Measuring Interdisciplinarity: A Graph-Based Analysis of Brazilian Academic Committees

Victor A. P. Mansueli; Jesús P. Mena-Chalco

Cómo citar este artículo:

Mansueli, Victor A. P.; **Mena-Chalco, Jesús P.** (2025). "Measuring Interdisciplinarity: A Graph-Based Analysis of Brazilian Academic Committees". *Infonomy*, 3(3) e25018.
https://doi.org/10.3145/infonomy.25.018

Victor A. P. Mansueli
https://orcid.org/0000-0002-8803-8876
https://directorioexit.info/ficha7247
Universidade Federal do ABC (UFABC)
Centro de Matemática, Computação e Cognição (CMCC)
Grupo de Pesquisa em Cientometria
Avenida dos Estados, 5001
Bairro Bangu 09280-560, Santo André, Brasil
victor.mansueli@ufabc.edu.br

Jesús P. Mena-Chalco
https://orcid.org/0000-0001-7509-5532
https://directorioexit.info/ficha7248
Universidade Federal do ABC (UFABC)
Centro de Matemática, Computação e Cognição (CMCC)
Grupo de Pesquisa em Cientometria
Avenida dos Estados, 5001
Bairro Bangu 09280-560, Santo André, Brasil
jesus.mena@ufabc.edu.br

Abstract

This study presents a scientometric analysis of interdisciplinary scientific collaboration in the composition of academic committees in Brazil. The research employed an extensive dataset, obtained from *CAPES* and *Sucupira* platforms, which store information on Brazilian academic output: master's and doctoral dissertations and theses. The collected data enabled the examination of relationships between 66 Knowledge Areas (KAs). Applying graphs to model the relationships among committee members, we distinguished each member according to their assigned role: candidate, evaluator, advisor. Three "academic graphs" were designed to represent these relationships: evaluations, invitations, and co-participations. As a method, we adapted a Knowledge Discovery in Database (KDD) process, systematizing the necessary steps to enable the desired measurement. Applying statistical techniques, the research results revealed significant patterns of interdisciplinarity in the composition of the committees. It was possible to identify KAs that collaborate more frequently (Ecology, Education, Linguistics, Zoology), as well as those that demonstrate less interaction, in addition to the relative role of each as a hub for other areas (Biochemistry, Education, Chemistry, and Sociology). This study contributes to a deeper understanding of the role of defense committees in the advancement of Brazilian science, emphasizing the relevance of interdisciplinary collaboration in the evaluation of academic output.

Keywords

Academic Committees; *CAPES*; Graph Theory; Interdisciplinarity; Knowledge Areas; Motifs; Scientometrics; Social Networks; Scientific collaboration; Doctoral dissertations; Theses; Knowledge Discovery; Patterns; Brazil.

1. Introduction

The evaluation of academic works, such as dissertations and theses, constitutes a fundamental pillar for the advancement of scientific research. Academic committees, comprising specialists from diverse scientific fields, play a crucial role in this process. However, the composition of these committees exhibits a notable diversity.

Given the prerogative of assuming this role of judgment, academic committees are, as a rule, composed of individuals qualified to evaluate the work in question, in addition to considering the presentation of the candidate-author (**Roberts**, 2010). The composition of these committees can vary both regarding the number of participating members, as well as their affiliations, areas of academic background, and research. The intended degree (Master's or Doctorate) and the program modality (academic, professional, or industrial) also influence the composition.

Interdisciplinarity, a central concept for this study, refers to the interaction between two or more disciplines, aiming at the integration of knowledge and methods (**Huutoniemi** *et al.*, 2010; **Karlqvist**, 1999; **Porter** *et al.*, 2007). In the composition of academic committees, interdisciplinarity can enrich the evaluation of academic works, providing a more comprehensive and critical perspective (**Morillo** *et al.,* 2001).

Considering the dynamic nature of academic committee composition, we propose that studies could be undertaken to analyze them empirically, which would yield a more profound understanding of their role as an integral agent of scientific knowledge. In this study, we consider these committees as a "temporary institution" wherein experience and expertise are shared among all participating members, not solely during the oral examination itself, but from the moment of their formation.

The members of a given academic committee establish relationships among themselves, contingent upon the role individually performed by each member: the evaluated, the evaluator (invited), or the advisor. One feasible approach to model these relationships lies in Graph Theory, the study and application of which have demonstrated increasing prominence in both the academic sphere and industry (**Benzi** *et al.*, 2016). By employing graphs in the modeling of academic committees, the relationships among the members can consequently be treated as a social network, allowing us to benefit from tools, software, methods, techniques, and algorithms that furnish significant information regarding the structure, characteristics, and functions of the network, at both a local level (a specific community, for instance) and globally. Furthermore, through the exploration of these graphs, we can detect and characterize recurring patterns in their topology, thereby affording a more profound understanding of the emergent interactions.

Graph analysis, in turn, emerges as a powerful tool for the study of interdisciplinary scientific collaboration (**Newman**, 2001). Graphs allow for the representation of relationships between different KAs, identifying patterns of collaboration and areas of greater interaction (**Koutra**; **Faloutsos**, 2017). They are composed of vertices (which represent entities) and edges (which represent the possible connections between these entities). When the edges originate from a source to a destination, we are dealing with a *directed graph*. In this study, we analyze the relationships among the members of academic committees to ascertain the extent of interaction between their respective KAs, and according to the role performed by each member. Consequently, we derive a measure of interdisciplinarity within the realm of scientific research.

In this study, we analyze the relationships among the members of academic committees to ascertain the extent of interaction between their respective KAs, and according to the role performed by each member. Consequently, we derive a measure of interdisciplinarity within the realm of scientific research

2. Academic Graphs

Three interactions manifest when we consider the composition of academic committees. Former members, having been evaluated, can return to academia fulfilling two other roles: as examiners/evaluators (upon receiving invitations to serve as judges of a manuscript), or indeed as advisors:

Evaluations: relationships established between members acting in the capacity of evaluators and members acting in the capacity of candidates (postgraduate students).

Invitations: formal procedures wherein an offer is extended by the supervising member to the evaluating members (academic or otherwise), with the aim of forming a specific academic committee for the subsequent evaluation of a manuscript. Invitations represent the relationships established between members acting in the capacity of advisors (*inviters*) and members acting in the capacity of evaluators (*invitees*).

Co-participations: relationships established among all members (candidate, evaluators, advisor) within an academic committee. That is to say, irrespective of the role assumed by each member, a connection exists among them all from the moment the committee is formed.

Herein, we collectively designate these iterations as *Academic Graphs*. To exemplify the modeling of academic committee members and their relationships, Figure 1 illustrates the prospective graphical representations for two distinct groups: one constituted for Dissertation A, and another for Thesis B. The vertices are identified by the designation of the members (M1 to M7).

