



Categorizing authors in co-authorship networks through textual information analysis: a comparative study of two scientific communities

Francesco Santelli² · Francesco Capone¹ · Domenico De Stefano²  · Alice Tontodimamma² · Luciana Lazzeretti¹

Received: 26 March 2025 / Accepted: 24 June 2026 / Published online: 25 July 2026
© The Author(s) 2026

Abstract

Scientific collaboration is a key driver of research productivity and innovation, enabling researchers to share knowledge, leverage expertise, and improve research outcomes. While co-authorship is commonly used to represent collaborative behaviour, most studies focus on specific single scientific fields or national networks. A significant limitation of these studies is the lack of comprehensive methods to classify all authors within co-authorship networks, particularly in relation to their field or macro-disciplinary affiliations. This study addresses these limitations by analysing co-authorship networks within two distinct Italian scientific target groups—Statistics and Management. Using a dataset of publications from 2012 to 2022, we develop a novel methodology that combines bibliographic and textual data, such as keywords and subject areas, to classify all authors in these communities. By constructing a two-mode network and applying a clustering algorithm, we categorize authors into distinct thematic areas, offering a more nuanced analysis of collaboration. Additionally, we compare the collaborative structures of the two target groups, revealing important similarities and differences in their research dynamics. This approach provides new insights into the evolution of scientific collaboration and offers a deeper understanding of how researchers from different fields engage in knowledge production.

Keywords Co-authorship · Two-mode networks · Textual analysis · Community detection

Introduction

In recent years, scientific collaboration has become a key driver of research innovation and productivity. Collaborations enable researchers to share ideas, pool resources, and benefit from collective expertise, thereby enhancing both the quality and quantity of research output (Beaver, 2001; Glänzel & Schubert, 2004). As a multifaceted phenomenon, scientific

✉ Domenico De Stefano
ddestefano@units.it

¹ University of Florence, Firenze, Italy

² University of Trieste, Trieste, Italy

collaboration can be examined through various forms of interaction, including informal exchanges (Olmos-Peñuela et al., 2014; Casaluze & D'Agata, 2025) and scientific support acknowledged in publications (Petrovich, 2022), joint participation in research projects (Fabbrucci Barbagli et al., 2025b), mentorship relationships (Cugmas et al., 2024), and other types of ties.

Within this framework, the most tangible and widely used proxy of collaboration in the literature is collaborative scientific production, as captured by bibliographic records. Based on these data, collaboration structures are typically modeled from a network perspective, where authors represent the nodes and the observable links between them represent their joint participation to one or more publications. This results in the construction of a proper author-by-author co-authorship network (Newman, 2001a, 2001b, 2004).

These networks of scientific collaboration are more than a product of shared effort; they facilitate the dissemination of knowledge across disciplines and international boundaries, making the entire research ecosystem more dynamic and robust (Lee & Bozeman, 2005). Moreover, scientific collaboration has been bolstered by government policies and initiatives, such as international exchange programs and funding incentives, which emphasize researcher mobility and interdisciplinary partnerships (Defazio et al., 2009). As a result, collaboration has become a significant metric in evaluating scientific performance, illustrating its essential role in modern research (De Stefano & Zaccarin, 2016; De Stefano et al., 2023).

Despite the growing emphasis on collaboration, a key challenge in the literature is that most studies examine co-authorship networks without delving deeply into the complex relationships and diversity within these networks (Newman, 2004; Ponomariov & Boardman, 2016). Co-authorship is commonly used as a proxy for collaboration, but it often only scratches the surface of the nuanced ways in which scholars interact. While co-authorship networks provide valuable insights into the structural dimensions of collaboration, they lack granularity in understanding the various types of collaborators or their roles (Moody, 2004; Yan & Guns, 2014). Furthermore, most studies are limited to examining single specific scientific field, leaving a gap in the comparative analysis of multiple research communities and the classification of authors based on their collaborative behaviours and fields of expertise (Savić et al., 2019).

A further limitation of existing studies that analyze the collaborations of a specific target group (e.g., scholars belonging to a given disciplinary field, country, or institution) concerns the lack of information on external co-authors within co-authorship networks. External authors are those co-authors who do not belong to the predefined target group of scholars under analysis for various reasons but who appear in the dataset through collaborative ties with such target group (see Sect. “Literature review” for a more precise definition).

While these external actors are expected to be positioned at the margins of the network, in real settings they may play a crucial role in shaping overall connectivity and facilitating links across different scientific domains (Pan & Saramäki, 2012).

Scientific collaboration frequently spans departments, institutions, and countries, making these external contributors essential for understanding the broader structure of knowledge exchange. However, a significant challenge arises from the limited availability of detailed information about them: while complete data (such as affiliation, role, and scientific field) are typically available for the target scholars, external co-authors are often described only through paper-related metadata (e.g., keywords, citations, and subject areas). This lack of auxiliary information hampers the ability to fully characterize their role within the network and limits the interpretability of collaboration patterns. Consequently,

there is a need for approaches that can effectively leverage available bibliometric data to better classify and understand these actors.

That is the reason why this work aims to identify an approach for dealing with a significant limitation of existing research in this context, that is the absence of comprehensive methods to classify all authors (including those not belonging to the target group under analysis) within these networks, particularly in relation to their field of research or disciplinary affiliations. Additionally, while the use of bibliometric data has made co-authorship networks easier to analyse, studies often fail to capture the full range of collaborations, including those involving non-core collaborators who play crucial roles in research but are outside the primary field of study (De Stefano et al., 2023).

In this study, we propose an approach to address these limitations by developing a method for classifying all authors retrieved from bibliographic data, including those not belonging to the target group under analysis, with respect to their field of research. The empirical application focuses on two distinct scientific communities, namely Italian Academic Statisticians and Management scholars (Capone & Lazzeretti, 2023; De Stefano et al., 2023).

Our primary goal is to use the information contained within bibliographic data, such as keywords and subject areas, to categorize each author in our dataset into specific thematic fields. By focusing on the entire publication output of these communities over a ten-year period (2012–2022), we aim to develop a novel methodology for the classification of authors that transcends the limitations of current network studies.

The analysis focuses on publications in the 2012–2022 time window. This choice is primarily motivated by data availability and consistency, as the Scopus/SciVal database provides more reliable and comparable coverage over a recent ten-year horizon. Limiting the observation period to this interval allows us to work with a dataset that is less affected by issues of incomplete indexing and changing coverage over time.

In addition, we consider this period sufficiently long to capture the recent evolution of collaboration patterns, including the emergence of new links with external authors and the development of research topics. The selected timeframe also aligns with significant changes in the Italian research evaluation system, such as the introduction and consolidation of policies related to the Valutazione della Qualità della Ricerca (VQR) (Research Quality Assessment) and the Abilitazione Scientifica Nazionale (ASN) (National Scientific Qualification), which may have influenced collaboration behaviors and publication strategies.

While earlier publications could provide insights into the historical formation of collaboration networks, they would likely reflect structures and institutional contexts that are less representative of current research dynamics. For this reason, we prioritize a more recent and homogeneous time window, while acknowledging that future research could extend the analysis to longer periods to investigate longitudinal changes in greater detail.

Furthermore, we seek to compare the collaborative behaviours and network structures of these two communities, providing valuable insights into the similarities and differences between them.

Our methodology involves constructing a two-mode network that integrates textual data (such as keywords and subject areas) with co-authorship data from the Scopus database. Then by means of a proper community detection algorithm we group all authors in our dataset into distinct thematic areas, which are the detected clusters or communities. The innovative aspect of our work lies in this multi-layered classification process, which to the best of our knowledge has not been applied in prior studies. By integrating textual data with network analysis, we move beyond the traditional focus on co-authorship alone,

providing a richer understanding of the collaborative structures within each target group. Our work provides fresh insights into the dynamics of scientific collaboration within and across disciplinary boundaries, contributing to the broader understanding of how scientific knowledge is produced and disseminated among the members of the two target groups under study. Finally, the paper represents the first attempt to use co-authorship information and bibliographic metadata to classify external co-authors to a specific target group of scholars.

The remainder of this paper is structured as follows. In Sect. “[Literature review](#)”, we provide an overview of co-authorship analysis studies relevant to the statistics and management communities, focusing on the existing challenges in classifying and analysing authors within these networks. Sect. “[Data and method](#)” details our data sources and methodological approach, including the construction of the two-mode network and the clustering algorithm used to categorize authors. Sect. “[Results](#)” presents the results of our analysis, including a comparison between the two scientific communities. Finally, Sect. “[Conclusion](#)” offers concluding remarks and suggestions for future research directions.

Literature review

Co-authorship analysis and scientific collaboration

In recent decades, co-authorship analysis has garnered increasing academic attention, leading to numerous fundamental contributions. Newman (2001a; Newman, 2001b) conducted one of the pioneering studies using social network analysis to examine the characteristics of large co-authorship networks in fields such as biology, medicine, physics, and computer science during the period from 1995 to 1999. Subsequently, Barabási et al. (2002) explored the dynamics and evolution of co-authorship networks in mathematics and neuroscience, introducing the concepts of preferential attachment and scale-free networks. Since then, co-authorship networks have been analyzed using various methodologies across multiple disciplines (Glänzel & Schubert, 2004; Kumar, 2015).

