Scientific Collaboration in Brazilian Software Engineering: A Longitudinal Analysis of the SBES Coauthorship Network (1987–2025)

Andréa Sabedra Bordin
Universidade Federal de Santa Catarina
Araranguá, Brasil
andrea.bordin@ufsc.br

Luiza Carolina Miranda Garcia
Universidade Federal de Santa Catarina
Araranguá, Brasil
luiza.cmg@posgrad.ufsc.br

Rafael Parizi
Instituto Federal Farroupilha
São Borja, Brasil
rafael.parizi@iifar.br

Abstract

The Brazilian Symposium on Software Engineering (SBES) reaches its 40th edition in 2026, marking nearly four decades of contributions to Software Engineering research in Brazil. To celebrate this milestone, this paper investigates the evolution of the Brazilian Software Engineering community through the analysis of the SBES coauthorship network. Using bibliographic data from 1,337 papers published across all 39 editions of the event (1987–2025), we constructed and analyzed the complete collaboration network from a Social Network Analysis (SNA) perspective. The study examines structural properties at the network, individual, edge, and group levels, including network evolution, prominent researchers, collaboration ties, and community structures. Additionally, we perform a thematic classification of the largest collaboration community according to SWEBOK knowledge areas. The results reveal the continuous growth and increasing integration of the SBES community, the emergence of a giant component encompassing more than 80% of all authors, the presence of small-world characteristics, and the coexistence of strong collaboration cores with multiple specialized communities. The findings provide a historical perspective on the development of scientific collaboration in Brazilian Software Engineering and contribute to understanding the structure and evolution of its research community over nearly four decades.

Keywords

SBES, Scientific Collaboration, Coauthorship Network, Social Network Analysis

1 Introduction

The Brazilian Symposium on Software Engineering (SBES, in Portuguese) is one of the main initiatives of the Brazilian Software Engineering community, with its first edition held in 1987. Since then, the event has been continuously organized, totaling 39 editions until 2025 and reaching its 40th edition in 2026. Over nearly four decades, the symposium has consolidated itself as an important forum for disseminating research results, promoting knowledge exchange between academia and industry, and discussing emerging trends in the field. Its trajectory reflects the evolution of Software Engineering in Brazil, as well as the growth and diversification of the scientific community itself.

In a scientific event with such a long and continuous trajectory, it is essential to understand how the scientific community has been structured and evolved over time within the national Software Engineering context. Across dozens of editions, the event accumulated a significant volume of scientific collaborations among researchers, reflecting interaction patterns, the formation of research groups, and the consolidation of collaboration networks.

In this context, bibliometric analysis based on coauthorship networks makes it possible to investigate structural aspects of the scientific community, such as connectivity, collaboration intensity, community formation, researcher centrality, and the evolution of collaborative relationships over time [17, 19]. Besides providing a historical perspective on the evolution of scientific collaboration in the field, studies of this nature contribute to understanding the maturity and integration level of the research community, highlighting how coauthorship relationships are associated with patterns of knowledge circulation, partnership formation, and the scientific development of the field [1].

To investigate the evolution of the Brazilian Software Engineering research community, we analyzed the scientific collaboration network that emerged across the 39 editions of SBES (1987–2025). The study adopts a Social Network Analysis (SNA) perspective [25] and examines the collaboration network at network, individual, edge and group levels. Additionally, a thematic classification of publications associated with the largest collaboration community was performed based on SWEBOK knowledge areas.

The study addresses five research questions concerning: (1) the evolution of the collaboration network structure; (2) the most collaborative and strategically positioned researchers over time; (3) the collaboration tie strength; (4) the characterization of collaboration communities; and (5) the predominant Software Engineering knowledge areas represented within the largest community.

The main contributions of this paper are: (i) the first longitudinal analysis of the SBES coauthorship network covering all 39 editions of the symposium; (ii) a comprehensive characterization of the evolution of the SBES collaboration network through multi-level analyses; (iii) a publicly available curated dataset containing bibliographic metadata, coauthorship network data, and associated network measures to support future research; and (iv) evidence-based implications and recommendations for strengthening scientific collaboration within the SBES community.