Committee	Manuscript	Supervisor	Evaluated	Evaluators	Co-participants
Dissertation A	1	M1	M2	M1, M3, M4	M1, M2, M3, M4
Thesis B	2	M5	M6	M1, M3, M7	M1, M3, M5, M6, M7

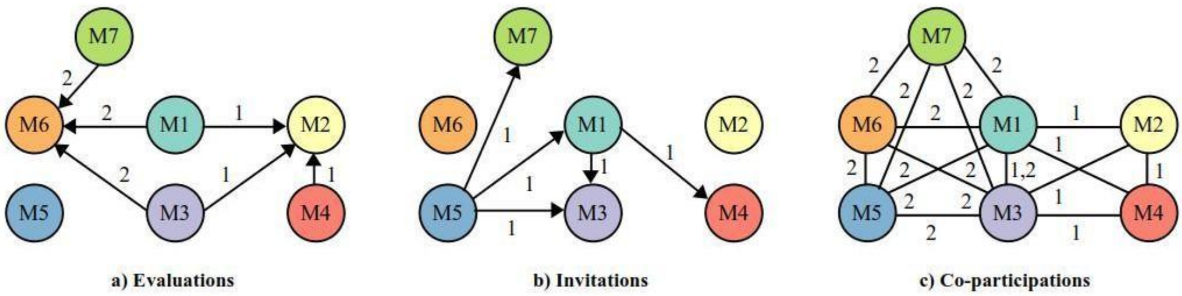


Figure 1. Composition of two hypothetical academic committees and their representations as academic graphs.

The academic **evaluations** graph is directed, with each edge originating from the evaluating member (source) and terminating at the evaluated member (destination). The attribute of the edges (label) indicates the identifier of the dissertation/thesis (manuscript). Within this academic graph (Figure 1(a)), we observe 6 edges, represented by the elements of the set
E={(M1,M2),(M1,M6),(M3,M2),(M3,M6),(M4,M2),(M7,M6)}.

In the **invitations** academic graph, which is also directed, each edge originates from the advisor (source) and terminates at the evaluator (destination). A member may fulfill both roles in distinct academic committees, as depicted in Figure 1(b): M1 serves as both an inviter (manuscript 1) and an invitee (manuscript 2). M3, conversely, was the sole member to receive two invitations (one from M1 and one from M5), with an in-degree of 2. M2 and M6 are isolated vertices within this graph, as they represent evaluated members who neither extend nor receive invitations. Within our illustrative example, only a singular invitation was extended/received in each relationship. In this academic graph, the set of edges is defined as

E={(M1,M3),(M1,M4),(M5,M1),(M5,M3),(M5,M7)}.

The **co-participations** graph stands as the singular instance among the academic graphs wherein the relationships are non-directed, with a connection established between each pair of vertices within an identical committee (Figure 1(c)). In this illustrative example, a clique[1] is present for each committee. Within this graph, concerning the co-participants of dissertation A (manuscript 1), we identify 6 edges within the set E={(M1,M2),(M1,M3),(M1,M4),(M2,M3),(M2,M4),(M3,M4)}. Conversely, for the co-participants of thesis B (manuscript 2), we observe 10 edges within the set E={(M1,M3),(M1,M5),(M1,M6),(M1,M7),(M3,M5),(M3,M6),(M3,M7),(M5,M6),(M5,M7),(M6,M7)}.

2.1. *Motifs* in Graphs

An important concept in graph analysis is that of *motifs*, which are patterns of subgraphs that repeat frequently within a network (**Milo** *et al.*, 2002; **Ribeiro** *et al.*, 2021). Figure 2 displays an example of a *motif* in a graph.

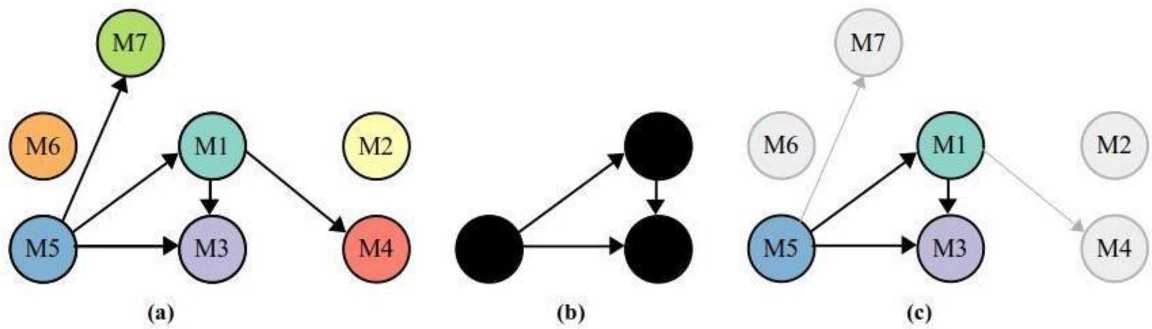


Figure 2. Illustration of **(a)** an original directed graph *G* with 7 vertices and 5 edges, **(b)** a *motif G'* of size 3, and **(c)** the highlighted occurrence *G''* of the *motif G'* in the original graph *G* (**Schwöbbermeyer**, 2008).

The nomenclature "network motif" was introduced to designate particular subgraphs that embody localized configurations of interconnection among the constituent entities of a network. These *motifs* were initially posited as statistically significant and over-represented patterns of localized interconnections within complex networks. In the context of biological networks, it is hypothesized that their structure was shaped through evolutionary processes, thereby imbuing these *motifs* with functional significance (**Schwöbbermeyer**, 2008).

The identification of *motifs* can reveal significant structures and functions within the network, providing insights into the organization and behavior of the system it represents.

The identification of *motifs* can reveal significant structures and functions within the network, providing insights into the organization and behavior of the system it represents

Within our investigation, we employed network *motifs* to extract the most recurrent relationships among the constituent members of the generated networks and, based on the predominant KA of each member, to enable the measurement of interdisciplinarity in the composition of academic committees.

3. Related Works

We delineate in this section relevant studies that bear a direct correlation with one or more topics under consideration within this study.

The entirety of these inquiries substantiates the prominence of *motifs* in the analysis of intricate networks, thereby affording valuable insights into the dynamics of interactions and the structural organization of diverse network topologies. The application of *motif* analysis methodologies elucidates pertinent information regarding the inherent characteristics of collaboration, communication, and the dissemination of information across a spectrum of contexts.

3.1. Bibliometric and/or Scientometric Studies

Yu *et al.* (2023) undertake an examination of the evolution of interdisciplinary citation networks through the application of colored network *motifs*, with a specific focus on the field of Perovskite Materials. The authors posit that colored *motifs*, wherein nodes are assigned colors representing distinct disciplines, possess the capacity to reveal intricate patterns of interdisciplinary collaboration. The source constructs an interdisciplinary citation network of publications within the studied field and analyzes the temporal evolution of both colored and uncolored *motifs*. The findings indicate a shift from knowledge integration towards knowledge diffusion within the field, accompanied by an increasing significance of *motifs* representing interactions among multiple disciplines.

A framework for studying linkage patterns and the evolution of domain knowledge structures using dynamic keyword co-occurrence networks (KCNs) was proposed by **Wang** *et al.* (2024). The authors advocate that analyzing the incremental evolution of KCNs, in conjunction with the properties of individual keywords and the emergence of network *motifs*, furnishes a comprehensive understanding of the dynamics of knowledge structure.