More recently, the literature has expanded to investigate aspects such as the impact of an author’s structural position within co-authorship networks on research performance, as well as the dynamics of publishing behavior (Cainelli et al., 2015; De Stefano et al., 2013; Fabbucci Barbagli et al., 2025a; Henriksen, 2016).

A key aspect of co-authorship network analysis is the identification of prominent authors who are strategically positioned, highly influential, and well-connected within the network. Notably, from this point onward, co-authorship analysis has been used to examine cohesion and connections within scientific communities (Capone & Lazzeretti, 2023; Kumar, 2015). This approach enables researchers to explore scientific communities and understand their publication patterns and evolution over time.

In this context, usually the interest is toward the analysis of interactions among members of one or more target scientific groups of scholars (and possibly bounded by countries or disciplinary affiliation) and the derived co-authorship networks may include different kinds of nodes, namely nodes representing scholars who belong of the target group and the so-called external co-authors who for several reasons do not belong to the specific target group. In this setting we can more precisely define what we intend as external co-authors. These represent nodes in the co-authorship networks that are individuals who fall outside the target population, including: (i) members of foreign scholarly communities;

(ii) non-tenured or non-permanent research personnel within universities (e.g., postdoctoral researchers, PhD candidates, and other non-faculty positions); (iii) researchers affiliated with non-university institutions (such as hospitals, public research organizations, and similar entities); (iv) scholars formally affiliated with Italian universities but belonging to disciplinary groups not included in the target population; and (v) individuals who were previously part of the target group but are no longer active during the observation period (e.g., retired scholars).

Co-authorship network analysis in statistics and management sectors

In this study, we concentrate on two specific disciplines, Management and Statistics scholars affiliated to Italian Universities in the period from 2012 to 2022. The focus on these two communities is based on both their commonalities and differences and the peculiarities of the Italian academic context make it an interesting case study.

Firstly, both communities belong to the same area in the official classification of the scientific areas provided by the Italian Ministry of University and Research (MUR).¹ Specifically, they are part of the broader *Economics and Statistical Sciences* category, namely Area 13.² As a result, it often happens that the members of both disciplines belong to the same departments, providing opportunities to share ideas, methodologies, and common practices. In 2012, scholars belonging to the Statistics and Management sectors accounted for 1,304 out of 4,802 researchers in Area 13 (27.2%). By 2022, this number increased to 1,573 out of 5,546 (28.4%). Overall, these two sectors consistently represent slightly more than one quarter of the scholars in Area 13.

However, these communities also differ in several key aspects, particularly in their publication practices and research outputs. The Italian Management community follows a publication style that lies between the social and natural sciences. They often publish scientific articles in national and international journals, usually in co-authorship, but also produce books (i.e., monographs), which are often authored by a single researcher (Capone & Lazzeretti, 2023). In contrast, the *Statistics* community follows a publication style more akin to the natural sciences, with journal articles being the predominant research output. Statistical researchers typically collaborate in groups, with an average of three to four co-authors per publication, often with an interdisciplinary composition. The number of co-authors and degree of interdisciplinarity vary across the different Statistics subfields (De Stefano et al., 2013, 2016).

Furthermore, the statistical community is of particular interest for several reasons on its own. Firstly, the field of Statistics holds a unique position within the scientific landscape. While its techniques, analytical formalization, and computational demands firmly place it within natural science disciplines, its broad collaborations with both natural and social sciences make it difficult to classify many statistical works as purely statistics-oriented. In some academic traditions, particularly in Italy, Statistics is even categorized within social sciences. However, its methodological versatility sets it apart from other disciplines, as statistical methods play a fundamental role across diverse fields (Cox, 1997; De Stefano et al., 2013). The Italian statistical landscape is characterized by five distinct subsectors, denoted by the acronym S.S.D (Settore Scientifico Disciplinare) at the time of the data collection of

¹ <https://www.mur.gov.it/>

² <https://cercauniversita.mur.gov.it/php5/settori/>

the present paper.³ In particular, the SSD shaping the Statistical academic community are the following: SECS-S/01-Statistics, SECS-S/02-Statistics for Experimental and Technological Research, SECS-S/03-Economic Statistics, SECS-S/04-Demography, SECS-S/05-Social Statistics.⁴

Each statistics SSD is characterized by its own trending topics and research areas, which can vary significantly. A key differentiating factor is the scientific background of collaborators, as it helps define the area of interest. External co-authors, ranging from natural science fields such as physics and engineering to medicine and social sciences, play a crucial role in shaping research directions. Moreover, in many studies, a single statistician collaborates with a larger group of researchers from other disciplines, making the classification of these external co-authors particularly relevant.

These differences in publication practices influence the structure of co-authorship networks (see Sect. “[Data and method](#)”).

Co-authorship studies have been conducted across a wide range of disciplines, including the above cited fields. Notably, significant research has focused on co-authorship in the broad area of the Social Sciences (De Stefano et al., 2011, 2013, 2017; Glänzel, 2002; Henriksen, 2016), particularly in fields such as economics and management (Casanueva & Larrinaga, 2013), Tourism (Hu & Racherla, 2008; Racherla & Hu, 2010) and Statistics (De Stefano et al., 2011, 2013).

The fields of Economics and Management have seen significant research into co-authorship dynamics. Casanueva & Larrinaga (2013) explored the concept of the “invisible college” (Crane, 1969) among Spanish accounting scholars by analysing the selection processes for Ph.D. panel members from 1994 to 2003. Their findings challenged the presence of invisible college dynamics, suggesting that high-profile scholars did not dominate the production of new publications, thus indicating the absence of preferential attachment in this academic community. Acedo et al. (2006) examined co-authorship trends in management and organizational studies, focusing on prominent international journals. They observed a rise in co-authored publications, reflecting a broader trend across various academic disciplines. Similarly, Beattie & Goodacre (2004) studied publishing patterns in the UK and Irish accounting and finance communities, using data from the British Research Register for 1998–1999. They noted that nearly two-thirds of articles were co-authored, with 25% involving contributors from outside the community. Merigó et al. (2016) took a different approach by conducting a bibliometric analysis of business and economics research using the ISI Web of Science database. Although they did not analyse co-authorship, they provided insights into the most cited papers, leading journals, influential institutions, and top-performing countries in the field. Podsakoff et al. (2008) examined factors influencing university and author impact within management literature over 25 years, identifying that a small number of institutions and scholars garnered most citations, reinforcing the idea of preferential attachment.

³ A new SSD classification has recently been adopted by the Italian Ministry of University and Research (starting from 2024). However, for the target scientific group analyzed in this paper, this change merely involves a modification of the identification code, while the organization of the subsectors remains unchanged.

⁴ It is worth noting that the management community is characterized by a single SSD, namely SECS-P/08 according to the official classification of the Italian Ministry of University and Research at the time of our data collection.

In Italy, co-authorship studies are less common. Early contributions include work by Cainelli et al., (2006, 2012; 2015) on co-authorship among economists. Plumper & Radaelli (2004) examined publications and citations in 89 political science journals from 1990 to 2002 for scholars affiliated with Italian institutions, though they did not focus specifically on co-authorship. De Stefano et al., (2013, 2017), De Stefano & Zaccarin (2016), and Fucella et al. (2016) were among the first to analyze co-authorship among the Italian statisticians, focusing on the internal network structures and their impact on academic performance. Menardi & De Stefano (2022) later emphasized the importance of prominent actors in shaping the community structures of the Italian academic statistician’s collaboration network.

Research specifically on Italian management scholars is even scarcer. Notable contributions include Lazzeretti (2001), who analysed the adoption of empirical statistical methodologies in management. Later, Lazzeretti et al. (2014) explored the “invisible college” within cluster research, mapping the key authors and historical developments of the field. Sedita et al. (2020) further utilized co-authorship analysis to study the evolution of major research themes within the broader community.

Based on the reviewed state of the art, several research gaps and challenges can be identified in the study of co-authorship and network data. Kumar (2015) highlights several issues in co-authorship data, including author name ambiguities, disambiguation problems, and the movement of scholars across institutions.

Furthermore, since this study focuses on scholars belonging to specific scientific disciplines—namely, Italian academic scholars in Statistics and Management—these issues are amplified. While we have information on the members of our target population, we lack additional details about their co-authors who do not belong to this group (whom we refer to as “external” co-authors). Despite this limitation, external co-authors may play a crucial role in shaping the overall co-authorship structure of a discipline. For instance, an external co-author may act as a bridge connecting two otherwise unconnected members of the target group. Additionally, external co-authors may provide insights into broader aspects of scientific collaboration, such as interdisciplinary tendencies or the emergence of new knowledge through evolving keywords and topics over time.

