

This item is the archived peer-reviewed author-version of:

Clustering by publication patterns of senior authors in the social sciences and humanities

Reference:

Verleysen Frederik, Weeren Arie.- Clustering by publication patterns of senior authors in the social sciences and humanities
Journal of informetrics - ISSN 1751-1577 - 10:1(2016), p. 254-272
Full text (Publishers DOI): <http://dx.doi.org/doi:10.1016/j.joi.2016.01.004>

Clustering by publication patterns of senior authors in the social sciences and humanities

Frederik T. Verleysen and Arie Weeren

Frederik.Verleysen@uantwerpen.be (corresponding author)

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

Arie.Weeren@uantwerpen.be

StatUA, University of Antwerp, Prinsstraat 13, 2000 Antwerp (Belgium)

Abstract

This study uses cluster analysis as a tool for mapping diversity of publication patterns in the social sciences and humanities (SSH). By algorithmic clustering of 1828 senior authors affiliated with 16 disciplines at five universities in Flanders, Belgium, based on the similarity of their publication patterns during 2000-2011, we distinguish two broad publication styles, both of which are present within each discipline. We conclude that diversity in SSH publication patterns cuts across disciplinary boundaries. Cluster analysis shows promise for application in research evaluation for the SSH.

Keywords: cluster analysis, bibliometrics, senior authors, social sciences, humanities, publication practices.

1. Introduction

Cluster analysis (CA) is a multivariate technique for classifying similar objects into groups. A major strength of algorithmic clustering lies in its ability to unambiguously classify cases by a multitude of attributes within otherwise opaque datasets, and to plot the results in a geometric representation (Kaufman & Rousseeuw, 1990; Johnson & Wichern, 1992). As the development of classification schemes has always played an essential role in science, there are many applications of CA in various scientific fields. Unsurprisingly, in information science as well, the use of CA has been advocated (Egghe & Rousseau, 1990). Well-known applications in this field involve the analysis of bibliographic networks and the cognitive-epistemological structure of the scientific system (Small, Sweeney, & Greenlee, 1985a; Small, Sweeney, & Greenlee, 1985b; Waltman, van Eck, & Noyons, 2010; van Eck & Waltman, 2014), or of a single field of research (Lin & Kaid, 2000; Persson et al., 2015). However, many more applications are conceivable. In 2005, for example,

Liu et al. have used clustering from a research evaluation perspective to identify groups of Chinese scientific research institutions (Liu, Li, Xu, & Shi, 2005). In the present article, we apply algorithmic clustering to publication patterns at the level of individual senior authors in the social sciences and humanities.

As bibliometric research of the social sciences and humanities (SSH) is growing to maturity, more attention is paid to their internal heterogeneity and dynamics, often within the context of national evaluation and/or performance-based research funding systems (Hicks, 2013; Engels, Ossenblok, & Spruyt, 2012; Hammarfelt & de Rijcke, 2015). Disciplinary publication cultures in the SSH vary to a considerable extent, as has been measured by the share of publication types (books in addition to journal articles), publication language (national or regional languages in addition to or instead of English), target audience (international or local, academic or non-academic), and the frequency of co-authorship. In summary, in many Western countries, most disciplines classified as social sciences tend to show a publication pattern in which international journals, the use of English and frequent co-authorship are starting to predominate, while, in comparison, most humanities disciplines remain more strongly oriented towards book publications and national journals, make use of national or regional languages, and often continue to prefer single authorship (Kyvik, 2003; Nederhof, 2006; Sivertsen, 2009; Sivertsen & Larsen, 2012; Ossenblok, Verleysen, & Engels, 2014; Verleysen & Engels, 2014; Puuska, 2014; Hammarfelt & de Rijcke, 2015).

While contrasting the publication patterns of disciplinary groups has greatly advanced the understanding of research and publication practices in the SSH, and has also contributed to the development of evaluation systems that are better adapted to a scholarly research environment (Hicks, 2013), there are limitations to this approach. Clearly, the aggregation of individual researchers' publications into statistics at the disciplinary level obfuscates intra-disciplinary diversity. The existence of formal disciplinary boundaries throughout science is the outcome of a historical process of cognitive and social structuring (Whitley, 2000; Bod, Maat, & Weststeijn, 2012), and while this process undoubtedly has had its influence on disciplinary *epistemic cultures* (Knorr Cetina, 1999), including publication practices, this in itself does not imply that boundaries between such cultures necessarily coincide with organizational demarcations. Although intra-disciplinary diversity is certainly deserving of more attention by bibliometric research on the SSH, a handful of studies on individual disciplines are available, demonstrating their cognitive fragmentation (Lin & Kaid, 2000; Ahlgren et al., 2015) and/or their (related) heterogeneity in terms of publication and citation patterns (Nederhof & Noyons, 1992; Nederhof, 2011; Chi, 2015). Our present study starts from the premise that individual authors in the SSH publish their work in a more idiosyncratic way than an analysis at the aggregation level of disciplines is able to reveal.

To corroborate this in a systematic way, we apply algorithmic clustering to a bibliographic dataset on the comprehensive peer-reviewed publication output of individual senior authors affiliated with 16 SSH disciplines at five universities in Flanders, the Northern Dutch-language part of Belgium. By letting a computer algorithm autonomously form groups of authors, based on the similarity of their publication patterns and regardless of disciplinary affiliations, we demonstrate that epistemic diversity reflected by publication patterns cuts across disciplinary boundaries. From a science policy and research evaluation perspective, both method and result of our analysis offer additional insights.

2. Data

Data used in our analysis is registered in the VABB-SHW, the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (or *Vlaams Academisch Bibliografisch Bestand voor de Sociale en Humane Wetenschappen*). Introduced in 2010, the VABB-SHW comprehensively registers all peer reviewed publications since the year 2000 by researchers affiliated with 16 SSH disciplines and two general categories at the five universities in Flanders (Belgium). The VABB-SHW is used in the regional research funding model for the five universities (Verleysen, Ghesquière, & Engels, 2014).

Five publication types are registered in the VABB-SHW: journal articles, monographs, edited books, book chapters and proceedings papers. For inclusion in the funding model, a weight is attributed to each 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 database. The first, VABB-WoS consists of references to 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 for 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 references which have additionally been selected 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 et al., 2012). Thus, in the results section below, with regard to journal articles and proceedings papers, a distinction is made between the subsets of VABB-WoS and VABB-GP. Book publications all stem from the VABB-GP subset.

For the present study the VABB-SHW dataset (N = 10181 authors) was delimited to the output of the most senior authors (n = 1828). These are defined in terms of publication productivity, i.e. as having published at least ten weighted outputs in at least four years. Junior (i.e. less productive) authors (n = 8353) were excluded from the analysis, because their less numerous and more

sporadic publications -typically a handful of articles or proceedings- have not yet had the chance to crystallize into a clearly discernible pattern. 6171 or 73.8% of junior authors have less than five weighted outputs in the twelve year time span. In the VABB-SHW, disciplines are defined based on the institutional affiliation of authors.

As input for the clustering algorithm (cfr. *infra*, section 3), a table was constructed listing the 1828 author names, their main disciplinary affiliation, as well as 11 variables mapping their publication output in the 2000-2011 time frame. 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 variables. The 11th variable is the fraction of weighted co-authored publications.

Data was processed in the Brocade library system of the University of Antwerp (<http://www.brocade.be/product-tour/brocade-library-services>), Microsoft Excel® and MATLAB®R2015a.

3. Method

In order to demonstrate diversity of publication patterns between and within disciplines, we present a contrastive analysis between publication patterns at the disciplinary level and those at the author level. A first step consisted of generating publication statistics from the VABB-SHW data (2000-2011) at the level of the 16 disciplines and two general categories. A second step entailed the clustering of individual senior authors, regardless of their disciplinary affiliation. For the cluster analysis, the steps below describe our method.

3.1. Principal components analysis

A principal components analysis (PCA) was conducted on the dataset. For two- or three-dimensional visualization purposes, we wish to retain the scores on the first two or three principal components. The result of PCA indicates that the first two components explain 91% of variance, and therefore can be retained for visualizations.

3.2. Selection of dissimilarity measure

The dissimilarity measure used is Mahalanobis distance or 'generalized squared interpoint distance' (Mahalanobis, 1936; Johnson & Wichern, 1992). The Mahalanobis distance of an

observation $x = (x_1, x_2, x_3, \dots, x_N)^T$ from a set of observations with mean $\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_N)^T$ and covariance matrix S is defined as:

$$D_m(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)}$$

Mahalanobis distance adjusts for correlation between variables, which makes it especially well-suited for application to often highly correlated publication data. It not only accounts for the variance of each variable, but also for that between variables. Geometrically, it does this by transforming the data into standardized uncorrelated data and subsequently computing the ordinary Euclidean distance for the transformed data. By doing this, Mahalanobis distance is the multivariate analog of a z-score (standardized variable) and provides a way to measure distances that takes into account the scale of the data (Wicklin, 2015).

3.3. Number of clusters: hierarchical clustering and silhouette

Initially, hierarchical clustering was used to assess the optimal number of clusters in the data. A dendrogram was generated to visualize the results. As this did not present a good visual clue on the number of clusters, a more formal alternative, the *silhouette* technique (Rousseeuw, 1987) was used to decide on the number of clusters. Results of silhouette indicate that the optimal number of clusters in our data is two.

3.4. Selection of clustering algorithm

The partitioning algorithm *k-Medoids* was selected, making use of Mahalanobis distance for computation of the dissimilarity matrix. *k-Medoids* is a more robust and more widely applicable alternative to the well-known *k-Means* algorithm. In order to obtain k clusters, *k-Medoids* first selects k most representative objects or *medoids* from the data set. Corresponding clusters are then determined by assigning all remaining objects to the nearest medoid, which must be centrally located in the clusters they define, i.e. their average distance to every other object within the same cluster must be minimal (Kaufman & Rousseeuw, 1987).

In order to label the two obtained clusters, the two medoids (most representative authors) and their respective publication patterns were identified in the dataset.

4. Results

4.1. Diversity among disciplines

At the disciplinary level, there is considerable diversity within the SSH as practiced at Flemish universities. In this section, publication patterns for the 1828 senior authors are described at the disciplinary level based on publication type, publication language and co-authorship occurrence.