The remainder of this paper is organized as follows. Sections 2 and 3 present the background and related work. Section 4 describes the research methodology. Section 5 reports the results of the coauthorship network and thematic classification. Section 6 discusses the findings, while Section 7 outlines the threats to validity. Finally, Section 8 concludes the paper and presents future research directions.

2 Bibliometric Analysis and Scientific Collaboration

Bibliometric analysis applies quantitative methods to investigate scientific production through publication metadata, such as authorship, citations, and collaboration patterns [21, 22]. These methods provide valuable insights into the development of scientific fields, enabling the identification of publication trends, influential actors, and patterns of scientific interaction. Among the different bibliometric approaches, coauthorship analysis has become one of the most widely adopted methods for studying scientific collaboration [19].

Scientific collaboration is a fundamental aspect of knowledge production, fostering the exchange of expertise, resources, and ideas among researchers. Since collaborative relationships are often reflected in joint publications, coauthorship data provide an observable and measurable representation of scientific collaboration. Consequently, coauthorship networks have been extensively used to investigate how scientific communities are structured, how researchers interact, and how collaboration evolves over time [17, 19].

2.1 Coauthorship Network Analysis

Coauthorship network analysis models scientific collaboration as a graph $G = (V, E)$, in which authors are represented as nodes and coauthorship relationships are represented as edges connecting pairs of researchers. This representation enables the application of Social Network Analysis (SNA) techniques to investigate the structural organization of scientific communities and the dynamics of collaboration [25].

The study of coauthorship networks gained prominence with the work of Newman [17], who demonstrated that scientific collaboration networks exhibit characteristic structural properties, including high clustering and small-world behavior. In such networks, researchers tend to form cohesive local collaboration groups while remaining connected to the broader community through relatively short collaboration paths. These properties make coauthorship networks particularly suitable for investigating how scientific communities emerge, evolve, and organize themselves over time [1, 8].

2.2 Social Network Analysis Measures

Social Network Analysis provides a set of measures that allow the characterization of collaboration networks from complementary perspectives. In this study, the analysis is organized into network-level, individual-level, edge-level, and group-level measures.

2.2.1 Network-Level Measures. Network-level measures describe the overall structure and cohesion of the collaboration network. Density measures the proportion of existing connections relative to all possible connections, providing an indication of the overall level of collaboration within the network. Diameter and average path length capture the distances between authors and provide insights into the efficiency of information flow across the network. Connected component analysis identifies subnetworks in which authors are directly or indirectly connected. In particular, the giant component, corresponding to the largest connected subnetwork, is commonly used to assess the degree of integration of a scientific community [25].

In addition, clustering coefficient and average path length can be used to investigate small-world behavior [26]. A network exhibits small-world characteristics when it combines high local clustering with short global distances between nodes, enabling efficient communication and knowledge dissemination across the community.

2.2.2 Individual-Level Measures. Individual-level measures focus on the position and importance of researchers within the network. Degree centrality corresponds to the number of collaborators directly connected to an author and is frequently used to identify highly collaborative researchers. Weighted degree extends this concept by considering the frequency of collaborations, capturing the intensity of collaborative relationships. Betweenness centrality measures the extent to which a researcher acts as an intermediary connecting different parts of the network, revealing authors who play strategic roles in integrating collaboration groups [9].

2.2.3 Edge-Level Measures. Edge-level measures focus on the relationships between pairs of researchers. In weighted coauthorship networks, edge weights represent the frequency of collaboration between two authors. These measures allow distinguishing between strong and weak collaborative ties, providing insights into the intensity, persistence, and stability of scientific collaborations [10]. Edge-level analyses are particularly useful for identifying long-term partnerships and core collaboration structures within a scientific community.

2.2.4 Group-Level Measures. Group-level measures investigate the existence of collaboration communities within the network. Community detection techniques identify clusters of authors who collaborate more frequently with one another than with the rest of the network [20]. Such communities often reflect research groups, institutional collaborations, advisor–student relationships, or thematic affinities. The analysis of communities provides insights into the internal organization and specialization of scientific communities.