To the best of our knowledge, there is a lack of literature addressing the classification of these external co-authors or leveraging bibliographic metadata from their publications to analyze their role. In this study, we take the latter approach, proposing a method to classify and extract groups of keywords and subject areas associated with the publications of both target group members and their external co-authors. External here means co-authors not belonging to the target group under study but that appears as co-authors, often covering an important role on scientific advances of a discipline and its interdisciplinary characterization as well as to the overall connectivity of the networks. The present paper represents the first attempt to combine co-authorship information and bibliographic metadata to classify co-authors external to a specific target group of scholars.

Data and method

Dataset description: statistics and management target groups

We have built two distinct datasets: one pertaining to the Statistics group and the other to the Management target group. These two datasets were constructed in the following way:

first, all the names of the Statisticians/Management scholars have been downloaded from the MUR (Ministry of University and Research) at 31/12 of all years from 2012 to 2022, in order to acquire information also on retired scholars or on scholars who were affiliated with an Italian university only for a few years. Then all Scopus IDs were manually retrieved for each Statisticians/Management scholars and finally using the Scopus IDs it was possible to download all the documents published by statisticians and management scholars from 2012 to 2022. In this way we obtained the dataset of scientific collaborations for the Statistics and Management communities in a decade. The dataset includes different types of scientific outputs indexed in Scopus, such as journal articles, conference papers, and book chapters. While publication practices may vary across disciplines and over time, where incentives have increasingly favored journal publications, these differences are not expected to substantially affect the structure of co-authorship networks, which remain the focus of our analysis. While shorter time spans or rolling-window approaches could capture more fine-grained temporal dynamics, they may also lead to more fragmented and less stable network structures (Sedita et al., 2020). Given the time required for research production and publication processes, a ten-year window allows for the identification of more stable and consolidated collaboration patterns, rather than temporal or shorter communities. For this reason, we opted for a time-window of 10 years.

The analysis focuses on the period 2012–2022. This choice is primarily driven by data availability, as the database used (SciVal/Scopus) provides consistent coverage over a ten-year window. Moreover, this time span is sufficiently long to capture the recent evolution of collaboration patterns and the emergence of new research links, while maintaining a focus on current dynamics.

Articles are generally the main type of publication over time, although there are some differences between the two communities. In Management, the share of articles increases from about 70 to 80% toward the end of the period. In Statistics, articles account for around 80% of publications at the beginning and rise to approximately 90%, highlighting a stronger tendency to publish in journals.

Overall, despite covering a decade, the distribution of publication types remains relatively stable. There is, of course, an increase in the number of articles published, but over time Scopus has also started to include other types of publications, such as books and book chapters.

The co-authorship network is constructed over the entire 2012–2022 period. This choice reflects the nature of scientific collaboration, which typically develops over extended time spans and is only partially observable within shorter intervals. Aggregating the network over a sufficiently long period allows us to capture stable and recurring collaboration patterns, while reducing the influence of short-lived or episodic interactions.

Furthermore, the aim of this study is to identify external co-authors and characterize their thematic positioning within the collaboration structure, rather than to model the temporal evolution of the network itself. In this perspective, a longer observation window provides a more robust basis for detecting meaningful communities and for assigning greater weight to persistent co-authorship ties.

We downloaded records for 836 statisticians and 899 management scholars from the MUR as of December 31, 2022. In this way, the reference population has been fixed to a specific time. In total, we collected 12,485 documents for the Statistics community and 11,898 documents for the Management community, with respect to the last decade of scientific production. Given the similar number of scholars and publications in both groups, their comparability in terms of size and scientific productivity is confirmed.

With respect to the external co-authors, we found 31.166 authors for Statistics and 16.400 for Management, underlining a higher average number of authors per publication for Statistics. In these datasets there are 806 authors affiliated with SSD (so we have a classification for these authors) and 30.360 authors unaffiliated for Statistics, while for Management 808 authors affiliated with SSD (so we have a classification for these authors) and 15.601 authors remain unaffiliated.

Before filtering hyper-authored papers, the average number of co-authors per paper was 8.31 in Statistics and 5.85 in Management. After excluding publications with more than 20 authors, the averages decrease to 4.22 and 3.47, respectively.

For each document we have several information that could be useful to classify authors such as abstracts of the published and the affiliation history, but for this work we only used subject areas and keywords.⁵ Keywords were pre-processed using standard normalization procedures: all terms were converted to lowercase and minor variations due to punctuation or special characters were harmonized (e.g., “covid-19”, “covid19”, and “covid 19” were treated as equivalent). No further cleaning (e.g., acronym expansion) was applied, as it was deemed unlikely to affect the results. Subject areas were not pre-processed, as they are provided in a standardized and unambiguous form by the database. The combination of author-provided keywords and Scopus subject areas allows us to integrate two complementary sources of information that differ in both origin and interpretative scope. Author keywords are directly assigned by researchers and typically reflect the specific content, methods, or case studies of individual publications. In contrast, Scopus subject areas are attributed at the journal level and provide a more standardized and externally defined classification of the broader disciplinary context.

Because of these different sources, the two types of information are not fully overlapping. Rather, they tend to exhibit a loose hierarchical relationship: subject areas capture the general domain of research, while keywords offer a more fine-grained and context-specific description of the topics addressed within that domain. At the same time, some generic keywords (e.g., “Italy”, “statistics”, “economics”) appear across multiple subject areas, reflecting their transversal nature.

For Statistics, there are 27.001 keywords, but 23.891 keywords with frequency equal to less than 2 were eliminated. While for Management community, there are 24.572 keywords and 21.716 keywords with frequency equal to less than 2 were eliminated.

This choice was motivated by both computational and substantive considerations. From a practical standpoint, removing very low-frequency terms reduces sparsity and improves the feasibility and stability of the analysis. From a substantive perspective, keywords appearing only once over a ten-year period provide limited evidence of recurring or shared research themes, and are therefore less informative for identifying meaningful patterns at the community level.

While such rare keywords may occasionally capture highly specific or emerging topics, their isolated nature makes it difficult to distinguish signal from noise in a systematic way. For this reason, we focus on keywords that exhibit at least minimal recurrence, ensuring a more reliable representation of the thematic structure of the network.

⁵ Scopus subject areas are broad disciplinary classifications assigned at the journal level, allowing publications to be grouped into macro-fields (Elsevier, 2023). Since journals (and therefore publications) may be assigned to multiple subject areas, publications classified in more than one category are included in each relevant subject area, consistently with a full counting approach.

To assess the potential loss of information, we conducted a qualitative inspection of the excluded keywords. This examination did not reveal recurring patterns or systematically relevant themes for the purposes of this study, suggesting that their exclusion entails only a marginal loss of information.

The Statistics community includes 310 subject areas. Figures 1 and 2 present the 20 most frequent keywords and the 20 most frequent subject areas, respectively. The Management community includes 274 subject areas. Figures 3 and 4 present the corresponding 20 most frequent keywords and subject areas.