Table 1 presents statistics on the share of the five weighted publication types registered in the VABB-SHW. For both journal articles and proceedings papers a distinction is made between WoS-indexed and GP-approved publications.

Table 1: share of weighted publication types for senior authors in 16 SSH disciplines and two general categories (2000-2011)

| Discipline | Journal Articles | | | | Monographs | | Edited Books | | Book Chapters | | Proceedings Papers | | | |
|------------------------|------------------|-------------|--------------|-------------|-------------|------------|--------------|------------|---------------|-------------|--------------------|------------|-------------|------------|
| | VABB-GP | | VABB-WoS | | VABB-GP | | VABB- GP | | VABB- GP | | VABB-GP | | VABB-WoS | |
| | # | % | # | % | # | % | # | % | # | % | # | % | # | % |
| Archeology | 348 | 41.8 | 291 | 35.0 | 72 | 8.6 | 9 | 1.0 | 82 | 9.8 | 21 | 2.5 | 8 | 0.9 |
| Art History | 278 | 39.9 | 194 | 27.8 | 64 | 9.1 | 26 | 3.7 | 103 | 14.7 | 15,5 | 2.2 | 15,5 | 2.2 |
| Communication Sciences | 377 | 45.5 | 227 | 27.4 | 32 | 3.8 | 28 | 3.3 | 148 | 17.8 | 2 | 0.2 | 13.5 | 1.6 |
| Criminology | 704 | 73.1 | 98 | 10.1 | 24 | 2.4 | 30 | 3.1 | 104 | 10.8 | 1 | 0.1 | 1.5 | 0.1 |
| Economics & Business | 2082 | 29.1 | 3981 | 55.7 | 324 | 4.5 | 38 | 0.5 | 481 | 6.7 | 49.5 | 0.6 | 180.5 | 2.5 |
| Educational Sciences | 1101 | 38.2 | 1398 | 48.5 | 48 | 1.6 | 30 | 1.0 | 258 | 8.9 | 28.5 | 0.9 | 13.5 | 0.4 |
| History | 597 | 48.8 | 190 | 15.5 | 60 | 4.9 | 60 | 4.9 | 301 | 24.6 | 13.5 | 1.1 | 1.5 | 0.1 |
| Law | 3642 | 78.0 | 212 | 4.5 | 252 | 5.4 | 60 | 1.2 | 491 | 10.5 | 5 | 0.1 | 3.5 | 0.0 |
| Linguistics | 849 | 32.7 | 605 | 23.3 | 220 | 8.4 | 159 | 6.1 | 704 | 27.1 | 41 | 1.5 | 11.5 | 0.4 |
| Literature | 575 | 38.3 | 162 | 10.8 | 180 | 12.0 | 132 | 8.8 | 430 | 28.7 | 18 | 1.2 | 1 | 0.0 |
| Philosophy | 862 | 34.4 | 936 | 37.4 | 156 | 6.2 | 74 | 2.9 | 447 | 17.8 | 4.5 | 0.1 | 21.5 | 0.8 |
| Political Sciences | 840 | 49.1 | 356 | 20.8 | 156 | 9.1 | 43 | 2.1 | 312 | 18,2 | 3 | 0.1 | 0.5 | 0.0 |
| Psychology | 1059 | 15.5 | 5425 | 79.9 | 40 | 0.5 | 12 | 0.1 | 207 | 3.0 | 27 | 0.3 | 19.5 | 0,2 |
| Social health Sciences | 1391 | 19.1 | 5690 | 78.5 | 20 | 0.2 | 6 | 0.0 | 107 | 1.4 | 5 | 0.0 | 27 | 0.3 |
| Sociology | 618 | 41.5 | 673 | 45.2 | 28 | 1.8 | 17 | 1.4 | 142 | 9.5 | 5.5 | 0.3 | 2.5 | 0.1 |
| Theology | 637 | 40.4 | 115 | 7.2 | 204 | 12.9 | 118 | 7.4 | 500 | 31.7 | 2 | 0.1 | 0.5 | 0.0 |
| SS General | 105 | 16.0 | 462 | 70.6 | 4 | 0.6 | 5 | 0.7 | 47 | 7.1 | 10.5 | 1.6 | 20 | 3.0 |
| H General | 325 | 38.3 | 144 | 16.9 | 120 | 14.1 | 34 | 4.0 | 204 | 24.0 | 13.5 | 1.5 | 7.5 | 0.8 |
| <i>Total</i> | <i>16390</i> | <i>35.9</i> | <i>21159</i> | <i>46.3</i> | <i>2004</i> | <i>4.4</i> | <i>881</i> | <i>1.9</i> | <i>5068</i> | <i>11.1</i> | <i>140.5</i> | <i>0.3</i> | <i>71.5</i> | <i>0.2</i> |

Overall, the journal article is clearly the most numerous publication type in all disciplines, though inter-disciplinary differences are considerable. The highest total share of articles (GP+WoS) is noted for Social Health sciences with 97.6% of its publication output. The least reliant on journal articles is Theology (47.6%). The shares of WoS-indexed and GP-approved articles, especially, show a large variation between disciplines: whereas Law and Criminology mostly rely on GP-approved journals (78% and 73.1% respectively), Social Health Sciences and Psychology show the strongest preference for WoS-indexed journals (78.5% and 79.9% respectively). The share of monographs varies between disciplines as well: Social Health Sciences only has 0.2%, whereas Theology has 12.9%. With regards to edited books, Literature (8.8%) is in lead, Social Health Sciences (<0.5%) comes last. Book chapters are especially important in several Humanities disciplines such as Theology (31.7%), Literature (28.7%), Linguistics (27.1%) and History (24.6%). Proceedings, finally, account by far for the smallest share of the five publication types. Leaving aside the general category off SS General, Art History notes the largest share (GP+WoS) of its total output (4.4%).

Since the year 2000, language use in the SSH in Flanders has seen a marked evolution: between 2000 and 2011, the total share of English in the whole VABB-SHW rose from 60.4% to 75.9%, and the share of other languages (Dutch and others) dropped from 39.6% to 24.1% (Engels et al., 2012).

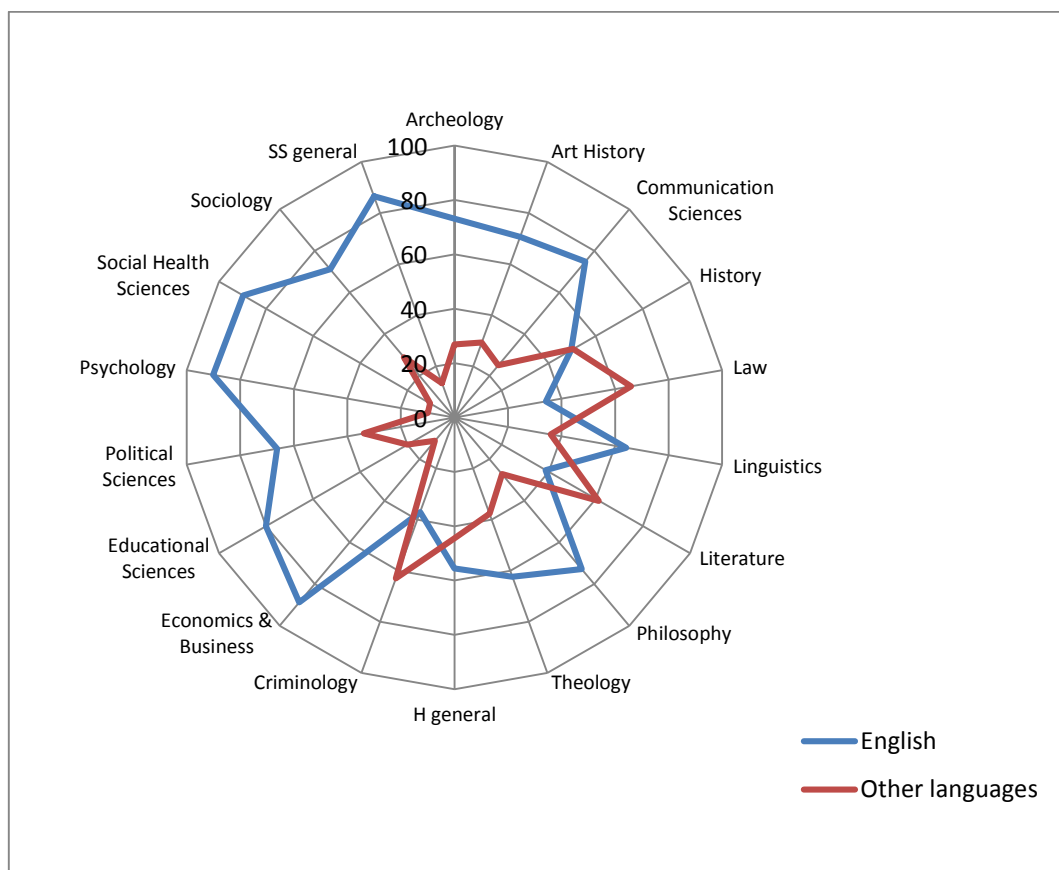


Figure 1: share of English and other publication languages for senior authors in 16 SSH disciplines and two general categories (2000-2011)

As regards the preference for specific publication languages, there are sizable differences between disciplines. In general, disciplines belonging to the social sciences rely more on English (e.g. Psychology 90.1%, Social Health Sciences 89.7%) than disciplines in the humanities. The notable exception among the social sciences is Criminology, where other languages (especially Dutch) take the lion's share of 63% of all publications. The humanities most reliant upon other languages are Law (classified as an H discipline in the VABB-SHW) (65.9%), Literature (61.3%) and History (50.4%).

Co-authorship patterns in Flemish SSH publications have previously been studied by Ossenblok et al. (2014). During the years 2000-2010, in the whole VABB-SHW single-authorship of journal articles and book chapters in the SSH has seen a steep decline from 55.6% to 36.8%. The rise of co-authorship was most profound in the social sciences (from 69.9% to 85.8%) and slightly less so in the humanities (from 22.2% to 34.1%).