3 Related Work

Scientific collaboration networks have been extensively studied to understand how research communities are structured and evolve over time. Pioneering work by Newman [17, 19] established many of the foundations of coauthorship network analysis, characterizing collaboration patterns across scientific disciplines and demonstrating the presence of structural properties such as giant components, small-world behavior, and highly connected researchers. Since then, coauthorship network analysis has become a widely adopted approach for investigating scientific communities, collaboration dynamics, and the dissemination of knowledge.

Several studies have analyzed scientific collaboration networks at different levels, including institutions, research areas, and scientific venues. In the Brazilian context, Mena-Chalco et al. [15] investigated the Brazilian scientific collaboration network through topological and bibliometric indicators, while Lima et al. [6] analyzed the coauthorship network of the Brazilian Symposium on Databases (SBBD), examining its structural evolution over three decades. Their study investigated network growth, collaboration patterns, central researchers, and community formation, providing a view of the Brazilian database research community.

Within Computer Science, several venue-centered studies have explored the evolution of scientific communities through coauthorship networks. Maia et al. [14] analyzed the collaboration network of the Brazilian Symposium on Computer Networks and Distributed Systems (SBRC), identifying collaboration communities and investigating structural properties over its first thirty editions. Similarly, Carvalho et al. [5] studied the Brazilian Symposium on Multimedia and the Web (WebMedia), examining both collaboration patterns and thematic evolution. Internationally, studies have also analyzed the coauthorship structures of major conferences such as ACM SIGMOD [16], SIGIR [23], and Digital Libraries conferences [13], revealing how collaboration networks evolve and mature within specific research communities.

Software Engineering has also been the subject of bibliometric investigations. Neto et al. [7] analyzed the internationalization of SBES by examining the participation of foreign researchers and institutions during its first twenty-four editions. However, to the best of our knowledge, no previous study has conducted a comprehensive longitudinal analysis of the SBES coauthorship network covering its entire history.

In contrast to previous work, this paper investigates the scientific collaboration network formed throughout the 39 editions of SBES, spanning the period from 1987 to 2025. We analyze the evolution of the network at multiple levels, including network-level properties, researcher centrality, collaboration intensity, connected components, and community structures. Furthermore, we examine how collaboration patterns evolved across four decades, identifying influential researchers, long-term collaboration dynamics, and the emergence of collaboration communities within the Brazilian Software Engineering research network. By providing a historical and structural perspective on scientific collaboration in SBES, this study contributes to understanding the development, integration, and maturity of the Brazilian Software Engineering community.

4 Methodology

This section describes the methodological procedures adopted to construct and analyze the SBES coauthorship network. The study follows a quantitative, bibliometric approach based on Social Network Analysis, and is organized into four stages, summarized in Figure 1: (i) data collection, (ii) dataset preparation, (iii) coauthorship network analysis at the network, individual, edge, and group levels and (iv) thematic classification of the largest collaboration community. The first three stages describe the structure and evolution of scientific collaboration, while the last examines the research content underlying the most cohesive collaboration core.

4.1 Data Collection

Data were collected through web scraping from the SBES proceedings available in the SBC OpenLib (SOL) and the DBLP Computer Science Bibliography. Proceedings from the first to the twenty-second editions of the event (1987–2008) and from the thirty-third to the thirty-ninth editions (2019–2025) were collected from the SOL platform. Proceedings from the editions held between 2009 and 2018 were collected from DBLP. In total, 1,337 publications from the 39 editions of the symposium were retrieved and analyzed.

4.2 Dataset Preparation

4.2.1 Author Name Deduplication. Semantic duplication of author names is common in bibliographic datasets and may result from inconsistent use of initials, inclusion or omission of name particles (e.g., “de”, “da”), abbreviation of middle names, or differences in author name standardization across publications. The deduplication process was performed using both automated and manual procedures. During the manual validation stage, two researchers independently examined potential duplicates using external sources such as the Brazilian Lattes Curriculum Platform. After the deduplication process, a unique canonical identifier was assigned to each author.