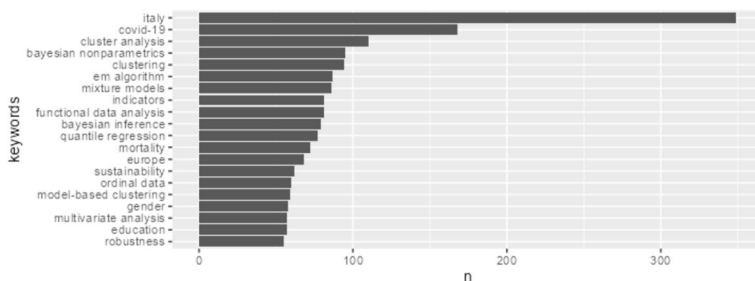


Fig. 1 The 20 most frequent keywords for Statistics Community

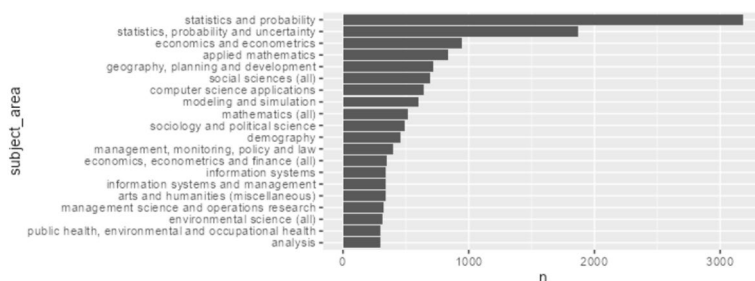


Fig. 2 The 20 most frequent subject areas for Statistics Community

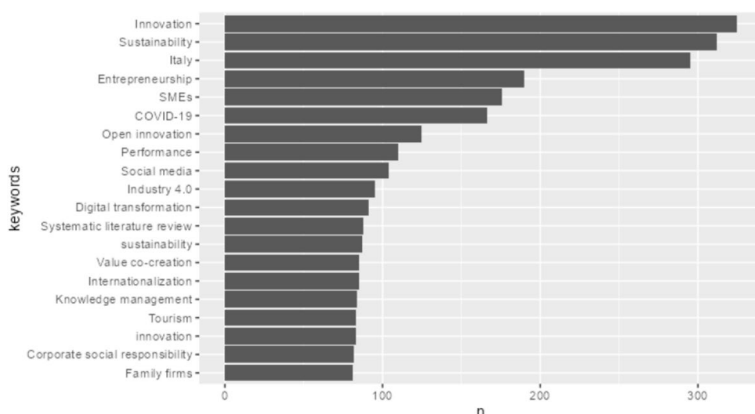


Fig. 3 The 20 most frequent Keywords for Management Community

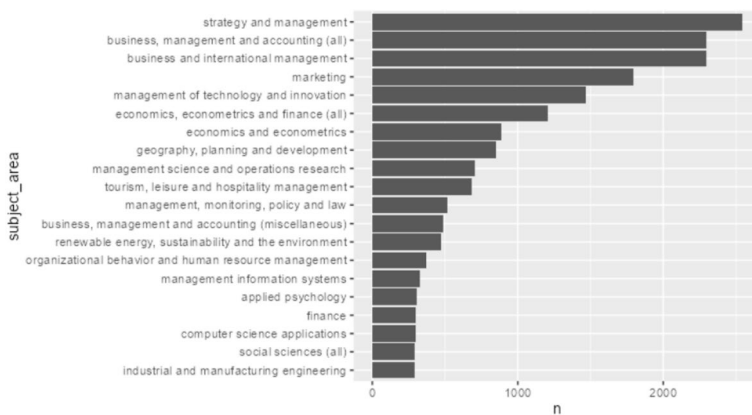


Fig. 4 The 20 most frequent subject areas for Management Community

Table 1 Co-authorship network main network indices for both scientific target groups after discarding publications with more than 20 authors

Descriptives	Statistics	Management
Vertices	18,835	9424
Edges	102,514	38,774
Edge density	0.00057	0.00087
Transitivity	0.3759	0.3033
N° of components	24	46
Largest component	18,576	8842
% of largest component	98.62%	93.82%
Geodesic distance	6.8028	6.9993
Diameter	23	28
Weight mean	1.4	1.6

Even when considering the evolution of most frequent keywords over-time, results suggest a certain stability in both academic fields, only few emergent topics arise such as Covid in 2019, and Artificial Intelligence and Machine Learning for Statistics in the last three years. Keywords and Subject areas are of course related, but they do not overlap as they come from two different sources. Subject areas are broad disciplinary categories assigned to journals by Scopus, providing a standardized way to classify research fields according to main journal topics. In contrast, author keywords are specific terms chosen by authors to describe the content of individual articles, thus are for sure more detailed when it comes to themes but less useful to identify broad topics. While subject areas remain fixed for a given journal whatever the paper is, author keywords vary across papers and are not related to the journals, so they vary to great extent even within the same journal, allowing for a very specific characterization of articles. That is why it could be beneficial to introduce both aspects into the analyses.

Co-authorships networks structure

The co-authorship networks were constructed using a full counting approach.⁶ Both networks have been constructed by filtering out works with more than 20 authors to ensure meaningful co-authorship relationships and deleting works usually related to disciplines (such as Medicine and Physics) where the number of authors is usually very high. The co-authorship networks in Statistics and Management exhibit distinct structural characteristics, reflecting different collaboration patterns within these disciplines (Table 1).

The Statistics network is notably larger even after such cut, with more vertices (18,835 vs. 9,424) and a higher number of edges (102,514 vs. 38,774), indicating a stronger tendency toward broader collaborations. This aligns with the observed lower edge density in the Statistics network (0.00057 vs. 0.00087 in Management), suggesting that statisticians are less interconnected in their research production. Additionally, the mean weight of collaboration (1.6 in Management vs. 1.4 in Statistics) implies that repeated co-authorships are slightly more frequent in Management, potentially pointing to more stable and recurrent research partnerships, while statisticians tend to open new collaborations more often.

Overall, both fields exhibit a relatively high and increasing level of collaboration with external authors, although with different patterns. In Statistics, the share of external co-authors is consistently high throughout the period, starting above 90% in 2012, followed by a decline until 2014, and then a gradual recovery. From 2018 onwards, the trend stabilizes at high levels (around 80–85%), suggesting a strong and persistent openness to external collaboration. In Management, the proportion of external co-authors is lower but shows a clearer upward trend over time. After a relatively stable phase between 2012 and 2014 (around 53–55%), the share increases, reaching over 70% by the end of the period. Despite some fluctuations, the overall pattern indicates a progressive strengthening of external collaboration.

Comparatively, Statistics maintains a higher level of external collaboration across all years, while Management exhibits a more pronounced growth trajectory. These patterns suggest structural differences in collaboration practices between the two domains, with Statistics characterized by a more established external network and *Management* showing a process of gradual expansion.

The Management network is more fragmented, with 46 connected components compared to 24 in Statistics, leading to a lower proportion of researchers in the largest component (93.82% vs. 98.62%). This suggests that Management researchers tend to form smaller, more isolated groups rather than integrate into a single, highly connected scientific community.

In terms of network cohesiveness, both fields exhibit similar average geodesic distances (6.80 in Statistics vs. 6.99 in Management). Still, the Management network has a slightly larger diameter (28 vs. 23), indicating that the longest shortest path between any two researchers is greater, potentially due to the presence of more peripheral research groups. Additionally, transitivity (0.3759 in Statistics vs. 0.3033 in Management) is higher in Statistics, suggesting a greater tendency for triadic closure, where researchers are more likely to collaborate with colleagues of their existing co-authors.

These patterns can also be interpreted in light of small-world network properties (Watts & Strogatz, 1998), which are typically characterized by high clustering and relatively

⁶ Specifically, for each publication, a link of weight 1 was assigned between every pair of co-authors. The links were not fractionally weighted as usually done in co-authorship analysis (De Stefano et al., 2016; 2023).

short path lengths. In this perspective, the higher transitivity observed in the Statistics network, combined with comparable geodesic distances, suggests a structure that is closer to a small-world configuration, whereas the Management network appears relatively more fragmented.

Overall, these findings highlight that Statistics fosters a more integrated and densely connected research environment. In contrast, Management exhibits a more fragmented structure with a higher occurrence of recurrent collaborations among the same researchers. This divergence in network topology could be influenced by specific sector norms, research funding structures, or differences in publishing practices between the two fields. Furthermore, statisticians tend to be more connected towards multidisciplinary collaborations.

Method: two-mode network using textual data and clustering algorithm

In this study, we focus on a bipartite (or two-mode) network (authors–keywords–subject areas). This implies that the nodes, which are the fundamental components of the network, belong to two different categories and connections between nodes of the same category are not allowed. Therefore, links only exist by connecting nodes of one category to nodes of the other category. In this case, scholars on one hand and keywords–subject areas on the other hand, are the two distinct sets of nodes. Keywords are usually specified by authors, while subject areas are usually linked to the journal aims and scope. Therefore, such two information, treated as textual data, come from different sources and give complementary information (see Figs. 1, 2, 3, and 4).

Links among nodes are defined as not directed, with symmetric relationships. Further, such links are conceived as weighted, with weights equal to the number of times the given scholar uses in her/his works the words of the other set of nodes. In this framework, we are interested in the pattern that can be found analysing simultaneously both set of nodes. The aim is to classify scholars and understand and interpret the classification taking advantage of the keywords and subject areas that are encompassed in the same group. For this purpose, the core of the analysis will be in the Community Detection phase for bipartite network.

Community Detection (see e.g. Newman, 2013) seeks to identify clusters of nodes that are tightly connected within communities and sparsely connected with nodes (both scholars and keywords–subject areas) from other communities. In other words, nodes have dense connections within their own community and fewer connections with members of other communities. Common community detection algorithms in one-mode networks are based on the concept of modularity (Newman, 2006), which measures the quality of community divisions in one-mode networks. Instead, we employ a two-mode community detection method (or equivalently, a community detection for a bipartite graph). This method is implemented in the DIRTLPawb+ algorithm (Beckett, 2016), originally proposed by Barber (2007), which identifies community partitions in a given two-mode network by maximizing weighted modularity in bipartite networks composed of two types of nodes.

The modularity is conceived in a bipartite graph as follows:

$$Q_b = \frac{1}{m} \sum_{u=1}^r \sum_{v=1}^c \left(\tilde{\mathbf{A}}_{uv} - \mathbf{P}_{uv} \right) \delta(g_u - h_v) = \frac{1}{m} \sum_{u=1}^r \sum_{v=1}^c \left(\tilde{\mathbf{A}}_{uv} - \frac{k_u d_v}{m} \right) \delta(g_u - h_v)$$

In this modularity representation, $\mathbf{P}$ is the null model matrix representing the probability of links between nodes, given the degree distributions of $\tilde{\mathbf{A}}$, the block diagonal affiliation

matrix with r nodes of the first type (scholars) and c nodes of the second type (keywords-subject areas); k is the node degree for 1st type nodes while d is the node degree for 2nd type nodes.