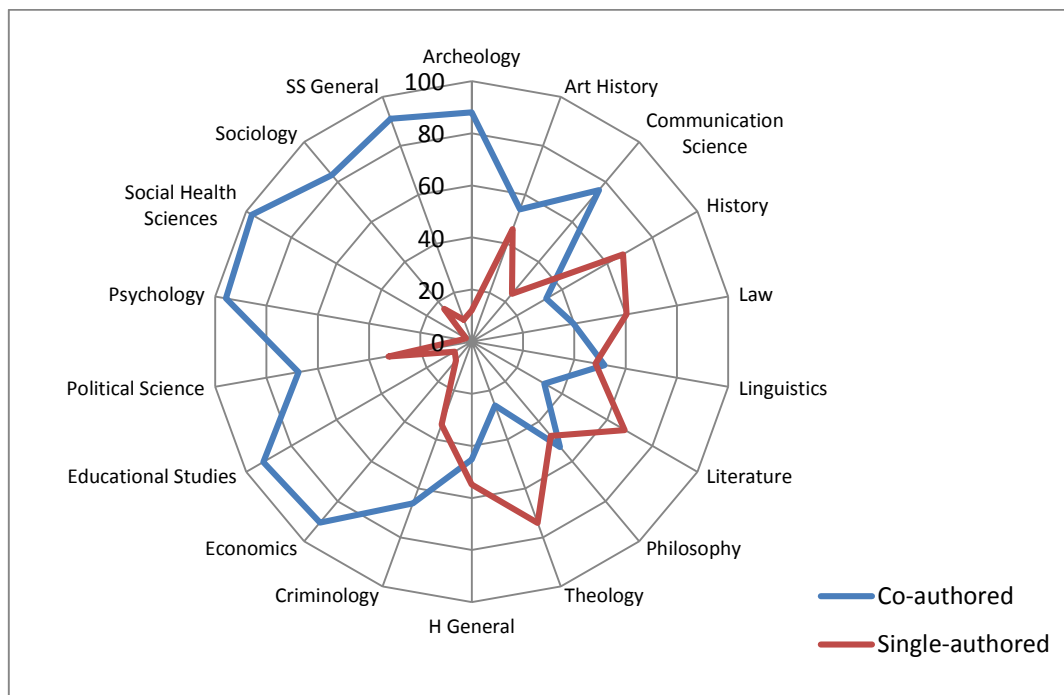


Figure 2: share of co-authored and single-authored publications for senior authors in 16 SSH disciplines and two general categories (2000-2011)

Here as well, diversity at the disciplinary level is considerable, with a clear divide between the social sciences and the humanities. In the 2000-2011 time frame, most disciplines belonging to the humanities still show a stronger preference for single authorship, while in the social sciences co-authorship is far more prevalent. Among the social sciences, co-authorship occurs most frequently in Social Health Sciences with 97.5% of its publications being co-authored, followed by Psychology (95.8%). The lowest share of co-authorship in the SS is noted for Criminology (66.1%), followed by Political Science (67.5%). The humanities disciplines most inclined to co-author publications are Archeology (88%) and Communication Sciences (76%). Least inclined are Theology (26.1%), followed by Literature (32.2%).

Summarizing results at the aggregation level of SSH disciplines, we conclude that publication patterns of senior authors in Flanders show considerable variation, but appear mostly congruent with the distinction made by the existing literature between the typical publication patterns of the social sciences and those of the humanities. For further interpretation of the results we refer to the *Discussion and conclusion* section below.

Though not the primary focus of the present article and requiring more formal statistical analysis, a comparison of the above statistics on senior authors with those on all VABB-SHW registered authors (Engels et al., 2012; Verleysen et al., 2014; Ossenblok et al., 2014), suggests a more

international (WoS-indexed, English-language) and more collaborative (co-authorship) publication profile for senior authors.

4.2. Diversity among authors

Algorithmically clustered according to the similarity of their publication patterns, the 1828 senior authors in the SSH are classified as belonging to two distinct groups, hereafter referred to as *Cluster One* and *Cluster Two* (Figure 3, black and grey, respectively; one dot = one author). Cluster One contains 926 authors, Cluster Two 902 . It is notable how Cluster One shows a more dense plotting of its data points than Cluster Two, which denotes a greater similarity of its constituent authors. Inversely, Cluster Two is the most heterogeneous of the two.

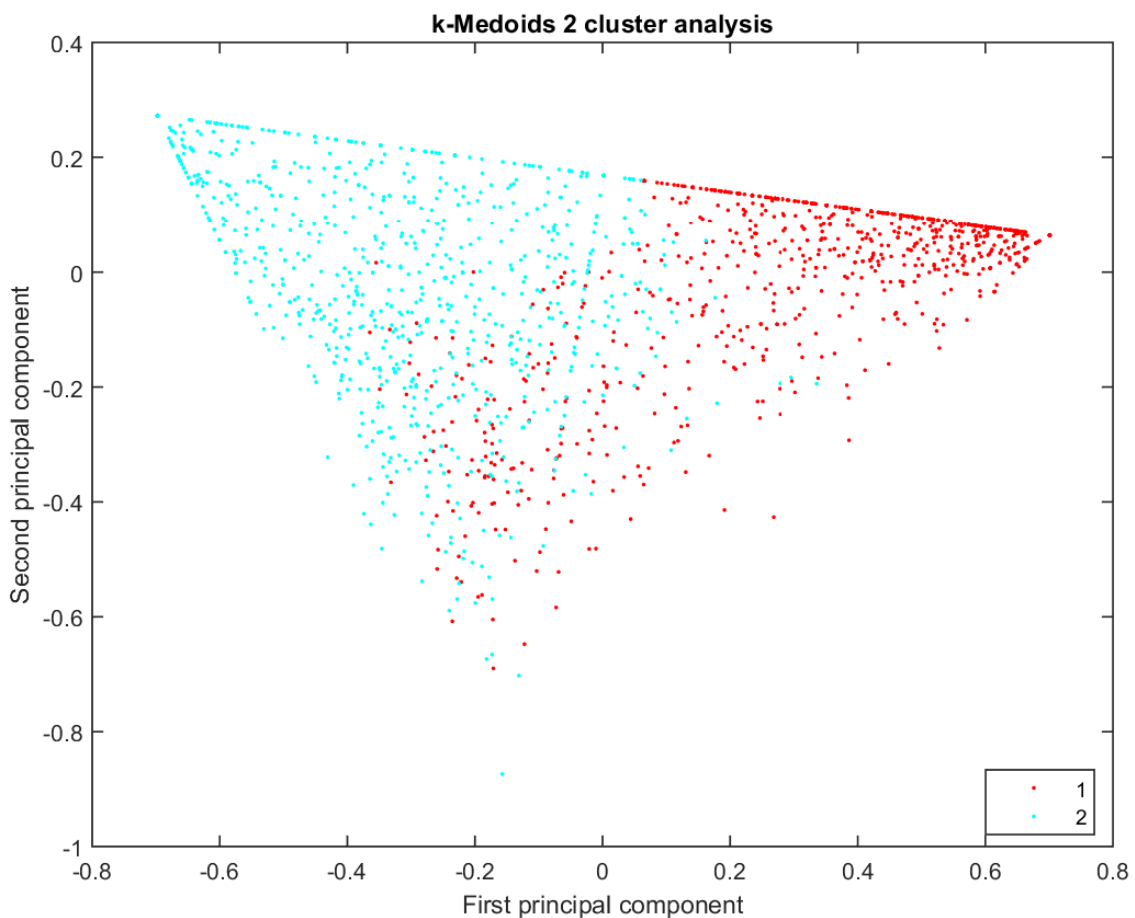


Figure 3: k-Medoids clustering of 1828 senior authors in the SSH

In order to label both clusters, we identify their respective medoids or most representative authors in the dataset. The medoid of Cluster One is an author whose main affiliation is with the discipline of Psychology. His/her publication pattern shows the following characteristics: a large share of WoS-indexed journal articles (82.6%), a smaller share of GP-approved journal articles

(14.1%), a modest share of book publications (an aggregated share of 2.7%), and a 100% share of publications in co-authorship. Cluster One, we may summarize, groups those authors who in a majority of their publications target an international audience of specialized academia, through the collaborative publication of mainly English language journal articles in high-profile journals, a large share of which are indexed by the WoS.

The medoid of Cluster Two is an author whose main affiliation is with the discipline of Philosophy. His/her publication pattern is characterized by a high share of GP-approved journal articles (69.2%), a lower share of WoS-indexed journal articles (23%), a relatively high share of book publications (aggregated share of 7.7%), and a lower share of publications in co-authorship (15.3%). Cluster Two thus groups those authors who are more strongly oriented towards national journals and also book publications, make more frequently use of other languages than English (mainly Dutch), and are less often inclined to co-author publications.

When the two clusters are re-plotted for only those authors affiliated with a specific discipline, it becomes apparent how the division between the two publication patterns cuts across SSH disciplines. Below are two examples of such plots, the first for political scientists (social scientists), the second for philosophers (humanities scholars). Plots for the other 14 disciplines and the two general categories are available in the appendix.