Focusing on structured databases, **Arroyo-Machado** (2024) underscored Wikidata's importance as an open data asset for interdisciplinary social science research, addressing social, economic, and cultural inquiries. Its graph-based framework is essential for illustrating the complex interrelations of entities and concepts, allowing for multifaceted analysis across disciplines such as History, Sociology, Economics, Political Science, and Anthropology. By structuring these intricate connections, researchers can identify patterns and relationships often missed by conventional approaches.

3.2. Social Networks Studies

A method for discovering *motifs* in real-world social networks has been proposed by **Romijn** *et al.* (2015). This method employs a combination of machine learning techniques and network analysis to identify recurring patterns within the network structure. This approach enables the efficient and effective analysis of large-scale social networks.

While traditional measures of centrality remain valuable instruments, the analysis of network *motifs* offers a distinct understanding of the dynamics governing information flow within online social networks. By identifying prevalent interaction patterns and the roles of individual nodes, researchers and practitioners can glean more profound insights into information dissemination, user behavior, and network dynamics (**Sinha** *et al.*, 2022).

4. Methodological Procedures

The application of data mining and knowledge discovery techniques (Knowledge Discovery in Databases - KDD) can assist in the identification of relevant patterns and insights within academic committees data (**Fayyad**, 2001; **Fayyad** *et al.*, 1996). KDD enables the extraction of useful and non-obvious knowledge from large volumes of data, which can be fundamental for the analysis of interdisciplinarity and the impact of scientific collaboration, given that academic repositories serve a uniquely informational purpose.

In this section, we present a computational method for knowledge discovery in datasets containing information on the composition of academic committees (Figure 3). The proposed method is structured within a Knowledge Discovery Process (KDP), which comprises the following stages:

We present a computational method for knowledge discovery in datasets containing information on the composition of academic committees. The proposed method is structured within a Knowledge Discovery Process (KDP)

Data Collection: this preliminary stage entails the extraction of salient data and information from one or more repositories. The methodology employed for this extraction is contingent upon the data's architectural framework and considerations pertaining to security protocols, privacy safeguards, and confidentiality. Data repositories may exist in diverse formats, encompassing unstructured data and distributed systems, potentially necessitating further integration procedures. Automation of the extraction process is often requisite, leveraging programming languages and utilities such as web scraping, scripting, and SQL. The data may be accessible in formats including plain text, web pages (HTML, XML, JSON), tabular datasets (CSV, TSV), relational database systems, and document archives. Following collection, ensuring the data's integrity is paramount, after which it is submitted to the subsequent KDP stage: *preprocessing*.

Preprocessing: the collected data undergoes preparation for analytical procedures. Despite the extensive data available from numerous online repositories, it is not always in a condition suitable for immediate analysis. Therefore, preprocessing is conducted to ensure data integrity through careful treatment and refinement. Common data quality issues include input errors, a lack of referential integrity constraints, inconsistent formatting, and data redundancies. The operations performed at this stage involve the removal of irrelevant data elements, the handling of missing data, text normalization, the elimination or replacement of non-standard characters, and case standardization. Furthermore, preprocessing includes the systematic organization and storage of the data in formats appropriate for subsequent analytical stages, such as tabular structures (CSV or TSV) and bibliographic citation formats (e.g., RIS).

Transformation / Visualization: this stage involves converting the preprocessed datasets into formats suitable for representing the interrelationships among members of academic committees within graph-based structures. Data is transformed into formats that facilitate both the storage and exchange of graph-oriented content. Several methodologies and algorithms for identifying *motifs* may be explored, potentially requiring the adoption of specific representational formats for the graphs in this stage (such as G, GDF, and GraphML). Upon completion, the data is structured in a configuration that reveals the interconnections among committee members, specific to each academic graph. Data visualization is also considered, employing graphical tools that provide functionalities for the empirical analysis of networks, including layout algorithms, filters, stylistic attributes, and color encodings. The labeling stage complements the graph models by incorporating pertinent information crucial for the following stages.

Labeling: this stage of the KDP encompasses two primary phases, both essential for quantifying interdisciplinarity within the constitution of academic committees. The initial phase involves the categorization of academic committees according to a specific KA (information regarding the KA of the manuscript may be present within the original dataset or obtained from external resources). The next phase entails the generation of attributes for the graph vertices, reflecting the specific KA of the members, based upon their overall participation in the committees. The labeling of members considers the primary KA for each individual, determined by the frequency of their participation in the committees (in instances of ties, a resolution criterion is implemented, such as alphabetical ordering of the area).

Graph Mining: algorithms are applied to the academic graphs, with due consideration given to the topology of the networks. The graph mining process is executed in two distinct phases, for each *motif* size under consideration:

Discovery and enumeration of characteristic motifs: during this phase, the principal objective is to discern the patterns of interconnection that exhibit the highest frequency of recurrence within the graph structures. The scope of *motif* discovery is delimited to specific dimensions owing to computational resource implications.

Search for motifs (subgraphs) within the original networks: the academic graphs undergo a subsequent processing phase to locate subgraphs that correspond to the *motifs* identified in the preceding phase. This search process considers the topology of the subgraphs, often employing matrix-based representations of the *motifs*. Following the identification of interconnections among members that reflect the structure of the *motifs*, ordered sets of elements are extracted to represent the interaction between the KAs. This extraction takes into account the semantic interpretation of each academic graph (which member invites, evaluates, or co-participates) and the principal KA of each member.

Post-processing: in this stage, the ordered sets of elements representing the interrelationships between disparate KAs are employed, in conjunction with their frequency of occurrence. The original dataset undergoes transformation to yield Weighted Graphs of Interdisciplinarity (WGIs), wherein the vertices denote the KAs and the edges indicate the interconnections between these domains. WGIs are generated in comprehensive configurations to facilitate visualization and statistical analysis.

Measurement: the final stage of the KDP is oriented towards the eliciting of knowledge derived from the WGIs. To this end, relevant properties and metrics are extracted from the graph structures, which are subsequently employed in the generation of feature vectors. Principal Component Analysis (PCA) is applied to the analysis of interdisciplinarity, considering diverse configurations of the graphs, encompassing both those incorporating and those excluding the influence of *motifs*. The measurement procedure entails the preparation of the feature vectors, the normalization of the constituent data, and the application of PCA to ascertain the KAs exhibiting the most significant relative importance within the compositional networks of academic committees.

This meticulously delineated methodology facilitates a comprehensive analysis of interdisciplinarity within the composition of academic committees, spanning the initial collection and preparation of data through to the ultimate measurement and interpretation of results.

5. Dataset

We employed a Brazilian dataset to discover knowledge from *CAPES*[1] Theses and Dissertations Catalog. This catalog constitutes a publicly accessible repository of quantitative information. According to the Coordination itself, the bibliographic information is directly provided by graduate programs nationwide, which assume responsibility for the veracity of the data.

We employed a Brazilian dataset to discover knowledge from *CAPES*[1] Theses and Dissertations Catalog. This catalog constitutes a publicly accessible repository of quantitative information

The total volume of data collected (information comprising over half a million manuscripts which include academic committee compositions) represents the Brazilian *stricto sensu* (postgraduate master's and doctoral courses) academic output over an eight-year period, spanning from 2013 to 2020.