1st type node labels are denoted by g , while h are the labels for 2nd type nodes, and the Kronecker delta function $\delta(g_u, h_v)$ is equal to one when nodes u and v are classified as being in the same community or zero otherwise. In this method, The DIRTLPawb+ algorithm maximizes the bipartite modularity proposed by Barber (2007). This formulation does not include an explicit resolution parameter (γ), and corresponds to the standard modularity definition where $\gamma=1$. Therefore, the detected community structure is not driven by an externally imposed resolution scale but solely by the network topology and the modularity optimization procedure. In this context, we denote as an affiliation matrix representation of the graph: it has as many rows as the number of nodes of the first type and as many columns as nodes of the second type. An example of such matrix formalization of the bipartite graph is given in the Fig. 5. We included in the analysis also scholars for whom we know they belong to the first target group (Statisticians) or second target group (Management scholars). The overall classification is made including the internal nodes; thus, we will use such information to figure out the interdisciplinarity edges of the two populations, and to use them as a sort of internal validation of the classification procedure. We expect that many scholars working in the same scientific discipline and on the same topics, or even belonging to the same institution, will likely be merged together in the community detection phase. If this will be the case, it will be a hint that also the external co-authors have been classified in a consistent way. For what concerns the weights, the distribution of edge weights is highly skewed in both datasets, with approximately 80% of edges in the Statistics network and 76% in the Management network having weight equal to one. These ties represent sporadic collaborations between authors who co-authored only one paper over the ten-year observation period. To focus on more stable collaboration relationships, we

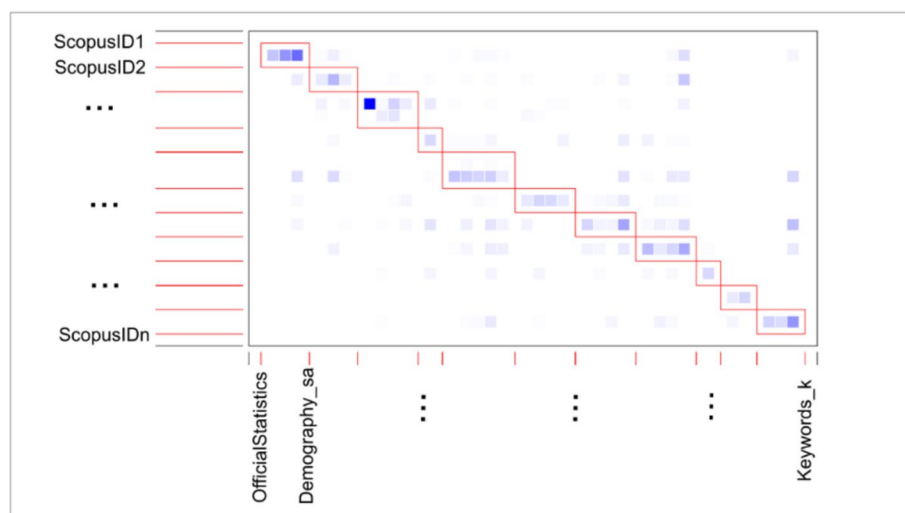


Fig. 5 Rows are authors identified by their Scopus ID, columns are keywords and subject areas, denoted as “sa” at the end. This incidence (affiliation) matrix is just an example. Modules merge together both rows and columns simultaneously. Cell color intensity corresponds to the magnitude of the entries in the affiliation matrix

therefore exclude edges with weight equal to one and retain only repeated collaborations ($\text{weight} \geq 2$).

Results

Comparison between the two sectors

In this section, we compare the two fields under study. The Statistics community is structured into 83 clusters, while the Management community comprises 47 clusters. A “cluster” represents a scientific thematic area to which researchers are assigned, based on specific keywords and subject areas. The table below summarizes the main characteristics of these clusters for both fields.

The analysis of co-authorship networks in the communities of Statisticians and Management scholars highlights significant differences in their collaborative structures (Table 2). The statistics community is characterized by a higher number of clusters (83 compared to 47) and slightly greater modularity (0.54 vs. 0.51), suggesting a more segmented network with distinct groups. This is also reflected in the average cluster size, which is smaller in statistics in terms of both nodes (71.95 vs. 98.83) and authors (46.93 vs. 63.55), indicating greater fragmentation compared to management, where groups appear larger and more cohesive. Differences also emerge in the distribution of authors within clusters: while the average percentage of authors per cluster is similar (55.3% for statistics and 57.5% for management), variability is greater in the former, with clusters ranging from a minimum of 23.8% to a maximum of 98.7% of authors, whereas in management, distribution is more

Table 2 Comparisons of the two communities' structure

	Statisticians communities	Management communities
N of clusters	83	47
Modularity	0.54	0.51
Avg N of nodes in cluster	71.95	98.83
Avg N of authors in cluster	46.93	63.55
Avg % of authors in cluster	55.3%	57.5%
Max % of authors in cluster	98.7%	96.9%
Min % of authors in cluster	23.8%	33.3%
Avg N of words in cluster	25.01	35.27
Avg % of words in cluster	44.7%	42.5%
Max % of words in cluster	76.1%	66.7%
Min % of words in cluster	1.2%	3.1%
Mean % internal authors in cluster	5.8%	9.5%
MAX internal %	18%	37.5%
MIN internal %	0%	0%
Less than 20	45	23
Between 21 and 100	27	16
100 +	11	8
Range	3–1420	3–1252

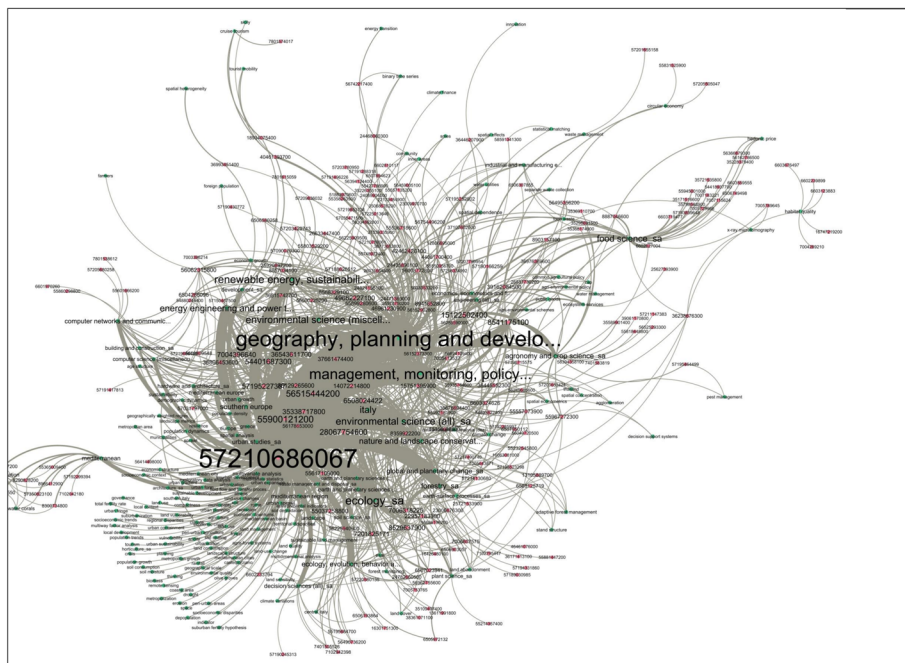


Fig. 6 Graph of clusters 13 in the Statistics target group. Source: our elaboration. Legend: Green nodes: subject areas/keywords, red nodes: authors

balanced (33.3–96.9%). Additionally, the management community exhibits greater thematic diversity, with a higher average number of words per cluster (35.27 vs. 25.01) and a broader distribution, suggesting a stronger interdisciplinary orientation. Finally, the percentage of internal authors per cluster is higher in management (9.5% vs. 5.8%), with peaks reaching 37.5%, while in statistics, the maximum is 18%, confirming a denser collaborative structure in the former.

These findings suggest that statistics, as a more technical and methodological discipline, tends to develop through more specialized and fragmented research groups, whereas management, with its more applied and multidisciplinary nature, fosters larger and more integrated collaborative networks.

Two clusters identified in the two target groups

The clusters identified by the proposed method appear to be consistent, regardless of their size. Specifically, in this section we present two examples to illustrate the similarities and differences between the two communities, primarily to qualitatively demonstrate the method’s effectiveness in categorizing researchers into macro areas.

In general, there are clusters characterized by many scholars and many subject areas and keywords, while others are defined by a limited number of scholars and few subject areas and keywords. Furthermore, each cluster exhibits a strong connection between keywords and subject areas, offering complementary insights. Here, we present the main clusters

for both sectors, noting that the unillustrated clusters exhibit patterns very similar to those described below.

Cluster 7 in the statistics target group

An especially noteworthy cluster within the Statistics target group consists of 479 scholars and 320 subject areas and keywords (Fig. 6 and Appendix 1). A compact version of the cluster is shown in Fig. 6, where nodes with a degree of less than 2 are excluded. The complete graph, which includes all nodes and relationships, is provided in Appendix 1. This is a two-mode graph, in which green nodes represent keywords/subject areas, while red nodes represent the authors of various publications. Node sizes are proportional to their degree.