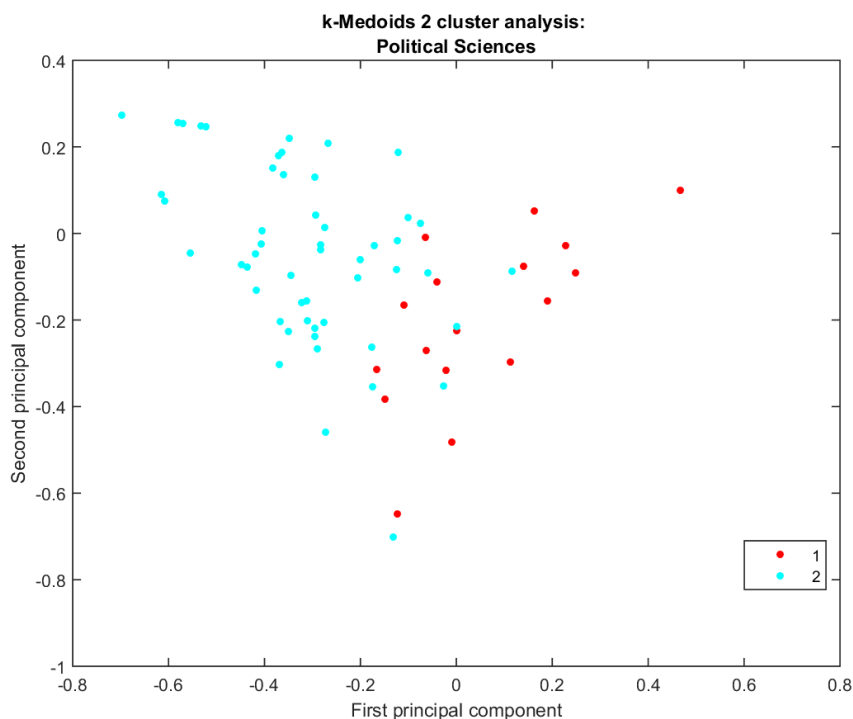


Figure 4: k-Medoids clustering of senior authors in Political Science (n=74)

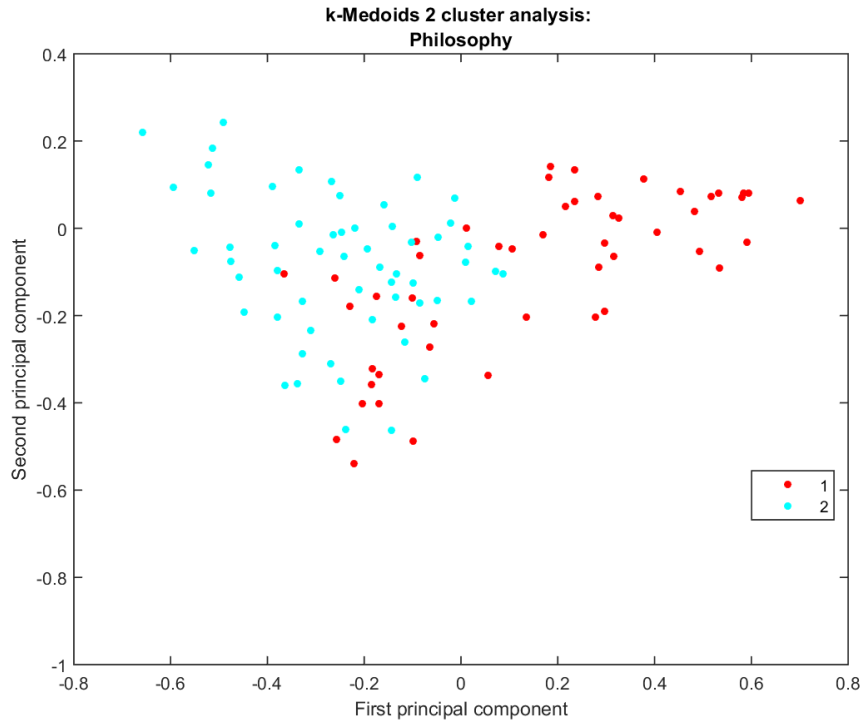


Figure 5: k-Medoids clustering of senior authors in Philosophy (n=108)

Summarizing the results of the clustering for all individual disciplines, Table 2 shows for each discipline the percentage of authors belonging to either of the two clusters. This demonstrates the considerable heterogeneity among authors within most disciplines.

Table 2: % of senior authors per discipline belonging to Cluster One or Cluster Two

| Discipline | n | % Cluster One | % Cluster Two |
|------------------------|-----|---------------|---------------|
| Archaeology | 32 | 40.6 | 59.4 |
| Communication Sciences | 35 | 25.7 | 74.3 |
| Criminology | 39 | 2.5 | 97.5 |
| Economics | 271 | 60.5 | 39.5 |
| Educational Sciences | 96 | 44.8 | 55.2 |
| History | 66 | 27.2 | 72.8 |
| History of Arts | 35 | 37.1 | 62.9 |
| Law | 206 | 3.8 | 96.2 |
| Linguistics | 121 | 47.1 | 52.9 |
| Literature | 62 | 51.6 | 48.4 |
| Philosophy | 108 | 47.2 | 52.8 |
| Political Sciences | 74 | 22.9 | 77.1 |
| Psychology | 199 | 90.9 | 9.1 |
| Social Health Sciences | 228 | 88.1 | 11.9 |

| | | | |
|------------|-----|------|------|
| Sociology | 57 | 47.7 | 52.3 |
| Theology | 54 | 18.5 | 81.5 |
| SS General | 102 | 66.7 | 33.3 |
| H General | 43 | 34.9 | 65.1 |

5. Discussion and conclusion

An analysis at the disciplinary or at the author level yields a different picture of publication patterns in the Social Sciences and Humanities (SSH). Publication statistics at the level of disciplines can be suggestive of a binary allocation of disciplinary groups to one of two publication styles: the first, often associated with the social sciences, focuses on co-authored English language articles in international, high-profile journals targeting an audience of specialized academia, while the second, associated with most humanities, is more strongly oriented towards single-authored book publications and articles in national or regional journals, makes use of other languages besides English, and targets a mixed audience of academia and interested lay readers. Clustering individual senior authors shows, however, that this division in fact cuts across all SSH disciplines in Flanders to a greater or lesser degree. While the authors of four disciplines are heavily concentrated in only one of the two clusters (Psychologists and Social Health Scientists in Cluster One; Criminologists and Law scholars in Cluster Two), authors affiliated with all other disciplines show a far more dispersed pattern. We conclude that a dichotomous classification of research and publication practices of the social sciences as opposed to those of the humanities is in most cases false; as well as that individual researchers affiliated with the same discipline can show substantially varying publication patterns.

These findings at the author level further substantiate and add detail to existing knowledge on the interaction between scholars and their target audiences. For Flanders, a recent analysis of places of publication and language in 12 SSH disciplines of scholarly books, peer reviewed and non-peer reviewed, identified two distinct target audiences: international academia on the one hand, and a mostly domestic intelligentsia on the other (Verleysen and Engels, 2014). Similarly, German political scientists belong to two networks: a local one, where publications are mostly in German and in regional, non-WoS indexed journals and in monographs, and an international one, where English and international journals predominate (Chi, 2015).

The principle explanation for intra-disciplinary heterogeneity in the SSH is the intrinsic diversity of many aspects of scholarly research. Most humanities and several social sciences are not only deeply fragmented with regard to target audience, but also to intellectual interest and approach, and conceptions of standards (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, for instance, research collaboration and co-authorship for publishing

journal articles is more easily achieved than in fields where qualitative methods are the norm (Kyvik, 1991; Moody, 2004).

A discussion of two telling examples from our clustering results can further illustrate the relation between specialization within disciplines and scholars' choice of publication channel. We will first discuss Sociologists, followed by Literature scholars.

The publication practices of senior Flemish sociologists show a distinctive clustering pattern, with 47.7% belonging to Cluster One and 52.3% to Cluster Two (cfr. Table 2). The cluster plot itself is also suggestive of a divided discipline in terms of publication preferences (cfr. appendix, figure 15). Specialization does indeed explain this division to a large extent. A sociological study from 2010 on publication practices in Flemish Sociology has found that, in more recent years, some communities of sociologists in Flanders have -somewhat belatedly compared to e.g. their Dutch colleagues- started a more active participation in international communication networks (Vanderstraeten, 2010). At the disciplinary level, this is attested to by a rise in WoS-indexed journal articles (Engels et al., 2012) and a growing share of English-language books published by international academic publishing houses (Verleysen & Engels, 2014). However, other communities of sociologists in Flanders have retained a more traditional focus on studies at the national or regional level, with articles mainly in three Dutch-language journals published in the Flanders or the Netherlands, which retain a relatively strong national profile and hardly attract an international authorship or readership (Vanderstraeten, 2010).

The clustering result for Literature, with a 51.6% and 48.4% distribution between Cluster One and Cluster Two (cfr. Table 2) is one of the more surprising among the 16 SSH disciplines. However, while one would expect Literature to have a relatively strong national focus in terms of publication channels and language use (the importance of object language), a substantial number of Flemish Literature scholars actually show a distinct international orientation. To corroborate our clustering result, we retrieved all 62 author names from the Literature subset, their main affiliation and publications from the websites of the five universities, in order to get an indication of their field of specialization. 17 researchers (27.2%) were found to belong to the subfield of Flemish-Dutch literature, which indeed corresponds to a more local orientation in terms of choice of publication channels. However, 20 (32%) are involved in British-American Literature and Literary Theory, corresponding with a much broader set of available international publication outlets. 9 (15.2%) are specialists of Latin-Greek literature, for which to a certain extent the same applies, whereas 16 (25.6%) are specialists of either French, German, Italian or Spanish literature. Here, the object language is likely to be important in publications, though for all of these four subfields the available number of international peer-reviewed publication outlets is still significantly larger than for a smaller language like Dutch. A second element explaining the result for Literature is the occurrence of co-authorship, one of the 11 variables used for our cluster

analysis. Senior Flemish Literature scholars are markedly more involved in research collaboration and ensuing co-authorship than has been found for their colleagues in other countries. Whereas in Canada, for instance, for 1980-2002, only 4% of Literature publications were co-authorships (Larivière et al., 2006), in Flanders, for all Literature authors in 2000-2010, this is 18% (Ossenblok et al., 2014), while for only the senior authors clustered in the present study (2000-2011), it is 32.2%.

Similar analyses of specialization could be made for most other SSH fields in Flanders. Added to this, from an organizational point of view, most SSH research at Flemish universities does not have a long or deeply embedded tradition of focused research groups which could in theory serve to coordinate and streamline larger research topics and projects. Instead, most SSH research is conducted in small, evolving units with a comparatively low degree of research coordination. Such organizational fluidity further facilitates intellectual fragmentation (Whitley, 2000).

There is also an exogenous element that can further explain the divide between individual authors' publication patterns: the national or local incentive structure for publishing academic research. In theory, where an incentive structure is in place, as is the case in Flanders, universities, their faculties/departments, research units and researchers are likely to respond in some way or another. So far, however, the evidence on the behavior-altering effects of incentives at the macro-level for publishing is mixed (Kyvik, 2003; Butler, 2010; Ossenblok, Engels, & Sivertsen, 2012; Hicks, 2013; Aagaard, Bloch, & Schneider, 2015). At lower levels, the extent to and the precise way in which institutional incentives trickle down to units and individual researchers remain largely unexplored, although it seems likely that researchers belonging to the same discipline or department can respond in varying ways to the incentives offered (Hammarfelt & de Rijcke, 2015). In the end, each individual makes his own balance of the inherent qualities and prestige of certain publication channels versus the rewards offered by the system for publishing in a specific way. In Flanders, the incentives to SSH scholars in the form of the VABB-SHW have only been in place since 2010 (Verleysen et al., 2014), which seems hardly likely to account by itself for the diversity among the authors clustered in the present study. However, already since 2003, publishing in WoS-indexed journals was incentivized by the Flemish funding model of that time (Ossenblok et al., 2012). As publishing in the WoS by scholars affiliated in Flanders has been shown to be strongly correlated with both the use of English (Engels et al., 2012) and co-authorship (Ossenblok et al., 2014), some influence of the incentive structure on the publication patterns of SSH scholars clustered in the present study is a distinct possibility. To assess the extent to which this has been the case, additional, both quantitative and qualitative research is called for.