The data pertaining to each manuscript were categorized according to Major Knowledge Areas (MKAs), as delineated in Table 1.

[1] Coordination for the Improvement of Higher Education Personnel: *https://catalogodeteses.capes.gov.br/catalogo-teses/#!/*

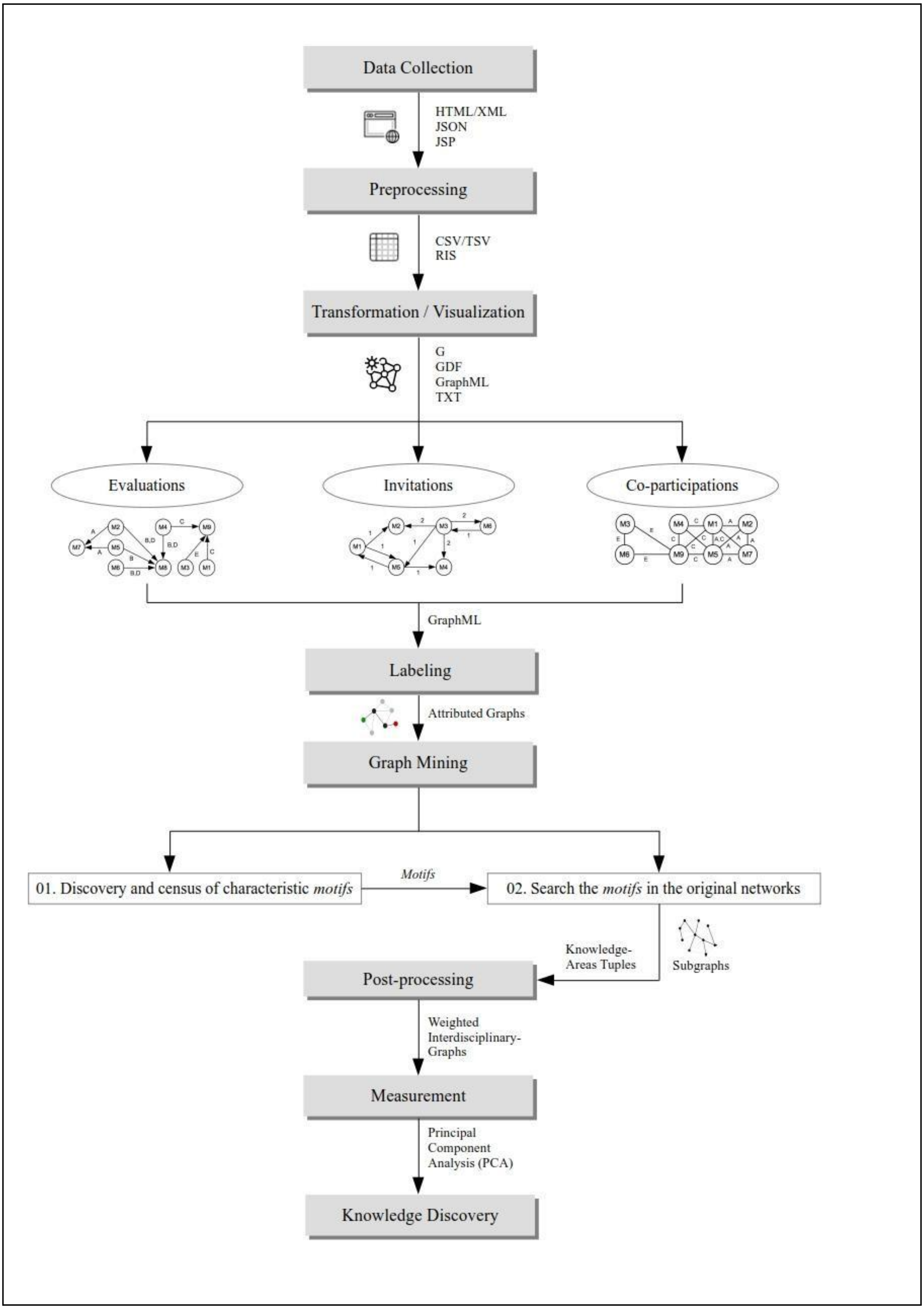


Figure 3. Flowchart of the constituent stages and phases of our KDP.

Table 1. Acronyms for the 9 MKAs and their respective meanings.

Acronym	Meaning
AGR	AGRICULTURAL SCIENCES
BIO	BIOLOGICAL SCIENCES
ENG	ENGINEERING
EXA	EXACT AND EARTH SCIENCES
HUM	HUMAN SCIENCES
LIN	LINGUISTICS, LITERATURE, AND ARTS
MUL	MULTIDISCIPLINARY
HEA	HEALTH SCIENCES
SOC	APPLIED SOCIAL SCIENCES

Table 2 displays the quantities of manuscript data collected, grouped and totaled by year and MKA. The MKA HUM exhibited the highest output with 105,081 manuscripts, closely followed by HEA (99,404). Conversely, the MKA LIN displayed the lowest number of academic outputs during the period (39,997), though with figures proximate to those of BIO (42,826). The cumulative quantity of manuscripts totaled 646,487.

Table 2. Summary of the CAPES manuscript data collection, grouped by year and MKA.

Year / MKA	2013	2014	2015	2016	2017	2018	2019	2020	Total	Percentage
AGR	7,531	7,833	7,796	8,408	8,149	8,542	8,676	7,201	**64,136**	9.92%
BIO	5,317	5,270	5,314	5,343	5,474	5,378	6,038	4,692	**42,826**	6.62%
ENG	7,373	7,587	8,268	8,993	9,225	9,618	9,920	7,858	**68,842**	10.65%
EXA	6,501	6,548	6,933	7,426	7,367	7,454	7,829	6,477	**56,535**	8.74%
HUM	11,077	11,377	12,488	13,397	13,931	14,294	15,461	13,056	**105,081**	16.25%
LIN	4,136	4,417	5,133	5,216	4,920	5,428	5,843	4,904	**39,997**	6.19%
MUL	6,646	7,456	8,486	9,973	10,986	11,999	13,334	11,757	**80,637**	12.47%
HEA	10,806	11,156	11,849	12,839	12,759	13,848	14,261	11,886	**99,404**	15.38%
SOC	8,716	9,430	10,029	11,319	11,900	12,819	13,280	11,536	**89,029**	13.77%
Total	**68,103**	**71,074**	**76,296**	**82,914**	**84,711**	**89,380**	**94,642**	**79,367**	**646,487**	**100.00%**

5.1. Knowledge Areas

Drawing upon an external resource, we proceeded to subdivide the 8 MKAs[2] into 66 KAs, predicated on the specific area of scholarly concentration provided in each manuscript's metadata.

[2] The MKA MUL was not considered, as its very definition encompasses manuscripts addressing more than one discipline.

Table 3 delineates these 66 distinct areas, alongside the respective quantity of manuscripts within each category, totaling 88,064 works that have been classified according to their thematic domain.

Table 3. Summary of manuscripts per KA (n=66), classified and aggregated by MKA. Note: discipline names were translated to their established English equivalents.