This community is characterized by topics such as environment, sustainability, ecology, geography, energy, and environmental science. A manual review confirms the internal consistency of the cluster, as the researchers included are actively engaged in these areas. Notably, the group brings together experts from diverse disciplinary backgrounds, reflecting the interdisciplinary nature of environmental and energy-related research.

In particular, the external co-authors involved in this cluster can be classified into different, yet related, thematic areas. The most prominent area that emerges is geography, planning, and development, which is naturally associated with keywords and publications related to specific geographic regions. Indeed, the main keywords in this area include Italy, Mediterranean Europe, and Southern Europe, as well as the Spatial Analysis methodology.

Additionally, the cluster also highlights a thematic area related to management, monitoring, and policy, which similarly incorporates a geographic dimension (with Italy again emerging as a key keyword). Furthermore, it is worth noting that Environmental Science and Ecology represent common thematic areas among the authors and co-authors in this cluster, with keywords such as climate change and nature and landscape conservation, as well as methodological keywords like spatial econometrics. Another important topic which emerges in this group is related to the *renewable energy*, *sustainability*, and *energy engineering and power technology*. Among other minor topics, the cluster also includes Food Science, Agronomy, and Crop Science, particularly linked to the keyword food waste.

Figure 6 reveals the presence of a prominent author who appears in numerous publications within this cluster.

From a classificatory perspective, this community includes scholars belonging to the Economic statistics subfield and statisticians specialized in spatial models linked to interdisciplinary scholars such as ecologists, environmental scientists, energy engineers, agronomists, and others related to the field of sustainability. Interestingly, this cluster can also reveal a group of statisticians that work together with scholars belonging to the management community under study. In fact, in the following section we will describe a cluster for the Management target group that partially contains similar subject areas and keywords with cluster 13 above described for statisticians.

Cluster 13 in the management target group

One of the largest clusters in management is related to sustainability, the environment, and the circular economy (Fig. 7–Appendix 2). A compact version of the cluster is shown in Fig. 7, where nodes with a degree of less than 2 are excluded. The complete graph, which includes all nodes and relationships, is provided in Appendix 2. As said, green nodes

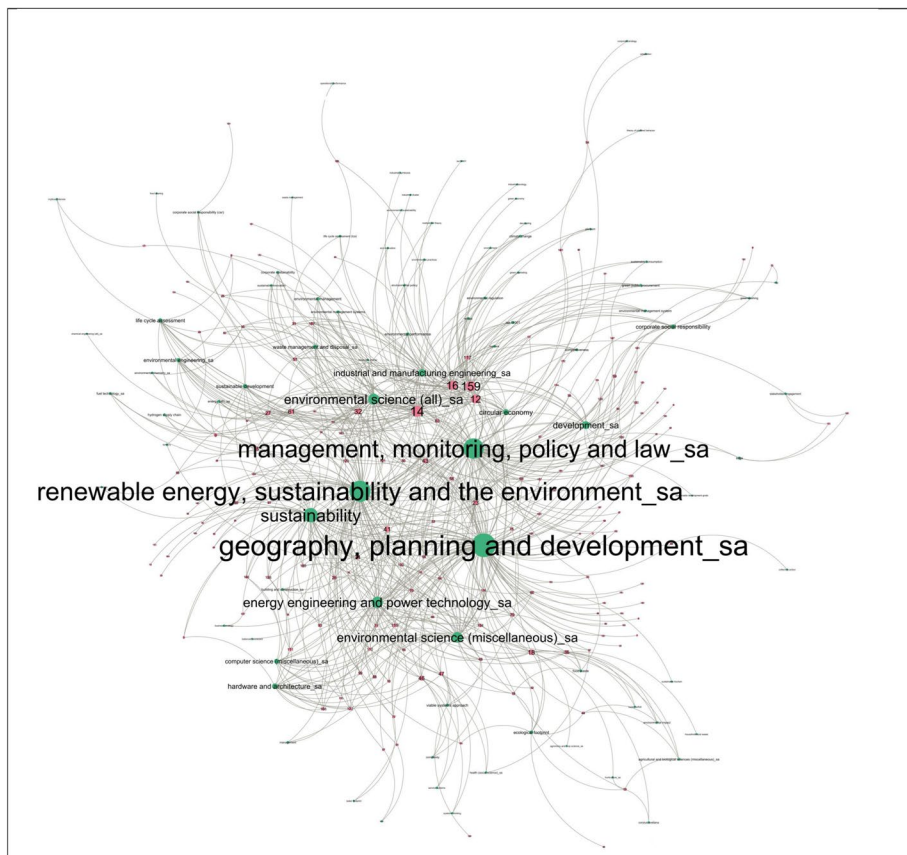


Fig. 7 Graph of Cluster 13 in the Management target group. Source: our elaboration. Legend: Green nodes: subject areas/keywords, red nodes: authors

represent keywords/subject areas, while red nodes represent the authors of various publications. Node sizes are proportional to their degree.

This cluster consists of 193 keywords and subject categories, along with 372 scholars, highlighting the significance and high engagement of the topics at hand. The group of keywords concerning environmental sustainability and the circular economy focuses on the importance of practices and strategies that reduce the environmental impact of human activities and promote responsible resource use. Within this cluster, we can highlight sub-themes such as the Circular Economy, with keywords like *circular economy*, *recycling*, *waste management*, *waste prevention*, etc.

A second area concerns Environmental Management, with keywords such as: *environmental management*, *environmental management systems*, *environmental regulation*, *ISO 14001*, *EMAS*. These terms pertain to the implementation of systems and regulations to manage and improve the environmental performance of organizations. The adoption of standards such as ISO 14001 and EMAS helps companies monitor and enhance their environmental sustainability.

Finally, there is an area of Sustainability and Innovation, with keywords such as *sustainable development*, *sustainable business model*, *eco-innovation*, *sustainable innovation*, *green practices*. These terms refer to the adoption of innovative practices that promote sustainability. Eco-innovation involves developing products and processes that reduce environmental impact, while sustainable business models aim to integrate sustainability as a fundamental part of the business strategy.

Equally important is the evaluation of Environmental Impact: *environment*, *environmental impacts*, *carbon footprint*, *water footprint*, *ecological footprint*. These terms concern the measurement and assessment of the impact of human activities on the environment. The carbon footprint measures greenhouse gas emissions, while water footprint and ecological footprint evaluate the use of water and biological resources.

This cluster is highly coherent, despite its size, featuring internationally recognized scholars specializing in the environment and sustainability, and it highlights the growing importance of these topics in recent years.

Conclusion

This study highlights the challenges of using co-authorship as an indicator of collaborative behavior, particularly the tendency of prior research to focus on specific disciplines or national networks while overlooking external authors. A key limitation of such studies is the lack of systematic methodologies for classifying all authors within co-authorship networks, especially in terms of their disciplinary or macro-disciplinary affiliations. To address this, we proposed an approach that leverages textual information—such as keywords and subject areas extracted from bibliographic records—to classify external co-authors of a target group of scholars. By constructing a two-mode network of authors and keywords/subject areas and applying community detection methods, we identified research communities that meaningfully reflect shared scientific interests. Our findings demonstrate that integrating co-authorship networks with textual analysis yields promising results for mapping scholarly collaboration.

Results provide evidence of the qualitative coherence of the identified communities. In particular, the detailed examination of selected clusters shows that they are internally consistent in terms of thematic focus, with smaller groups often including scholars from the same department who frequently collaborate. While only a subset of communities is analyzed in depth, these examples suggest that the method is able to capture meaningful co-authorship and thematic groupings. Moreover, the detected communities do not display evident internal substructures, which is consistent with the identification of relatively well-defined research clusters. To further assess the plausibility of the detected communities, we complemented the structural analysis with both qualitative and quantitative evidence. From a qualitative standpoint, we manually inspected numerous communities, checking the alignment between authors and the identified topics. Several communities display a clear internal coherence in terms of research groups and themes (for example, first small group identified in the Statistics field encompasses subject areas as Biology, Chemistry and Molecular Biology), keywords, and institutional affiliations. In multiple cases, clusters correspond to identifiable groups of scholars working on closely related topics or within the same disciplinary or organizational contexts, supporting the interpretability of the partition. In addition, we conducted a quantitative assessment of semantic coherence based on subject areas associated with authors. Using SciBERT-based embeddings, we computed

cosine similarity both within and across communities. Results indicate that the average within-community similarity is higher than the between-community similarity (approximately 0.3 vs. 0.2, respectively), consistently across both the Statistics and Management domains. While the difference is not large, it suggests that the detected communities tend to group together semantically related topics and researchers.

