

Dalian University of Technology, China

Erjia Yan

Drexel University, USA

Liying Yang

National Science Library, Chinese Academy of Sciences, China

Ying Ye (Fred Y. Ye)

Nanjing University, China

Jan Youtie

Georgia Institute of Technology, USA

Marcia L. Zeng

Kent State University, USA

Zhixiong Zhang

National Science Library, Chinese Academy of Sciences, China

Dangzhi Zhao

University of Alberta, Canada

Yuxiang Zhao

Nanjing University of Science and Technology, China

Donghua Zhu

Beijing Institute of Technology, China

MANAGING EDITOR

Fenghong Liu

National Science Library, Chinese Academy of Sciences, China

EDITORIAL STAFF

Fenghong Liu, Jing Cao, Lin Peng

National Science Library, Chinese Academy of Sciences, China

Copyright ©2016. All rights are reserved by the Editorial Office of *Journal of Data and Information Science (JDIS)*, National Science Library, Chinese Academy of Sciences. Address: No 33, Beisihuan Xilu, Zhongguancun, Haidian District, Beijing 100190, P.R.China.

Tel: 86-10-82624454 or 86-10-82626611 ex. 6628. Fax: 86-10-82624454. Email: jdis@mail.las.ac.cn. Website: <http://www.jdis.org>.

Published by: National Science Library, Chinese Academy of Sciences, No 33, Beisihuan Xilu, Zhongguancun, Haidian District, Beijing 100190, P. R. China

Edited by: Editorial Office of *Journal of Data and Information Science (JDIS)*, No 33, Beisihuan Xilu, Zhongguancun, Haidian District, Beijing 100190, P. R. China

Printed by: Beijing KEXIN Printing Co. Ltd., Beijing 102208, P.R. China.

Tel: 86-10-62903036. **Fax:** 86-0-62805493

Co-Editors-in-Chief: Xiaolin Zhang, Ronald Rousseau & Ying Ding

Typesetting: Charlesworth (Beijing) Information Services Co Ltd., Room 1105, Building No. 9, Jianwai SOHO, Chaoyang District, Beijing 100022, P.R. China
Tel/Fax: 86-10-58698392.

Distributed by: Editorial Office of *Journal of Data and Information Science (JDIS)*, No 33, Beisihuan Xilu, Zhongguancun, Haidian District, Beijing 100190, P.R. China

Subscription: RMB 200/Issue, RMB 800/Volume in China per year; US\$ 199/Volume outside of China (including air shipping).

Distributional Code: 82-563

JDIS

JOURNAL OF DATA AND INFORMATION SCIENCE (QUARTERLY)

Volume 1 Number 4 2016

CONTENTS

PERSPECTIVES

- 1 | Data-driven Discovery: A New Era of Exploiting the Literature and Data
Ying Ding & Kyle Stirling

EXPERT REVIEWS

- 10 | Under-reporting of Adverse Events in the Biomedical Literature
Ronald N. Kostoff

RESEARCH PAPERS

- 33 | Mapping Diversity of Publication Patterns in the Social Sciences and Humanities: An Approach Making Use of Fuzzy Cluster Analysis
Frederik T. Verleysen & Arie Weeren

- 60 | Open Peer Review in Scientific Publishing: A Web Mining Study of *PeerJ* Authors and Reviewers
Peiling Wang, Sukjin You, Rath Manasa & Dietmar Wolfram
- 81 | Topic Detection Based on Weak Tie Analysis: A Case Study of LIS Research
Ling Wei, Haiyun Xu, Zhenmeng Wang, Kun Dong, Chao Wang & Shu Fang

CONTENTS INDEX

- 102 | Contents Index to Volume 1

ISSN 2096-157X



ISSN 2096-157X

CN 10-1394/G2