4.2.2 Generation of Coauthorship Pairs. After author names were normalized, all coauthorship relationships were processed to identify pairs of coauthors and the frequency of collaboration between them. As a result, a GEXF file was generated containing the list of unique authors (identifier and name) and the weighted coauthorship relationships between author pairs.

4.3 Coauthorship Network Analysis

The structural analysis of the SBES coauthorship network was performed using the Gephi software [2], which provides a set of Social Network Analysis (SNA) metrics for investigating the organization and evolution of collaboration networks. The analysis was conducted at four complementary levels: network-level, individual-level, edge-level and group-level.

At the network level, structural measures such as density, diameter, connected components, and giant component were calculated to characterize the overall connectivity and cohesion of the scientific collaboration network. In addition, longitudinal analyses were performed by dividing the dataset into four temporal periods of approximately ten years each, enabling the investigation of how the network structure evolved over time.

At the individual level, centrality measures were computed to identify influential and strategically positioned researchers within the collaboration network. Specifically, degree centrality, weighted degree centrality, and betweenness centrality were analyzed to investigate collaboration breadth, collaboration intensity, and intermediary roles in the network. Temporal analyses of these measures were also conducted to identify changes in researcher prominence and collaboration dynamics across different periods.

An edge-level collaboration tie strength analysis was additionally performed using weighted edges representing the frequency of coauthorship relationships between pairs of researchers. This analysis allowed us to examine strong and recurrent collaboration ties, which may indicate stable scientific partnerships and consolidated collaboration groups.

Finally, at the group level, we applied community detection techniques based on modularity optimization to identify densely connected collaboration groups within the network’s giant component. These communities reflect cohesive research groups, institutional partnerships, or thematic clusters within the Brazilian Software Engineering community. We performed community detection using the Louvain algorithm [3], which maximizes the modularity metric (Q). Modularity evaluates whether the density of connections within communities exceeds what a random network layout

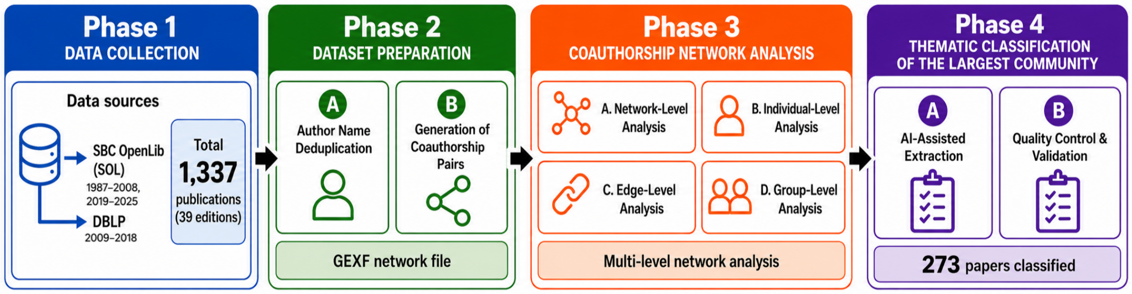

Figure 1: Overview of the methodological procedures, from data collection and coauthorship network analysis to the thematic classification of the largest community.

would yield, where values closer to 1 indicate well-defined, distinct substructures.

We executed the analysis in Gephi software, defining the coauthorship frequencies as edge weights and maintaining a default resolution of 1.0 to balance community granularity. To optimize performance and ensure robustness against local optima, we enabled both the *fast local moves* and *random walk* options. Upon convergence, the algorithm assigned a *modularityClass* attribute to each node, mapping every author to their respective community.

4.4 Thematic Classification of the Largest Community

While network analysis characterizes the structure of collaboration, it does not reveal the scientific content that holds each group together. To address this gap, we conducted a thematic classification of the largest community identified within the network's giant component, which comprises 273 papers. This community represents the main collaboration core identified in the SBES network.