There are a number of limitations to the method and results presented in this article. First, the classification of disciplines of the VABB-SHW is solely based on departmental author affiliation.

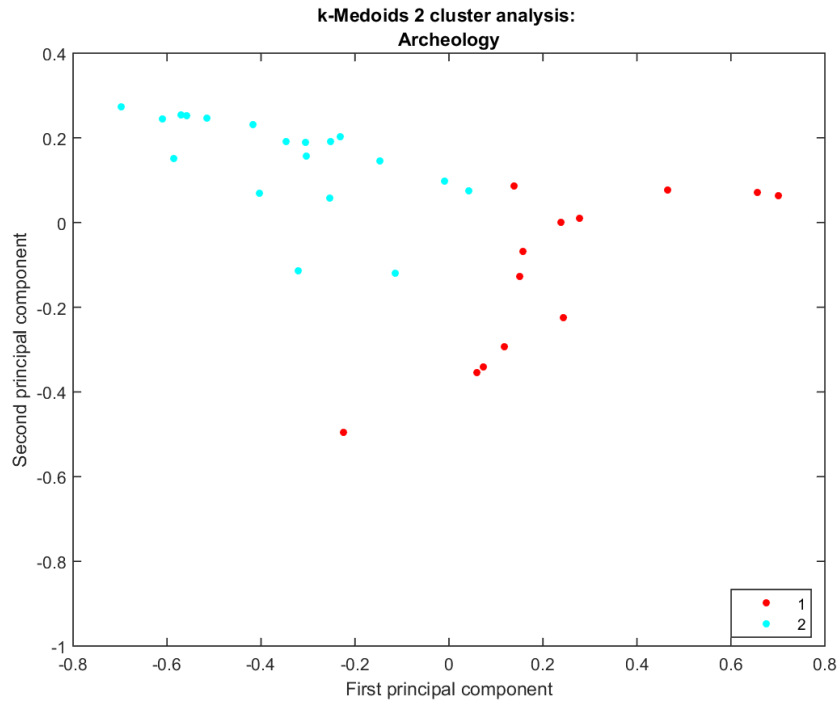
This has consequences for the clustering result for individual disciplines, as not all academics affiliated with a university department can be expected to have been trained in a corresponding field, nor to share its main research interests or publication preferences. It is therefore likely that an alternative (cognitive) classification into disciplines of Flemish SSH authors would yield a somewhat different result. Such a classification would also better account for the publication patterns of newly emerging transdisciplinary fields known for their international orientation. Second, the analysis in this article is based only on academic, peer reviewed publications. Surely, the ratio between the two clusters -which themselves would not exist in their current form- would be different were non-peer reviewed publications included as well. Especially for some humanities disciplines, where communicating with non-peers in society at large forms an important part of scholarly output, and where the distinction between academic and enlightenment literature remains a more gradual one than is the case for most social sciences (Verleysen & Engels, 2014), the picture obtained would likely not be the same. At the present stage, however, including VABB-SHW data on non-peer reviewed publications would result in a number of methodological and interpretative problems. As they are not counted for inclusion in the Flemish funding model, not only is data quality of non-peer reviewed publications in the VABB-SHW much lower, but the published material itself is also highly heterogeneous and of unequal academic stature. For our current analysis, their inclusion would therefore clutter the picture of publication patterns in the SSH rather than refine it in a meaningful way. A third limitation relates to the specific clustering techniques used. In the present study, a *hard* partitioning algorithm was used, i.e. an algorithm that allocates each author to a single cluster. In order to label the two clusters we have looked at the publication patterns of their respective most representative authors. While this has the advantage of clarity, we must state that the publication patterns of many scholarly authors should realistically be viewed as located somewhere on a scale of belonging between both clusters' extremes. For many disciplines the visual plots suggest that a number of their authors could be located near the thresholds separating both clusters and their underlying publication patterns. Further experimentation with regard to the applicability to SSH bibliometrics of *fuzzy* clustering algorithms, resulting in cluster membership coefficients (Kaufman & Rousseeuw, 1990), could therefore yield additional, more nuanced insights on diversity in scholarly publication patterns.

On the whole, as a method for research on SSH publication practices, cluster analysis (CA) shows promise. Clustering techniques also seem well applicable to quantitative aspects of research evaluation in the SSH at lower aggregation levels. As a factual bottom-up approach, CA could contribute to a sound empirical basis for research evaluation exercises in varying contexts, both with regard to the identification of the proper level or unit of assessment, as well as to that of the criteria (outputs) to be assessed in evaluations. As such, CA could serve as a complement to other, more qualitative bottom-up approaches to the development of procedures and criteria for research evaluation in the SSH (Ochsner, Hug, & Daniel, 2013). Wherever bibliographic data on

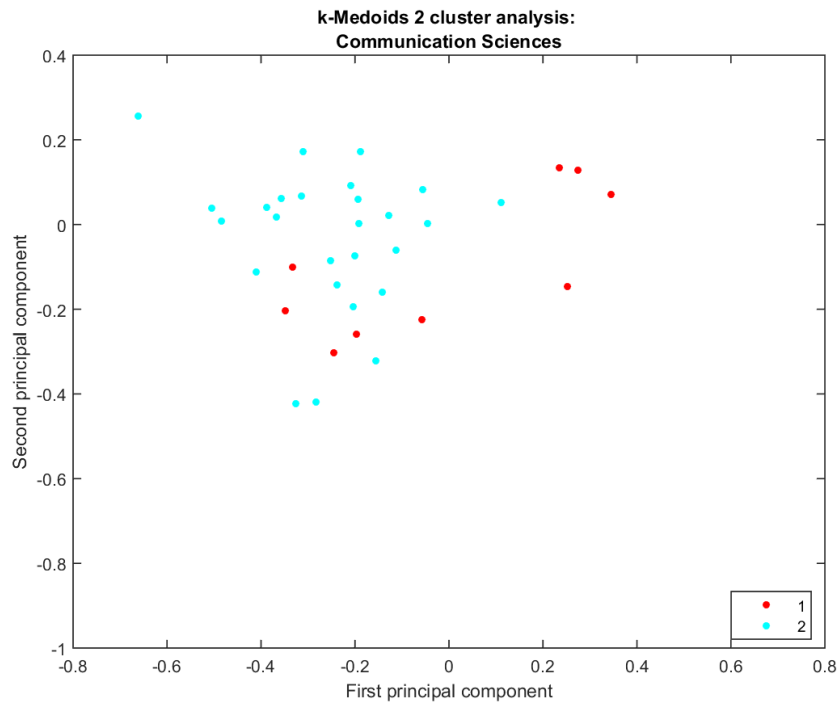
publication patterns of a sufficient volume, quality and granularity is available, application of CA is conceivable. Methodologically it is essential that the attributes of the cases (authors or other units) to be clustered are conceptually derived from the actual context, i.e. the scholarly research environment, to which CA is to be applied. Attributes which sharply differentiate publication patterns in the SSH in Flanders may or may not be applicable to other countries or institutions.

Acknowledgements

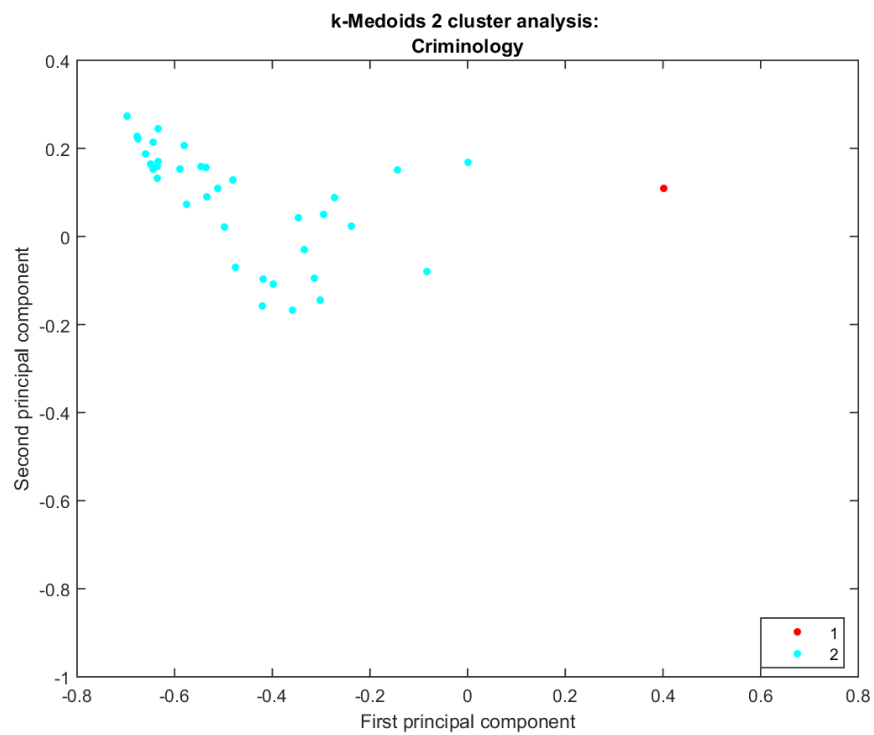
The authors thank Trycken Ossenblok for making available the dataset used in this article, its adaptation and initial processing. They also thank Tim Engels, Raf Guns, Ronald Rousseau and two anonymous reviewers for their valuable comments and suggestions. The Expertise Center for O&O Monitoring (ECOOM) is funded by the Flemish government.