MKA [total]	KA (Knowledge Area)	Manuscripts
AGR [6]	Agronomy	307
	Food Science and Technology	2,402
	Agricultural Engineering	305
	Veterinary Medicine	559
	Fisheries Resources and Fisheries Engineering	239
	Animal Science	660
Subtotal		**4,472**
BIO [14]	Biophysics	454
	General Biology	78
	Biochemistry	969
	Biotechnology	380
	Botany	388
	Ecology	2,417
	Pharmacology	1,146
	Physiology	1,586
	Genetics	1,092
	Immunology	701
	Microbiology	915
	Morphology	30
	Parasitology	273
	Zoology	1,729
Subtotal		**12,158**
ENG [9]	Biomedical Engineering	669
	Energy Engineering	229
	Production Engineering	622
	Transportation Engineering	194
	Electrical Engineering	291
	Mechanical Engineering	12
	Naval and Ocean Engineering	338
	Chemical Engineering	1,508
	Sanitary Engineering	9

Subtotal		**3,872**
EXA [8]	Astronomy	227
	Computer Science	6,091
	Physics	2,371
	Geosciences	57
	Mathematics	3,044
	Oceanography	119
	Probability and Statistics	141
	Chemistry	5,610
Subtotal		**17,660**
HUM [8]	Anthropology	1,165
	Archeology	488
	Political Science	1,057
	Education	22,298
	Philosophy	3,550
	History	43
	Psychology	3,568
	Sociology	2,735
Subtotal		**34,904**
LIN [3]	Arts	553
	Literature	87
	Linguistics	2,036
Subtotal		**2,676**
HEA [8]	Physical Education	4
	Nursing	639
	Pharmacy	527
	Speech Therapy	84
	Medicine	489
	Nutrition	560
	Dentistry	893
	Public Health	3,239
Subtotal		**6,435**
SOC [10]	Administration	2,110
	Architecture and Urbanism	105
	Information Science	72
	Communication	907
	Demography	298
	Economics	1,649
	Home Economics	45
	Museology	76
	Urban and Regional Planning	502
	Social Service	123
Subtotal		**5,887**
Total		**88,064**

Subsequent to the labeling of each manuscript according to its respective KA, this same area was attributed to all members constituting its committee. Following this, we quantified the participations of all members across all the committees and determined their primary KA. With the members thus labeled, this enabled the measurement of interdisciplinarity based on the relationships and roles assumed by them (candidate, evaluator, advisor) within the emergent graph networks. It is important to note that only those regions within these networks where characteristic *motifs* were detected were considered for subsequent analysis.

5.2. Discovered *Motifs*

To discover *motifs* of sizes 4 and 5 (phase 01 of Graph Mining stage from Figure 3), we employed the G-Tries algorithm (**Ribeiro**, 2010), which is integrated within the Motif-Discovery App in the graph visualization and analysis software *Cytoscape*.

To illustrate the discovery of *motifs* within the networks, Table 4 presents the 5 most frequent *motifs*, identified after processing the 66 evaluation academic graphs (one for each KA) for sizes 4 and 5.

Regarding the search for *motifs* within the original networks (phase 02), mathematical tools, such as the SageMath software package, were utilized to perform this phase. From the returned subgraphs corresponding to the *motifs*, we extracted relationship pairs (tuples) between distinct Kas based on the primary area of each member, replicating the processing for each academic graph, KA, and *motif* size.

6. Main Results and Conclusions

In this section, we present our most significant findings. Initially, we provide a visual analysis of interdisciplinary interconnections through the utilization of Weighted Graphs of Interdisciplinarity (WGIs). Subsequently, employing the Principal Component Analysis (PCA) statistical technique, we discuss the principal aspects observed regarding the relative role of each KA in scatter plots, preceding our concluding remarks.

We provide a visual analysis of interdisciplinary interconnections through the utilization of Weighted Graphs of Interdisciplinarity (WGIs). Subsequently, employing the Principal Component Analysis (PCA) statistical technique, we discuss the principal aspects observed regarding the relative role of each KA in scatter plot

6.1. WGIs

In this subsection, we present and analyze solely the WGIs derived from the academic **evaluation** graph, considering *motif* sizes of 4 and 5. Both graphs encompass vertices, each representing a distinct KA, with a minimum edge weight exceeding 100.

To facilitate the visual identification of components within the graph layout definitions, the MKAs were discretized by color palette (Table 5). Consequently, the vertices representing KAs belonging to the same MKA share a common color.

Table 4. The 5 most frequent *motifs* discovered in the evaluations academic graphs, distributed by size and their frequency of occurrence across the total number of Kas.

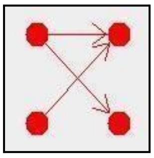
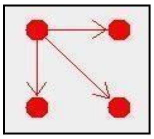
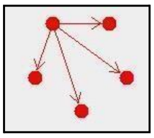
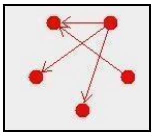
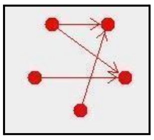
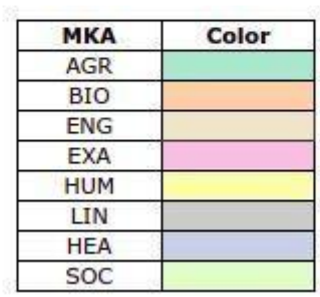
Size	Visual Representation	Occurrences / Total
4		27 / 66
4		39 / 66
5		19 / 66
5		46 / 66
5		01 / 66

Table 5. Color palette for the identification of KAs (vertices) in the graphs, according to their respective MKAs.



MKA	Color
AGR	
BIO	
ENG	
EXA	
HUM	
LIN	
HEA	
SOC	

The resulting graph, derived from the connections of *motifs* of size 4, is displayed in Figure 4. It comprises 22 vertices and 29 edges, exhibiting a minimum edge weight of 101 (tuple (Biochemistry, Microbiology)) and a maximum of 668 (tuple (Ecology, Zoology)).