We used the paper titles as the textual unit of analysis. To support the extraction and organization process, we utilized the ChatGPT large language model (GPT-5.5, OpenAI, May 2026), followed by thorough manual validation by two researchers. We structured this individualized analysis of each title into four distinct extraction levels. First, we identified literal terms by directly extracting words from the title related to technical objects, methods, models, artifacts, practices, and SE activities. Second, we interpreted and translated these terms into broader thematic concepts. Third, we synthesized these concepts into an intermediate category describing the main thematic focus of the paper. Finally, we classified each paper according to the 18 knowledge areas of the most recent Software Engineering Body of Knowledge (SWEBOK v4) guide [24], assigning a primary SWEBOK knowledge area.

To ensure traceability and methodological rigor, we recorded a justification for each association. In cases where the title did not allow a reliable connection to a SWEBOK area, we left the classification blank. Additionally, the researchers assigned a confidence indicator (High, Medium, or Low) for each classification, based on the semantic clarity of the title in relation to the assigned area. Finally, a third researcher reviewed the complete classification to ensure and further reinforce the rigor of the process.

5 Results

This section presents the results organized according to the research questions introduced earlier: (1) the evolution of the collaboration network structure; (2) the most collaborative and strategically positioned researchers over time; (3) the collaboration tie strength; (4) the characterization of collaboration communities; and (5) the predominant Software Engineering knowledge areas within the largest community.

5.1 Evolution of the collaboration network structure

The coauthorship network across the thirty-nine editions of SBES consists of 2,298 authors (nodes) and 6,927 coauthorship relationships (edges). The network has a diameter of 11 and a density of 0.003. The network comprises 145 connected components, including a giant (core) component with 1,864 nodes and 6,424 edges, as shown in Figure 2. This shows that a single component accounts for approximately 81.1% of all nodes in the network, suggesting a high level of structural integration in which most authors are indirectly connected through collaboration paths.

In contrast, the remaining 144 components are considerably smaller, as shown in Table 1 that presents the distribution of connected components according to their size (number of authors/nodes). In particular, components composed of only two authors represent the largest proportion of components in the network (31.72%), followed by components with three authors (22.76%) and isolated authors (20.69%). Together, these small structures account for most of the connected components identified in the network. This distribution suggests that, despite the existence of a highly integrated collaboration core, the network also contains a large peripheral region formed by isolated research groups. Such a pattern is commonly observed in scientific communities, where a dominant giant component coexists with numerous small and episodic collaboration groups[17, 19].

To assess whether the SBES coauthorship network exhibits **small-world behavior**, we compared its giant component with an equivalent random graph containing the same number of nodes (1,864) and edges (6,424). A small-world network is characterized by the coexistence of two seemingly contradictory properties: high local clustering and short global distances between nodes [26]. In scientific collaboration networks, this means that researchers tend