The absence of clearly identifiable substructures within the detected communities can be interpreted in light of the structural properties of co-authorship networks. In our data, many communities exhibit a configuration that is consistent with a core–periphery structure, where one or a few highly connected scholars occupy a central position and act as bridges among a set of less connected collaborators. These central actors frequently belong to the target disciplinary groups (e.g., *Statistics* or *Management*), and their co-authorship ties connect otherwise weakly interconnected authors.

Under this configuration, the internal organization of communities is not characterized by the presence of multiple dense and well-separated subgroups, but rather by a leader-based pattern in which peripheral nodes are primarily linked through the central core (as also found in previous analysis on the same target group, see Geremia et al., 2025). As a result, potential subdivisions tend to lack the internal density and separation that would justify their identification as distinct subcommunities. This interpretation is supported by graphical inspection of the networks, as well as by network-level indicators that do not suggest the presence of secondary modular structures within the identified clusters.

Overall, these elements provide converging evidence, qualitative, semantic, and structural, on the coherence of the detected communities, while acknowledging that the assessment is based on a subset of possible validation approaches.

Interestingly, the method effectively classifies both Italian and foreign co-authors in a meaningful way, indicating that no systematic bias related to national or international affiliations emerges in the categorization of external co-authors.

Another notable observation is the distinct distribution of subject areas and keywords across the network. While subject areas tend to be associated with broader thematic classifications, keywords provide more granular insights into specific research topics. The fact that these two types of textual information occupy regions of the network that are usually somehow close but do not overlap, suggests that their combined use enhances the classification process by providing complementary perspectives. This reinforces the idea that integrating multiple textual features improves the ability to map scientific collaborations and delineate research communities with greater accuracy.

However, several limitations remain. First, our approach currently relies on a restricted set of textual features (i.e. keywords and subject areas), while additional sources of information—such as abstracts or the affiliation history—could improve classification accuracy. Second, the way in which the network is constructed and segmented requires further refinement, as different methodological choices in defining links and cutting the network may influence the detected communities. Third, while we employ two-mode community detection, alternative methods could be explored to improve the robustness of the classification.

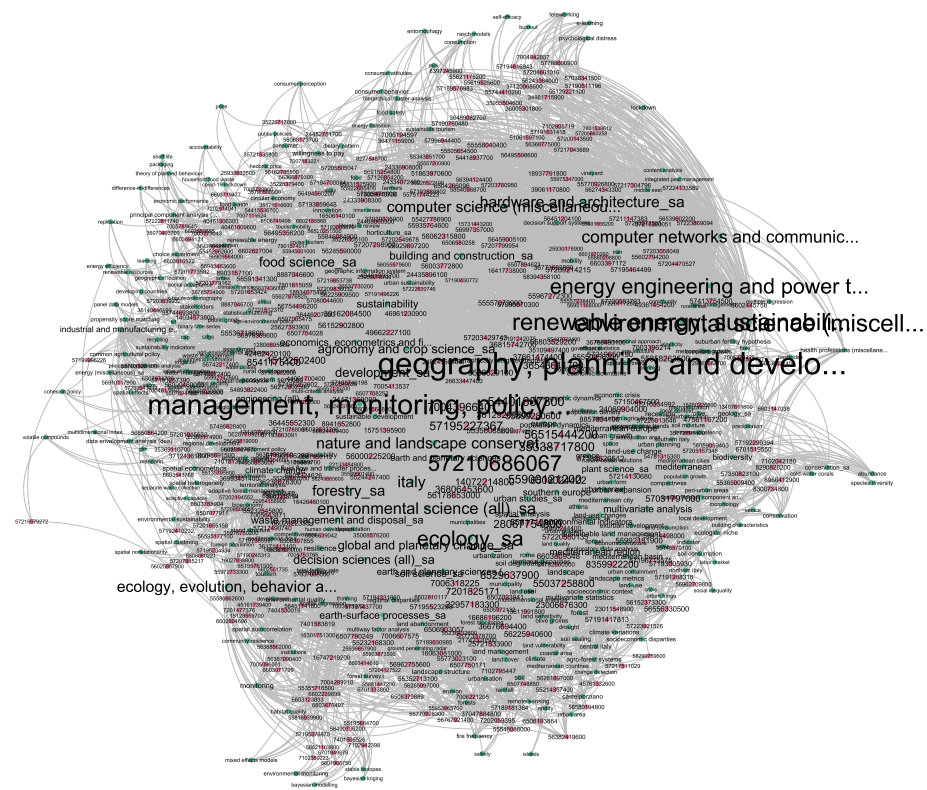
Moreover, co-authorship represents only a proxy for scientific collaboration and does not capture other forms of interaction, such as informal exchanges or contributions acknowledged outside formal authorship. Alternative approaches, for instance based on acknowledgements, have been proposed to map broader collaboration structures (Petrovich, 2022). Other studies have also adopted contributorship-based approaches, which better capture the heterogeneity of authors' roles and contributions than traditional co-authorship data (e.g., Larivière et al., 2021). At present, however, information on authors' roles and contributions is available only for a limited set of journals and remains difficult to use for comprehensive

Future research could also further investigate the presence of small-world properties (Watts & Strogatz, 1998) in co-authorship networks, in order to assess whether differences in clustering and path lengths contribute to the observed structural heterogeneity across fields.

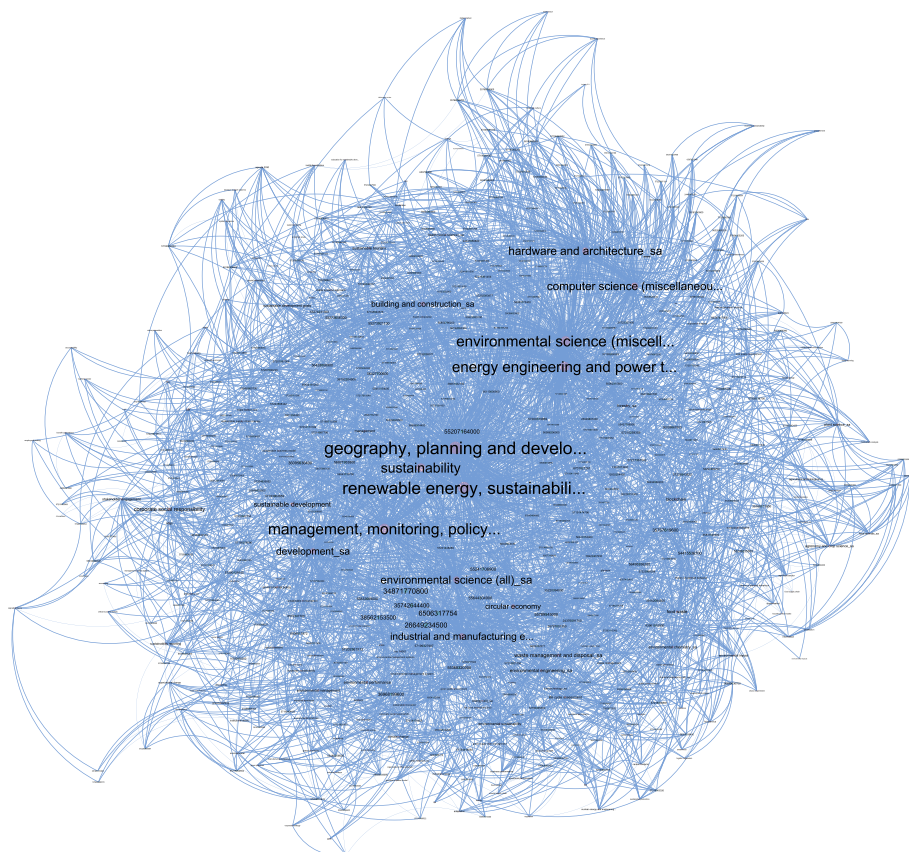
Finally, external validation remains an open challenge: incorporating auxiliary information could help assess the reliability of the clustering results and provide further insights into the effectiveness of our approach. For the case of Statisticians, the usage of the sub-sectors could be a potential auxiliary variable to introduce for validation purposes, but it is not available for the Management group.

Future work will focus on addressing these limitations. In particular, we will quantitatively evaluate the effectiveness of the proposed method, compare it with alternative classification techniques, and explore additional sources of textual information to refine the identification of research external co-authors within co-authorship networks starting from a specific academic community.

Appendix 1



Appendix 2



Acknowledgements We acknowledge financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 104 published on 2.2.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union-NextGenerationEU. Project Title “Models and methods for the analysis of collaboration networks”, CUP J53D23011540006, Principal Investigator (PI) prof. Domenico De Stefano.