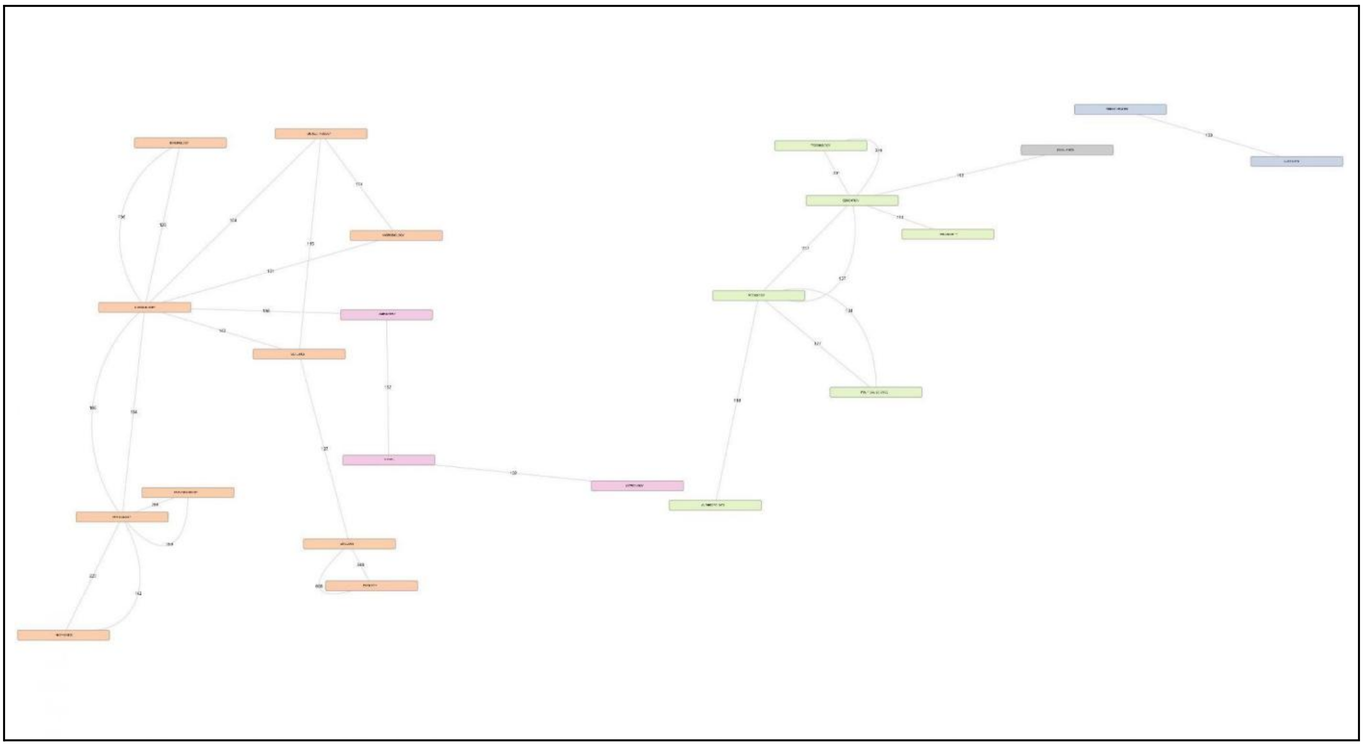


Figure 4. Evaluations WGI for *motifs* of size 4.

For *motifs* of size 5, the resulting graph is displayed in Figure 5. It comprises 23 vertices and 34 edges, exhibiting a minimum edge weight of 101 (tuple (Astronomy, Physics)) and a maximum of 663 (tuple (Ecology, Zoology)).

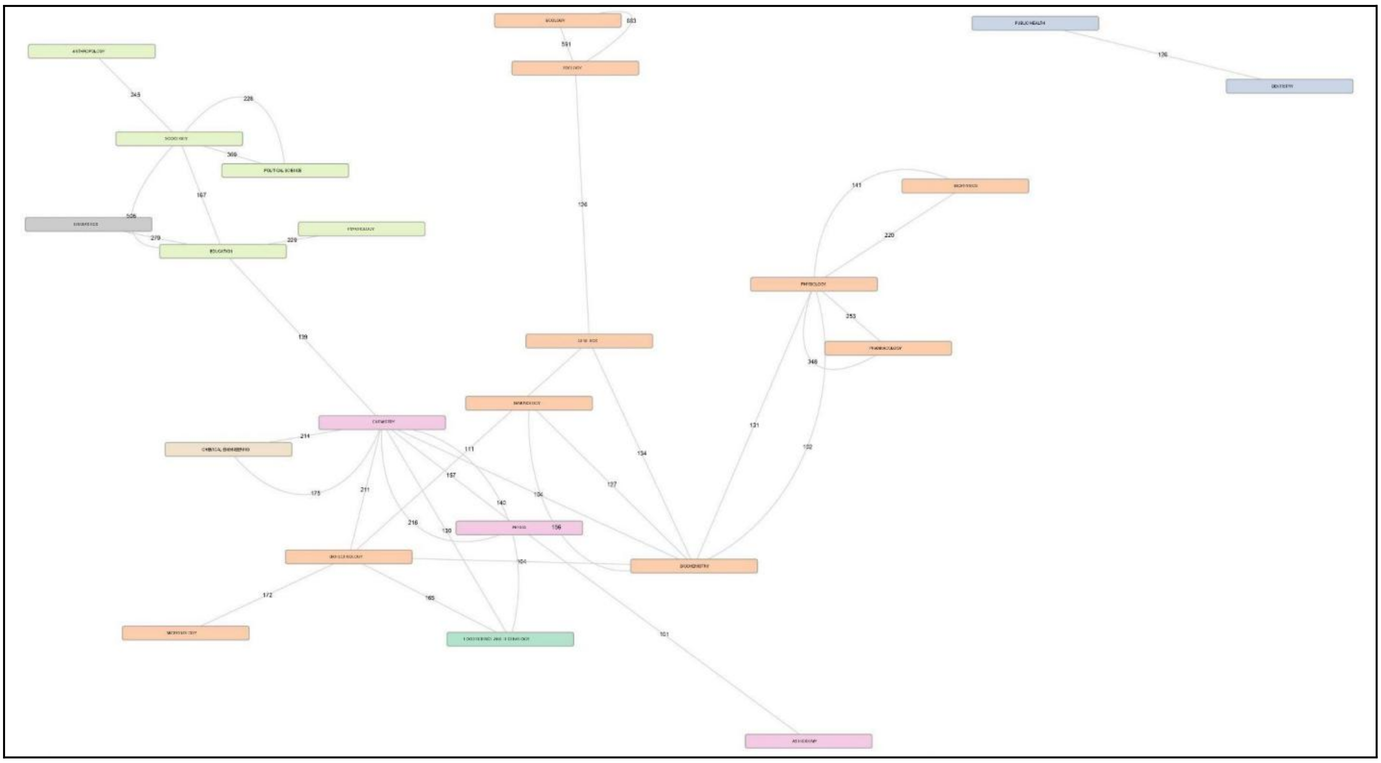


Figure 5. Evaluations WGI for *motifs* of size 5.

6.2. Scatter Plots

In this subsection, we present the findings derived from the measurement of the WGIs, subsequent to the validation of the KDP on our dataset. PCA was applied to feature vectors for two versions: one with *motifs* of sizes 4 and 5 combined, and one without *motifs*.

11 characteristics of the WGIs were considered, encompassing both local and global properties and metrics. These are enumerated below:

1. Clustering coefficient
2. Assortativity
3. Centrality
4. Density
5. Degree
6. Betweenness centrality
7. Maximum edge weight
8. Minimum edge weight
9. Number of edges
10. Number of vertices
11. Transitivity

It is noteworthy that, as a direct reflection of the greater number of labeled academic committees, the KA *Education* exhibited a differentiated behavior in relation to other areas. This fact led us to treat it as an outlier in the measurement analyses, as it would not be equitable to determine its relative importance: during the *Labeling* stage (Figure 3), the values obtained for it were significantly higher.

Figure 6 presents the version of the scatter plot with the combined *motifs*. Taking both axes into consideration, the KAS that exhibited the most prominence were *Sociology*, *Social Service*, and *Biochemistry*.