Appendix

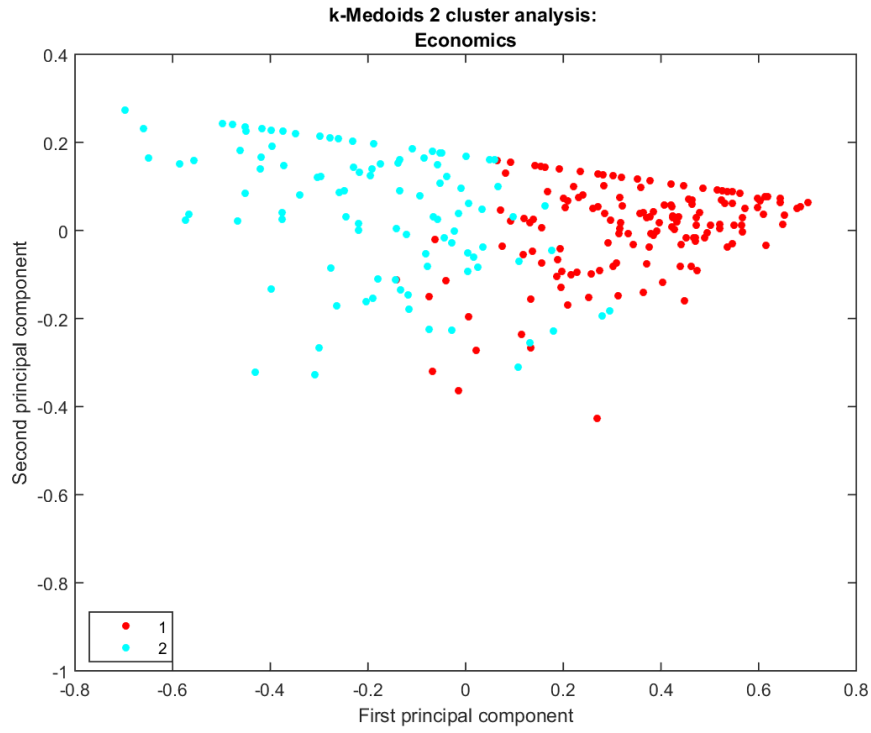
Appendix figure 1: k-Medoids clustering of senior authors in Archeology (n=32)



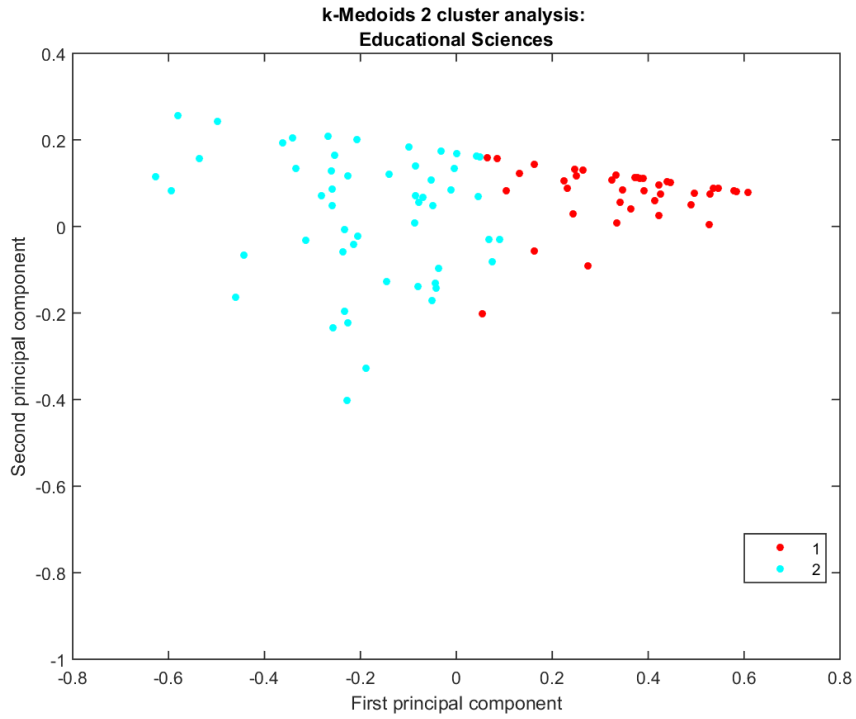
Appendix figure 2: k-Medoids clustering of senior authors in Communication Sciences (n=35)



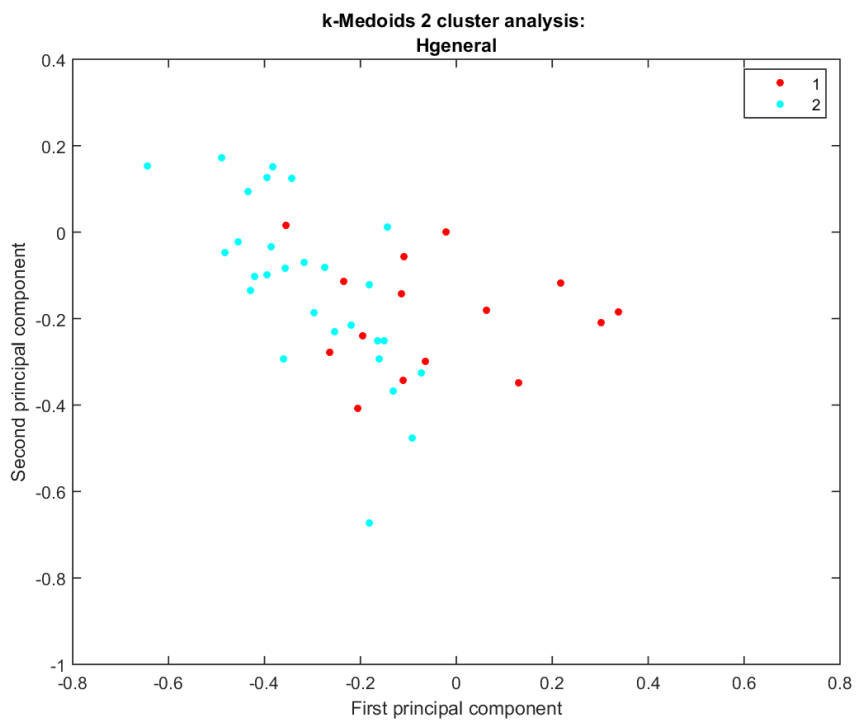
Appendix figure 3: k-Medoids clustering of senior authors in Criminology (n=39)



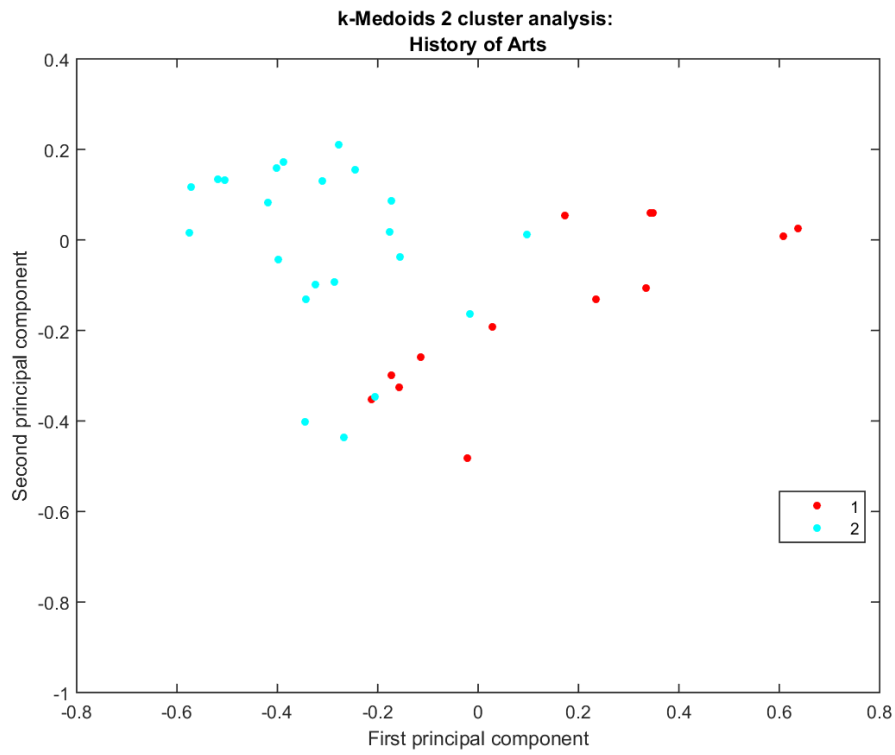
Appendix figure 4: k-Medoids clustering of senior authors in Economics (n=271)