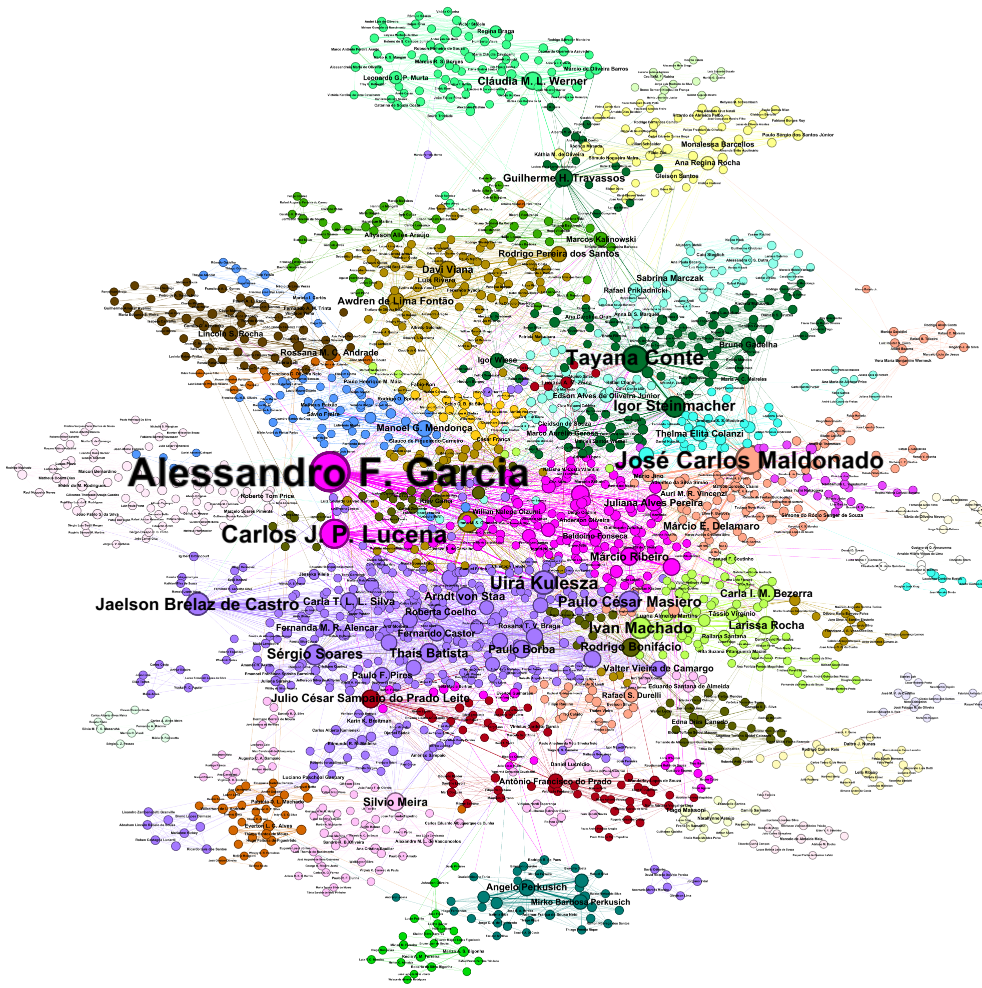

Figure 2: Giant component from SBES Coauthorship Network (1987–2025)

Table 1: Distribution of connected components by size

Authors/Nodes	Nu. of Components	Proportion (%)
1864	1	0.69
19	1	0.69
17	1	0.69
11	2	1.38
10	1	0.69
7	3	2.07
6	4	2.76
5	8	5.52
4	15	10.34
3	33	22.76
2	46	31.72
1	30	20.69

to form tightly connected collaboration groups, while remaining connected to the broader community through relatively few intermediate collaborators [19].

The giant component presented an average clustering coefficient of 0.839, whereas the random graph exhibited a coefficient of only 0.004. Thus, the observed clustering is approximately 210 times higher than expected by chance. Similarly, the giant component contained 14,408 triangles, compared with only 59 triangles in the random graph, indicating a strong tendency for authors to form cohesive collaboration groups.

At the same time, the average path length of the SBES giant component was 4.54, only slightly higher than the 4.12 observed in the random graph. Consequently, although researchers are organized into densely connected local communities, the overall network remains highly integrated, allowing any two authors to be connected through a relatively small number of collaboration steps.

Using the small-world coefficient proposed by Humphries and Gurney [11], the SBES giant component achieved a value of approximately 190.7, far exceeding the threshold of 1 typically used to characterize small-world networks. These results indicate that the SBES collaboration network exhibits a pronounced small-world structure, combining high local cohesion with efficient global connectivity. Such a configuration may favor knowledge circulation, collaboration opportunities, and community integration.

5.1.1 Longitudinal analysis of network structure. The longitudinal analysis reveals a clear evolution of the coauthorship network over time, as shown in Table 2. The number of authors (nodes) and collaborations (edges) increases substantially across periods, with a particularly sharp growth in the most recent years. This expansion is accompanied by changes in network cohesion. While earlier periods are characterized by fragmentation (number of components) and low collaboration intensity (average degree), the network reaches its highest level of cohesion in the 2007–2016 period, as indicated by the peak in density and average degree, and the lowest number of components.