Funding Open access funding provided by Università degli Studi di Trieste within the CRUI-CARE Agreement. Funding was provided by Ministero dell’Università e della Ricerca, 2022MSL3AY, Domenico De Stefano.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Acedo, F. J., Barroso, C., Casanueva, C., & Galán, J. L. (2006). Co-authorship in management and organizational studies: An empirical and network analysis. *Journal of Management Studies*, 43(5), 957–983.
- Barabási, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica a, Statistical Mechanics and Its Applications*, 311(3–4), 590–614.
- Barber, M. J. (2007). Modularity and community detection in bipartite networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 76(6), 066102.
- Beattie, V., & Goodacre, A. (2004). Publishing patterns within the UK accounting and finance academic community. *The British Accounting Review*, 36(1), 7–44.
- Beaver, D. (2001). Reflections on scientific collaboration (and its study): Past, present, and future. *Scientometrics*, 52(3), 365–377.
- Beckett, S. J. (2016). Improved community detection in weighted bipartite networks. *Royal Society Open Science*, 3(1), Article 140536.
- Cainelli, G., de Felice, A., Lamonarca, M., & Zoboli, R. (2006). The publications of Italian economists in ECONLIT. Quantitative assessment and implications for research evaluation. *Economia Politica*, 23(3), 385–424.
- Cainelli, G., Maggioni, M. A., Uberti, T. E., & De Felice, A. (2012). Co-authorship and productivity among Italian economists. *Applied Economics Letters*, 19(16), 1609–1613.
- Cainelli, G., Maggioni, M. A., Uberti, T. E., & De Felice, A. (2015). The strength of strong ties: How co-authorship affect productivity of academic economists? *Scientometrics*, 102(1), 673–699.
- Capone, F., & Lazzeretti, L. (2023). United we stand, divided we fall. A co-authorship analysis of management scholars in Italy. *Sinergie*, 41(2), 57–76.
- Casaluce, R., & D'Agata, R. (2025). Impact of Formal and Informal Collaborations on Academic Performance in Italy. In T. Di Battista, Y. Calleo, A. Cucco *DSSR 2025 - Towards a holistic understanding of society: bridging Social Sciences, Statistics and Computational Sciences* (Pescara 19 - 21 February, 2025). ISBN: 9781326620653.
- Casanueva, C., & Larrinaga, C. (2013). The (uncertain) invisible college of Spanish accounting scholars. *Critical Perspectives on Accounting*, 24(1), 19–31.
- Cox, D. R. (1997). The current position of statistics: A personal view. *International Statistical Review*, 65(3), 261–276.
- Crane, D. (1969). Social structure in a group of scientists: A test of the “invisible college” hypothesis. *American Sociological Review*. <https://doi.org/10.2307/2092499>
- Cugmas, M., Mali, F., & Kronegger, L. (2024). Longitudinal patterns of scientific collaboration in doctoral studies. *Scientometrics*, 129(2), 1055–1077.
- De Stefano, D., & Zaccarin, S. (2016). Co-authorship networks and scientific performance: An empirical analysis using the generalized extreme value distribution. *Journal of Applied Statistics*, 43(1), 262–279.
- De Stefano, D., Fuccella, V., Vitale, M. P., & Zaccarin, S. (2013). The use of different data sources in the analysis of co-authorship networks and scientific performance. *Social Networks*, 35(3), 370–381.
- De Stefano, D., Fuccella, V., Vitale, M. P., & Zaccarin, S. (2016). Improving co-authorship network structures by combining multiple data sources: Evidence from Italian academic statisticians. *Scientometrics*, 107(1), 167–184.
- De Stefano, D., Fuccella, V., Vitale, M. P., & Zaccarin, S. (2023). Quality issues in co-authorship data of a national scientific community. *Network Science*, 11(1), 98–112.
- De Stefano, D., Giordano, G., & Vitale, M. P. (2011). Issues in the analysis of co-authorship networks. *Quality & Quantity*, 45(5), 1091–1107.
- De Stefano, D., Vitale, M. P., & Zaccarin, S. (2017). Community Structure in Co-authorship Networks: The Case of Italian Statisticians. *Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society* (pp. 65–72). Springer.
- Defazio, D., Lockett, A., & Wright, M. (2009). Funding incentives, collaborative dynamics and scientific productivity: Evidence from the EU framework program. *Research Policy*, 38(2), 293–305.
- Elsevier (2023). Scopus Content Coverage Guide. Available at: https://service.elsevier.com/app/answers/detail/a_id/15181/supporthub/scopus/.
- Fabbrucci Barbagli, A. G., De Stefano, D., Santelli, F., & Zaccarin, S. (2025a). Unveiling collaboration persistence and interactions among Italian academic statisticians through relational hyperevent models. *Statistical Methods & Applications*, 34(5), 1149–1167.

- Fabbrucci Barbagli, A.G., Lerner, J., & Boudourides, M. (2025b). Co-authors and Co-principal Investigators: Relational Hyperevent Models for the Co-evolution of Scientific Publications and Grant Funding. In *Scientific Meeting of the Italian Statistical Society*. Cham: Springer Nature Switzerland, 289–295.
- Fuccella, V., De Stefano, D., Vitale, M. P., & Zaccarin, S. (2016). Improving co-authorship network structures by combining multiple data sources: Evidence from Italian academic statisticians. *Scientometrics*, 107(1), 167–184.
- Geremia, S., De Stefano, D., & Fop, M. (2025). Community-level core-periphery structures in collaboration networks. *arXiv preprint arXiv:2511.19305*.
- Glänzel, W. (2002). Co-authorship patterns and trends in the sciences (1980–1998): A bibliometric study with implications for database indexing and search strategies. *Library Trends*, 50(3), 461–473.
- Glänzel, W., & Schubert, A. (2004). Analysing scientific networks through co-authorship. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 257–276). Springer.
- Henriksen, D. (2016). The rise in co-authorship in the social sciences (1980–2013). *Scientometrics*, 107(2), 455–476.
- Hu, C., & Racherla, P. (2008). Visual representation of knowledge networks: A social network analysis of hospitality research domain. *International Journal of Hospitality Management*, 27(2), 302–312.
- Kumar, S. (2015). Co-authorship networks: A review of the literature. *Aslib Journal of Information Management*, 67(1), 55–73.
- Larivière, V., Pontille, D., & Sugimoto, C. R. (2021). Investigating the division of scientific labor using the contributor roles taxonomy (CRediT). *Quantitative Science Studies*, 2(1), 111–128.
- Lazzeretti, L. (2001). Metodologie statistiche e management research: Una content analysis comparativa su alcune riviste manageriali. *Sinergie*, 55, 235–276.
- Lazzeretti, L., Sedita, S., & Caloffi, A. (2014). Founders and disseminators of cluster research. *Journal of Economic Geography*, 14(1), 21–43.
- Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, 35(5), 673–702.
- Menardi, G., & De Stefano, D. (2022). Density-based clustering of social networks. *Journal of the Royal Statistical Society Series a: Statistics in Society*, 185(3), 1004–1029.
- Merigó, J. M., Rocafort, A., & Aznar-Alarcón, J. P. (2016). Bibliometric overview of business & economics research. *Journal of Business Economics and Management*, 17(3), 397–413.
- Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, 69(2), 213–238.
- Newman, M. E. J. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64(1), 016132.
- Newman, M. E. J. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E*, 64(1), Article 016131.
- Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), 8577–8582.
- Newman, M. E. J. (2013). Community detection and graph partitioning. *Europhysics Letters*, 103(2), Article 28003.
- Newman, M.E.J. (2004). Who is the best connected scientist? A study of scientific co-authorship networks. *Complex Networks*, 337–370.
- Olmos-Peñuela, J., Molas-Gallart, J., & Castro-Martínez, E. (2014). Informal collaborations between social sciences and humanities researchers and non-academic partners. *Science and Public Policy*, 41(4), 493–506.
- Pan, R. K., & Saramäki, J. (2012). The strength of strong ties in scientific collaboration networks. *Europhysics Letters*, 97(1), Article 18007.
- Petrovich, E. (2022). Acknowledgments-based networks for mapping the social structure of research fields. A case study on recent analytic philosophy. *Synthese*, 200(3), Article 204.
- Plümper, T., & Radaelli, C. (2004). Publish or perish? Publications and citations of Italian political scientists in international political science journals, 1990–2002. *Journal of European Public Policy*, 11(6), 1112–1127.
- Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P., & Bachrach, D. G. (2008). Scholarly influence in the field of management: A bibliometric analysis of the determinants of university and author impact in the management literature in the past quarter century. *Journal of Management*, 34(4), 641–720.
- Ponomarev, B., & Boardman, C. (2016). What is co-authorship? *Scientometrics*, 109, 1939–1963.
- Racherla, P., & Hu, C. (2010). A social network perspective of tourism research collaborations. *Annals of Tourism Research*, 37(4), 1012–1034.

- Savić, M., Ivanović, M., Jain, L. C., Savić, M., Ivanović, M., & Jain, L. C. (2019). Extraction of Co-authorship Networks. Complex Networks in *Software, Knowledge, and Social Systems*, 193–234.
- Sedita, S. R., Caloffi, A., & Lazzeretti, L. (2020). The invisible college of cluster research: A bibliometric core–periphery analysis of the literature. *Industry and Innovation*, 27(5), 562–584.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684), 440–442.
- Yan, E., & Guns, R. (2014). Predicting and recommending collaborations: An author-, institution-, and country-level analysis. *Journal of Informetrics*, 8(2), 295–309.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.