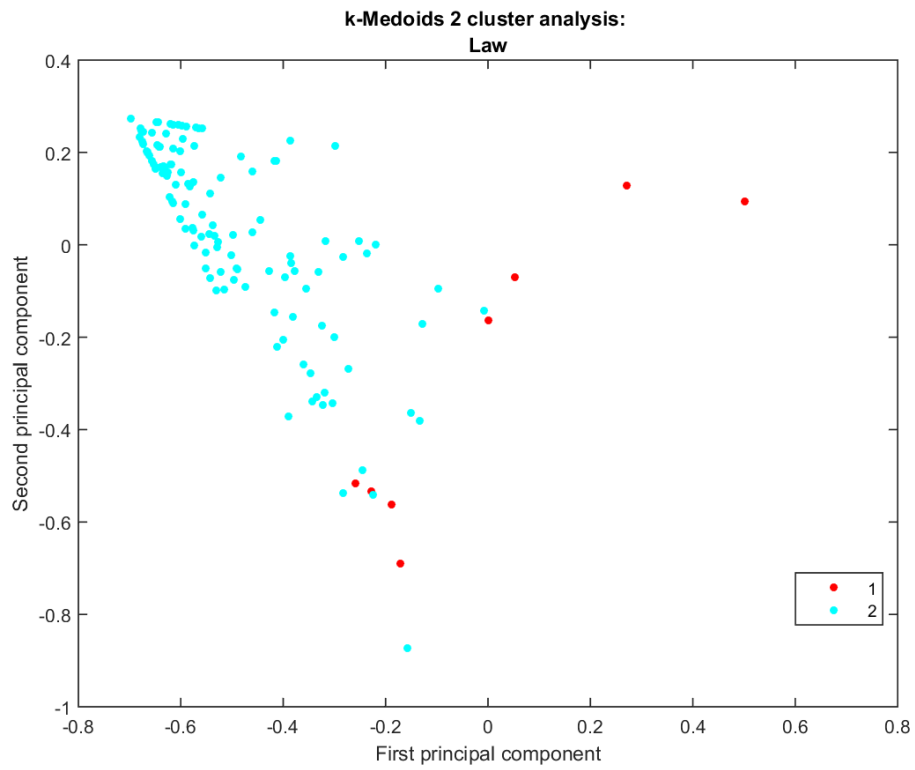


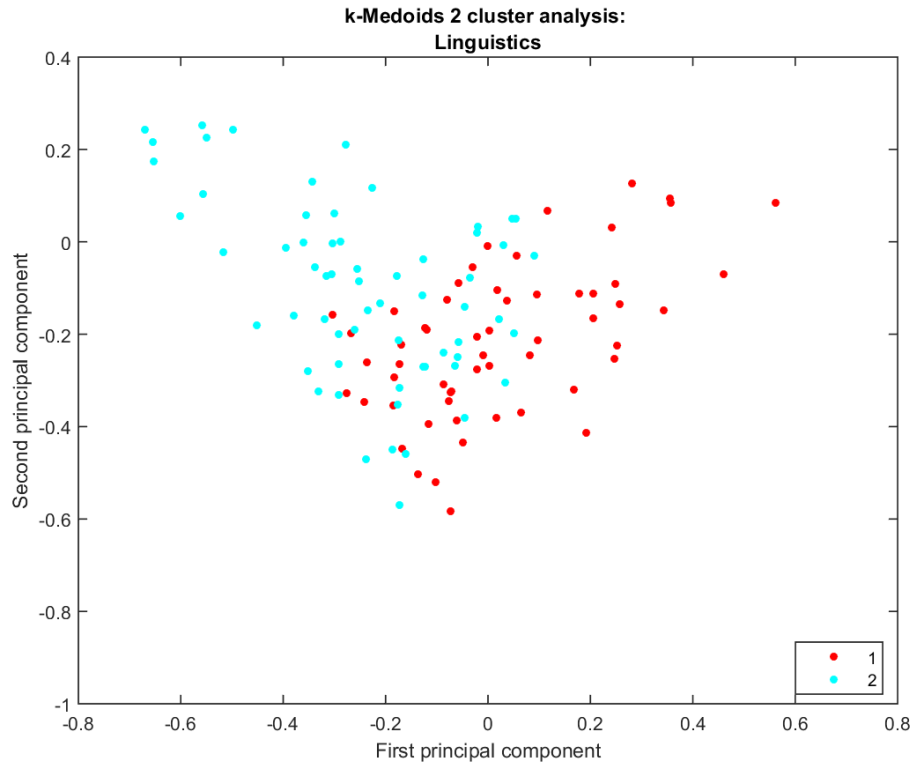
Appendix figure 5: k-Medoids clustering of senior authors in Educational Sciences (n=96)



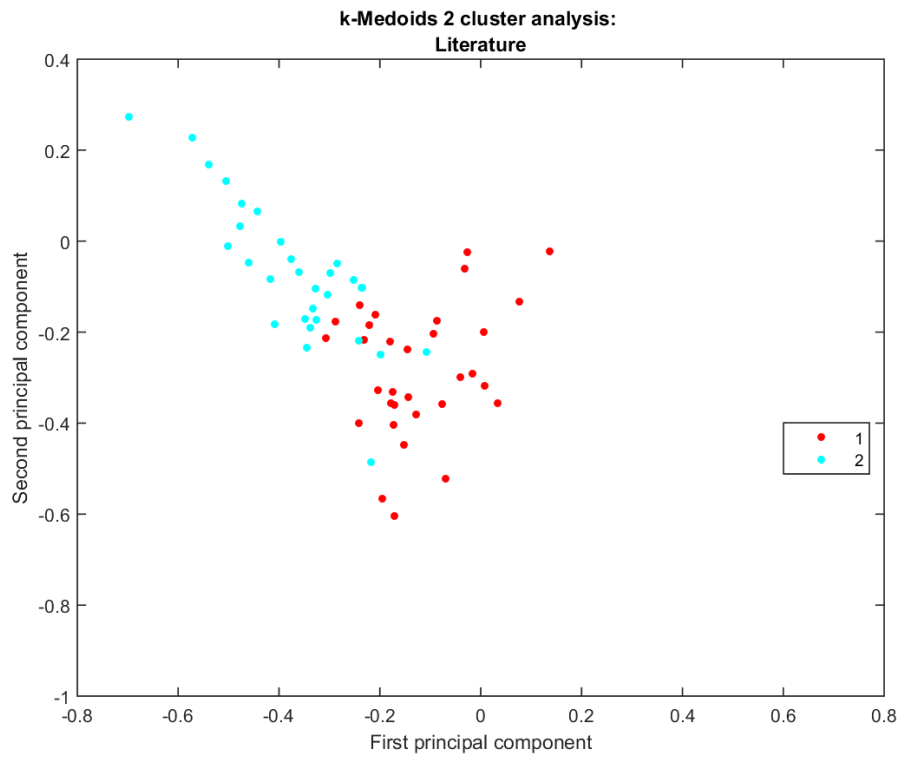
Appendix figure 6: k-Medoids clustering of senior authors in Humanities General (n=43)



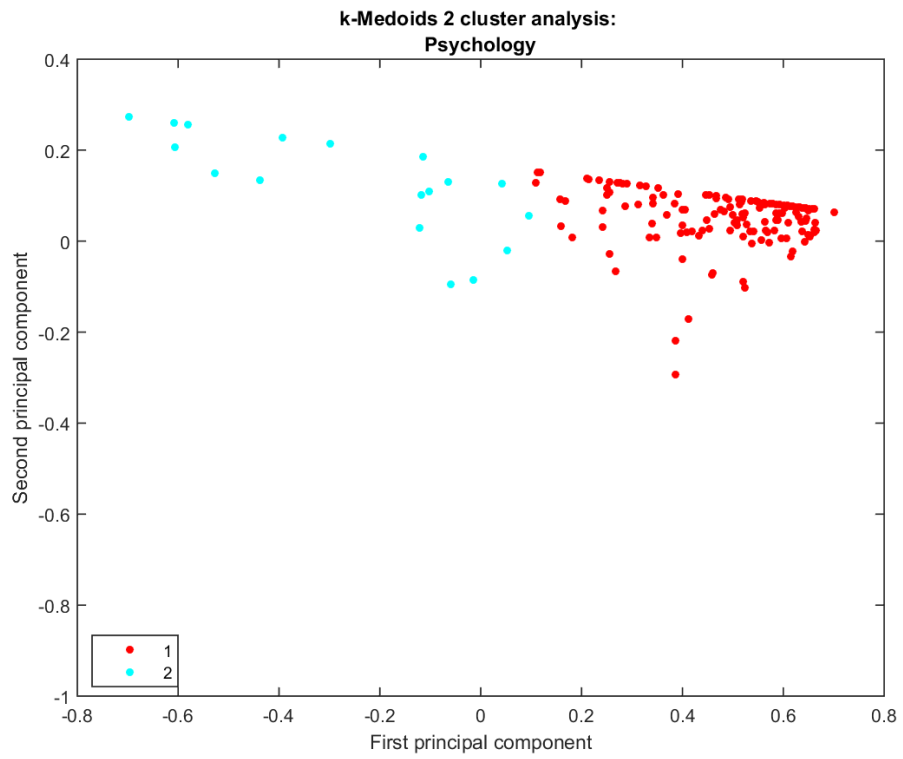




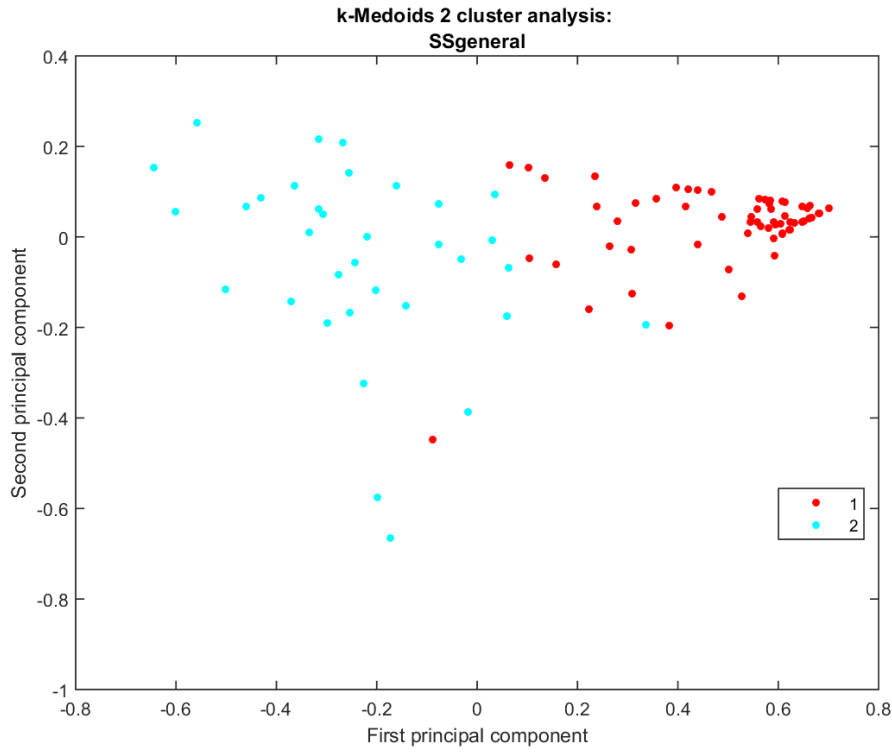
Appendix figure 10: k-Medoids clustering of senior authors in Linguistics (n=121)