Figure 7 presents the version of the scatter plot without *motifs*. Taking both axes into consideration, the KAS that exhibited the most prominence were *Biotechnology*, *Political Science*, and *Chemistry*.

We can observe a greater concentration of KAS in the central-left region of the plot in this version, which did not utilize *motifs* to extract more recurrent relationships.

6.3. Conclusions

In this study, we sought a more profound understanding of the composition of academic committees. This composition arises from the distinct roles that members may assume over time (today's evaluated individuals (advisees) may become tomorrow's evaluators and advisors).

Through the analysis of *motifs*, we determined the associations among these members and their frequency of recurrence, which aided us in detecting common situations that, given an appropriate contextual interpretation, may not be conducive to the dissemination of knowledge across different areas (e.g., unidisciplinary committees).

Based on the foregoing, we conclude that the utilization of *motifs* constituted a preponderant factor in achieving a more refined measurement of interdisciplinarity, by extracting from the networks solely those relationships that were demonstrably non-fortuitous among the Kas of the members comprising the academic committees.

Our Principal Component Analysis (PCA), based on the topological properties and metrics of the WGIs, highlighted prominent areas such as Biochemistry, Education, Chemistry, and Sociology. Furthermore, the PCA provided strong evidence suggesting that the versions incorporating *motifs* (whether combined or not) more accurately characterized the areas within the global networks.

Our Principal Component Analysis (PCA), based on the topological properties and metrics of the WGIs, highlighted prominent areas such as Biochemistry, Education, Chemistry, and Sociology. Furthermore, the PCA provided strong evidence suggesting that the versions incorporating *motifs* (whether combined or not) more accurately characterized the areas within the global networks

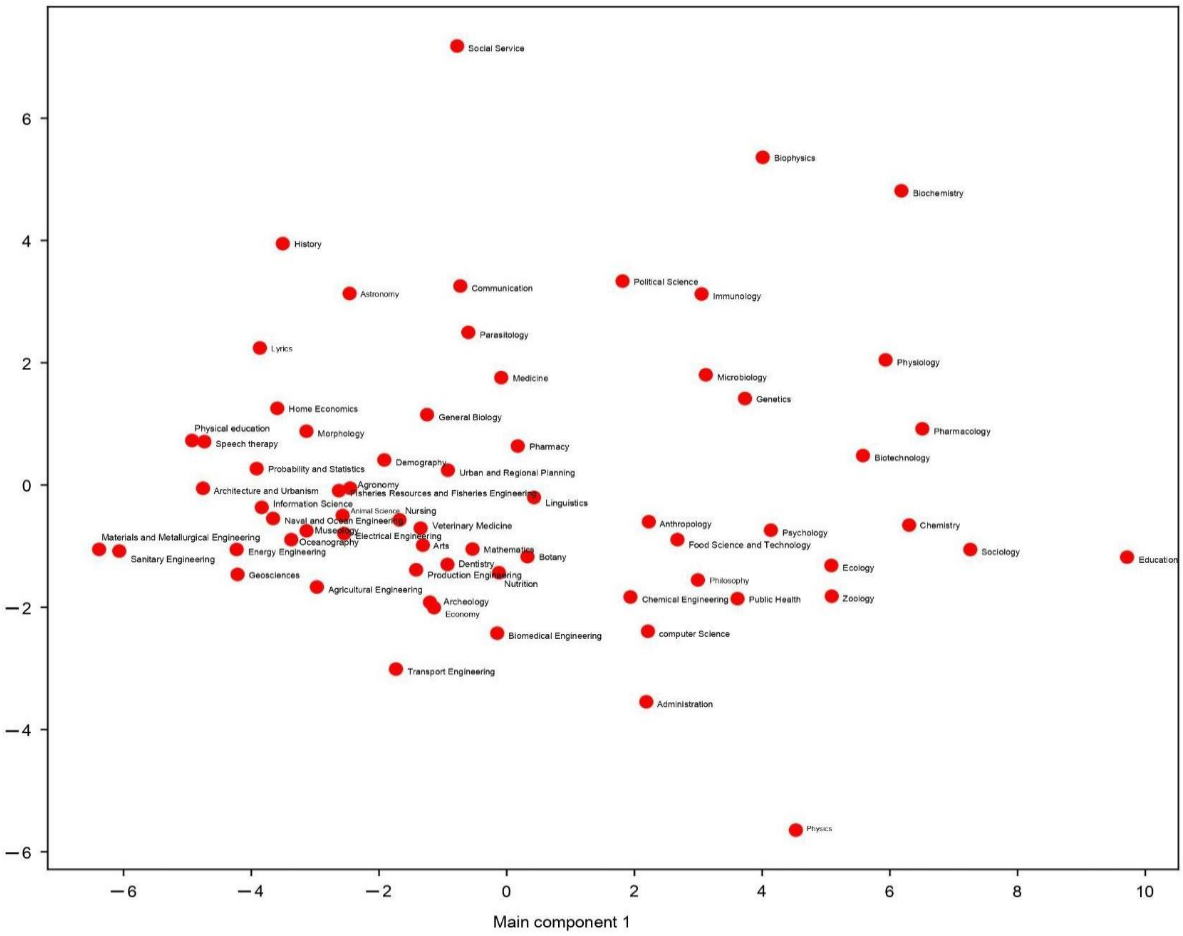


Figure 6. Scatter plot illustrating the distribution of the 66 KAS based on the combined characteristics of the WGIs for *motifs* of sizes 4 and 5.