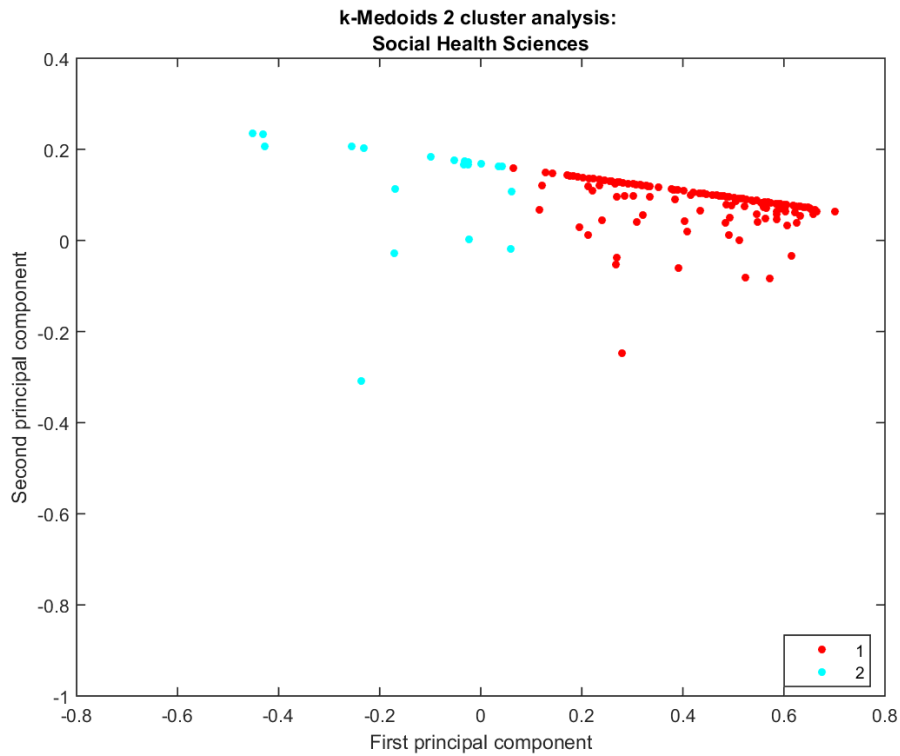
Appendix figure 11: k-Medoids clustering of senior authors in Literature (n=62)



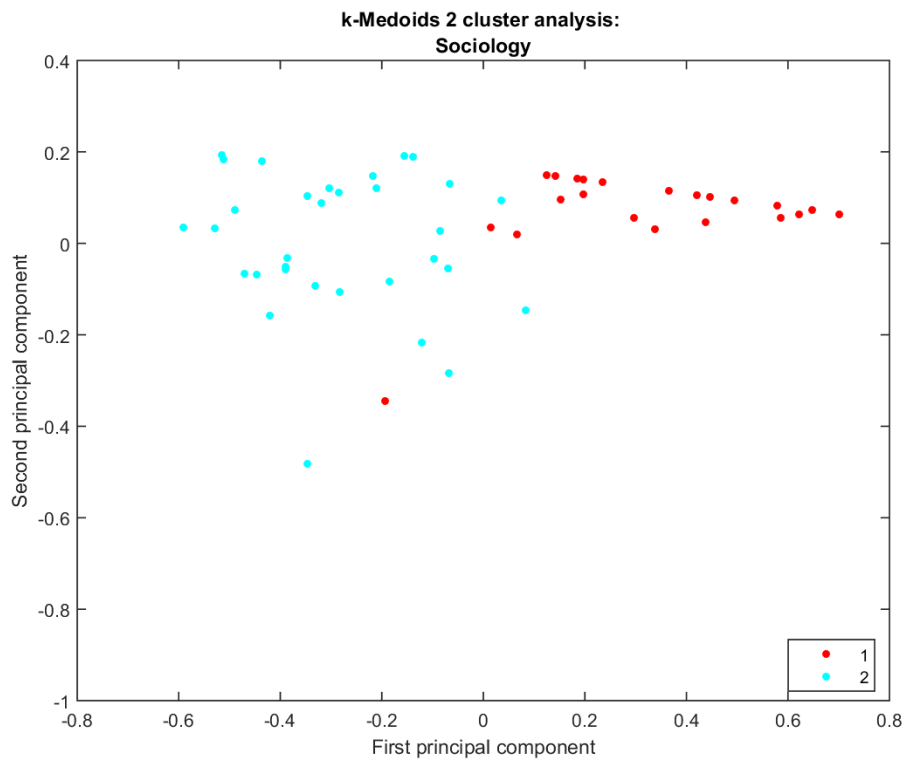
Appendix figure 12: k-Medoids clustering of senior authors in Psychology (n=199)



Appendix figure 13: k-Medoids clustering of senior authors in Social Sciences General (n=102)



Appendix figure 14: k-Medoids clustering of senior authors in Social Health Sciences (n=102)



Reference List

- Aagaard, K., Bloch, C. W., & Schneider, J. W. (2015). Impacts of performancebased research funding systems: the case of the Norwegian Publication Indicator. *Research Evaluation*, 24, 106-117.
- Ahlgren, P., Pagin, P., Persson, O. & Svedberg, M. (2015), Bibliometric analysis of two subdomains in philosophy: free will and sorites. *Scientometrics*, 103, 47-73.
- Bod, R., Maat, J., & Weststeijn, T. (2012). The making of the humanities. Volume II: from early modern to modern disciplines. Amsterdam University Press.
- Butler, L. (2010). Impact of performance-based research funding systems: A review of the concerns and the evidence. In OECD (Ed.), *Performance-based funding for public research in tertiary education institutions: Workshop proceedings* (pp. 127-165). Paris: OECD-publishing.
- Chi, P.-S. (2015). Changing publication and citation patterns in political science in Germany. *Scientometrics*, 105, 1833-1848.
- Egghe, L. & Rousseau, R. (1990). *Introduction to informetrics. Quantitative methods in library, documentation and information science*. Amsterdam: 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, 373-390.
- Hammarfelt, B. & de Rijcke, S. (2015). Accountability in context: Effects of research evaluation systems on publication practices, disciplinary norms and individual working routines in the faculty of Arts at Uppsala University. *Research Evaluation*, 24, 63-77.
- 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.
- Hicks, D. (2013). One size doesn't fit all: on the co-evolution of national evaluation systems and social science publishing. *Confero*, 1, 67-90.
- Johnson, R. A. & Wichern, D. W. (1992). *Applied multivariate statistical analysis*. (Third edition ed.) New Jersey: Prentice Hall.
- Kaufman, L. & Rousseeuw, P. J. (1987). Clustering by means of medoids. In Y.Dodge (Ed.), *Statistical analysis based on the L₁-norm and related methods* (pp. 405-416). North-Holland.
- Kaufman, L. & Rousseeuw, P. J. (1990). *Finding groups in data: an introduction to cluster analysis*. New York: Wiley.

- Knorr Cetina, K. (1999). *Epistemic cultures. How the sciences make knowledge*. Cambridge, Massachusetts: Harvard University Press.
- Kyvik, S. (1991). *Productivity in academia: Scientific publishing at Norwegian universities*. Trøgstad, Norway: Norwegian University Press.
- Kyvik, S. (2003). Changing trends in publishing behaviour among university faculty, 1980-2000. *Scientometrics*, 58, 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.
- Liu, J., Li, J., Xu, W., & Shi, Y. (2005). Data mining approach in scientific research organisations evaluation via clustering. In *Data mining and knowledge management*. (pp. 128-134). Springer.
- Mahalanobis, P. C. (1936). On the generalized distance in statistics. In (pp. 49-55).
- 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, 81-100.
- Nederhof, A.J. (2011). A bibliometric study of productivity and impact of modern language and literature research. *Research Evaluation*, 20(2), 117-129.
doi:10.3152/095820211X12941371876508.852
- Nederhof, A.J. & Noyons, E.C.M. (1992). International comparison of departments' research performance in the humanities. *Journal of the American Society for Information Science & Technology*, 43,3, 249-256.
- Ochsner, M., Hug, S. E., & Daniel, H. D. (2013). Criteria for assessing research quality in the humanities: a Delphi study among scholars of English literature, German literature and art history. *Research Evaluation*, 22, 369-383.
- Ossenblok, T. L. B., Engels, T. C. E., & Sivertsen, G. (2012). The representation of the social sciences and humanities in the Web of Science. A comparison of publication patterns and incentive structures in Flanders and Norway (2005-9). *Research Evaluation*, 21, 280-290.
- Ossenblok, T. L. B., Verleysen, F. T., & Engels, T. C. E. (2013). Co-authorship of journal articles and book chapters in the social sciences and humanities (2000-2010). *Journal of the American Society for Information Science & Technology*, 65, 882-897.
- Puuska, H.-M. (2014). *Scholarly publishing patterns in Finland. A comparison of disciplinary groups*. School of Information Sciences, University of Tampere, Finland.
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the validation and interpretation of cluster analysis. *Computational and applied mathematics*, 20, 53-65.

- 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). ISSI.
- Sivertsen, G. & Larsen, B. (2012). Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: an empirical analysis of the potential. *Scientometrics*, *91*, 567-575.
- Small, H., Sweeney, E., & Greenlee, E. (1985a). Clustering the Science Citation Index using co-citations I. A comparison of methods. *Scientometrics*, *7*, 391-409.
- Small, H., Sweeney, E., & Greenlee, E. (1985b). Clustering the Science Citation Index using co-citations II. Mapping Science. *Scientometrics*, *8*, 321-340.
- Vanderstraeten, R. (2010). Scientific Communication: Sociology Journals and Publication Practices. *Sociology*, *44*(3), 559-576.
- van Eck, N. J. & Waltman, L. (2014). Visualizing bibliometric networks. In Y.Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact* (pp. 285-320). Springer International.
- Verleysen, F. T. & Engels, T. C. E. (2014). Internationalization of peer reviewed and non-peer reviewed book publications in the social sciences and humanities. *Scientometrics*, *101*, 1431-1444.
- Verleysen, F. T., Ghesquière, P., & Engels, T. C. E. (2014). The objectives, design and selection process of the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (VABB-SHW). In W.Blockmans & al. (Eds.), *The use and abuse of bibliometrics* (pp. 115-125). Academiae Europaea; Portland Press.
- Waltman, L., van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, *4*, 629-635.
- Whitley, R. (2000). *The intellectual and social organization of the sciences*. Oxford: Oxford University Press.
- Wicklin, R. (2015). What is Mahalanobis distance? Blogs.sas.com.
<http://blogs.sas.com/content/iml/2012/02/15/what-is-mahalanobis-distance.html>. Date last accessed: 21 October 2015.