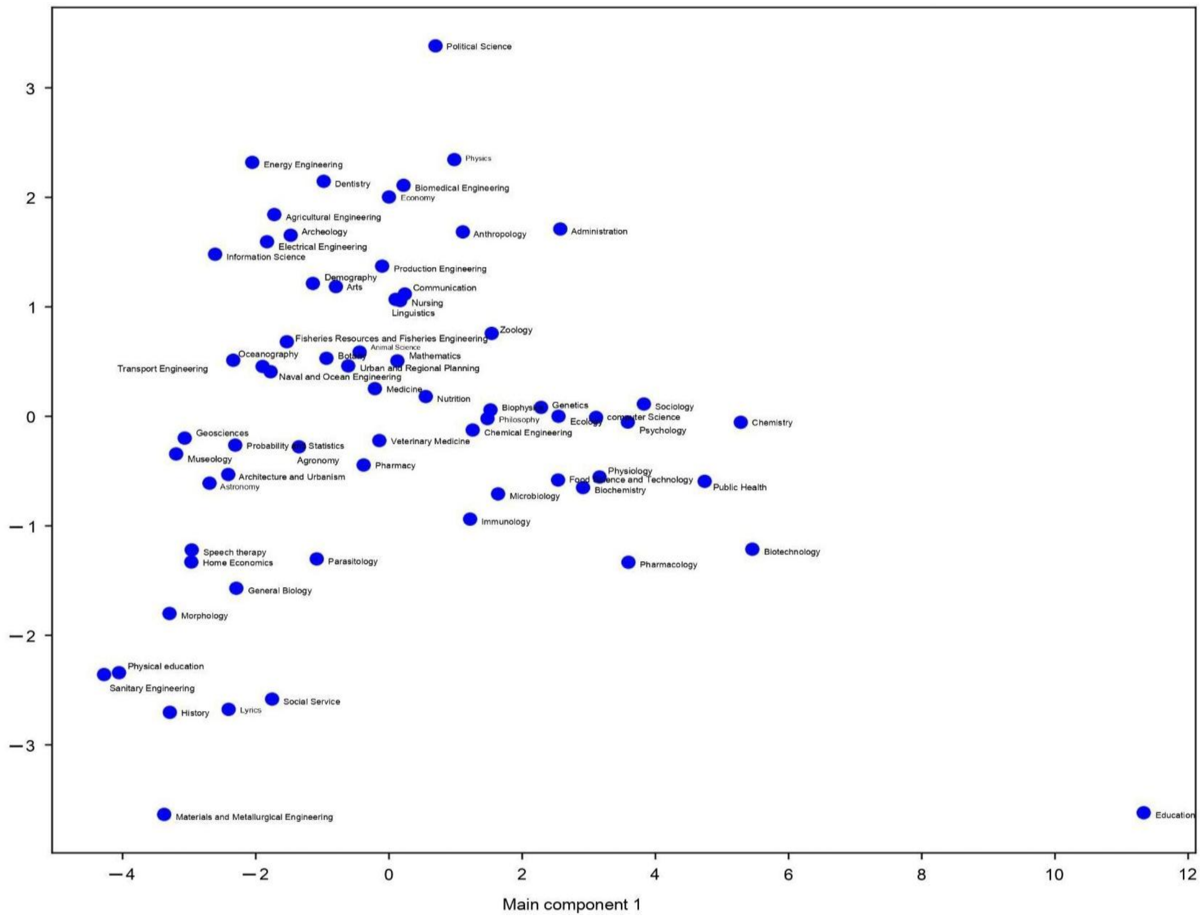


Figure 7. Scatter plot illustrating the distribution of the 66 KAS without the utilization of *motifs*.

The theme of interdisciplinarity was addressed with the aim of elucidating the extent to which disciplines have engaged in dialogue within the 19Brazilian academic environment, at least at the final stage of the evaluation of dissertations and theses that, it is hoped, contribute to the corpus of their respective fields.

7. References

Arroyo-Machado, W. (2024). Exploring Wikidata for Social Science research. *Infonomy*, 2(2)e24034.
https://doi.org/10.3145/infonomy.24.034

Benzi, K.; **Ricaud, B.**; **Vandergheynst, P.** (2016). Principal patterns on graphs: Discovering coherent structures in datasets. *IEEE Transactions on Signal and Information Processing over Networks*, 2(2)160-173.
https://ieeexplore.ieee.org/document/7398168

Fayyad, U. (2001). Knowledge discovery in databases: An overview. In *Relational Data Mining*. Springer Berlin Heidelberg, 28-47.
https://doi.org/10.1007/978-3-662-04599-2_2

Fayyad, U.; **Piatetsky-Shapiro, G.**; **Smyth, P.** (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3).
https://doi.org/10.1609/aimag.v17i3.1230

Huutoniemi, K.; **Thompson Klein, J.**; **Bruun, H.**; **Hukkinen, J.** (2010). Analyzing interdisciplinarity: Typology and indicators. *Research Policy*, 39(1)79-88. *https://doi.org/10.1016/j.respol.2009.09.011*

Karlqvist, A. (1999). Going beyond disciplines: The meanings of interdisciplinarity. *Policy Sciences*, 32(4)379-383.
https://doi.org/10.1023/A:1004736204322

Koutra, D.; **Faloutsos, C.** (2017). *Individual and Collective Graph Mining: Principles, Algorithms, and Applications*. Morgan & Claypool.
https://doi.org/10.1007/978-3-031-01911-1

Milo, R.; **Shen-Orr, S.**; **Itzkovitz, S.**; **Kashtan, N.**; **Chklovskii, D.**; **Alon, U.** (2002). Network motifs: Simple building blocks of complex networks. *Science*, 298(5594)824-827.
https://www.science.org/doi/10.1126/science.298.5594.824

Morillo, F.; **Bordons, M.**; **Gómez, I.** (2001). An approach to interdisciplinarity through bibliometric indicators. *Scientometrics*, 51(1)203-222.
https://doi.org/10.1023/A:1010529114941

Newman, M. E. J. (2001). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E*, 64(1).
https://doi.org/10.1103/PhysRevE.64.016131

Porter, A. L.; **Cohen, A. S.**; **David Roessner, J.**; **Perreault, M.** (2007). Measuring researcher interdisciplinarity. *Scientometrics*, 72(1)117-147.
https://doi.org/10.1007/s11192-007-1700-5

Ribeiro, P.; **Paredes, P.**; **Silva, M. E. P.**; **Aparicio, D.**; **Silva, F.** (2021). A survey on subgraph counting: Concepts, algorithms, and applications to network motifs and graphlets. *ACM Computing Surveys*, 54(2).
https://doi.org/10.1145/3433652

Ribeiro, P.; **Silva, F.** (2010). G-tries: An efficient data structure for discovering network motifs. In: *Proceedings of the 2010 ACM Symposium on Applied Computing*. Sierre, Switzerland: Association for Computing Machinery, 1559-1566. *https://doi.org/10.1145/1774088.1774422*

Roberts, C. M. (2010). *The Dissertation Journey: A Practical and Comprehensive Guide to Planning, Writing, and Defending your Dissertation*. Corwin Press, 2nd edition. *https://www.corwin.com/books/dissertation-journey-4e-283256*

Romijn, L.; **Nualláin, B. Ó.**; **Torenvliet, L.** (2015). Discovering motifs in real-world social networks. In: *SOFSEM 2015: Theory and Practice of Computer Science*. Pec pod Sneûkou, Czech Republic: Springer Berlin Heidelberg, (8939)463-474.
https://doi.org/10.1007/978-3-662-46078-8_38

Schwöbbermeyer, H. (2008). Network motifs. In: *Analysis of Biological Networks*. Gatersleben, Germany: John Wiley & Sons, 85-111.
https://www.wiley.com/en-br/Analysis+of+Biological+Networks-p-9780470041444

Sinha, S.; **Bhattacharya, S.**; **Roy, S.** (2022). Impact of second-order network motif on online social networks. *The Journal of Supercomputing*, 78(4)5450-5478.
https://doi.org/10.1007/s11227-021-04079-7

Wang, Q.; **Zou, B.**; **Jin, J.**; **Wang, Y.** (2024). Studying the linkage patterns and incremental evolution of domain knowledge structure: a perspective of structure deconstruction. *Scientometrics*, 129(2024)4249-4274.
https://doi.org/10.1007/s11192-024-05059-3

Yu, Q.; **Tao, R**.; **Jiang, S.** (2023). Exploring the evolution of interdisciplinary citation network by the colored network motifs: the case of perovskite materials. *Scientometrics*, 128(8)4421-4446.
https://doi.org/10.1007/s11192-023-04777-4