Although the number of components increased again in the last period, this result is likely associated with the rapid expansion of the community and the emergence of new collaboration groups. Similarly, the decline in density observed after 2016 should be interpreted in light of the substantial increase in network size, since according to Newman [18], density tends to decrease in large-scale networks despite continued growth in collaborative activity. Overall, these results indicate the evolution of SBES from a relatively fragmented community into a large and highly interconnected collaboration network.

5.2 Most Collaborative and Strategically Positioned Researchers

At the individual level, we identified the most collaborative and strategic (“bridge”) authors using centrality measures. Table 3 presents the most collaborative authors ranked by weighted degree (WDeg), along with their degree centrality (Deg). The table includes 11 entries because two authors are tied at rank 9. Weighted degree captures both the number of collaborators and the frequency of collaboration. Alessandro F. Garcia has the highest value (235), followed by Carlos J. P. Lucena (156), José C. Maldonado (153), and Tayana Conte (137). These results are consistent with the presence of strong and recurrent collaboration ties, long-term scientific partnerships and cohesive research groups. Degree centrality, in turn, measures the number of distinct collaborators connected to each researcher. Alessandro F. Garcia also occupies the first position according to this metric (120 collaborators), followed by Carlos J. P. Lucena (89), Tayana Conte (85), Jaelson Brelaz de Castro (80), and Sérgio Soares (77). These results indicate that these researchers established collaborations with a large number of distinct authors throughout the history of SBES, suggesting broad visibility and extensive participation within the scientific community.

The comparison between Weighted Degree and Degree provides insights into the collaboration profiles of the most central researchers. Alessandro F. Garcia, Carlos J. P. Lucena, and Tayana Conte exhibit high values in both measures, indicating a combination of broad collaboration networks and strong long-term partnerships. In contrast, authors such as José Carlos Maldonado and Ivan Machado present Weighted Degree values that are disproportionately high relative to their Degree Centrality, suggesting collaboration patterns characterized by frequent interactions within cohesive research groups. Conversely, researchers such as Sérgio Soares and, to a lesser extent, Jaelson Brelaz de Castro display high Degree Centrality combined with comparatively lower Weighted Degree values, indicating broader collaboration diversity with less repeated interaction.

Table 2: Network structural properties over time

Period	Nodes	Edges	Components	Avg. Degree	Density
1987–1996	323	381	82	2.3	0.007
1997–2006	478	935	75	3.9	0.008
2007–2016	539	1840	38	6.8	0.013
2017–2025	1353	4037	65	5.9	0.004

Table 3: Top authors by Weighted Degree

Author’s Name	WDeg Rank	WDeg	Deg Rank	Deg
Alessandro F. Garcia	1	235	1	120
Carlos J. P. Lucena	2	156	2	89
José Carlos Maldonado	3	153	8	66
Tayana Conte	4	137	3	85
Uirá Kulesza	5	119	6	68
Jaelson Brelaz de Castro	6	113	4	80
Ivan Machado	7	104	16	51
Igor Steinmacher	8	99	12	60
Christina Chavez	9	97	9	65
Paulo César Masiero	9	97	11	61
Sérgio Soares	10	95	5	77

To identify the most strategic authors, we used betweenness centrality, which measures the extent to which an author acts as a bridge between different collaboration communities, facilitating collaboration across the network. Table 4 shows the top ten authors with the highest betweenness centrality, normalized such that the maximum possible value is 1.0.

Table 4: Top authors by Betweenness Centrality

Rank	Name	Betweenness Centrality
1	Alessandro F. Garcia	0.097
2	Carlos J. P. Lucena	0.083
3	Tayana Conte	0.059
4	Silvio Meira	0.056
5	Eduardo Figueiredo	0.040
6	Sérgio Soares	0.039
7	Cláudia M. L. Werner	0.038
8	José Carlos Maldonado	0.037
9	Jaelson Brelaz de Castro	0.036
10	Guilherme H. Travassos	0.036

First, Alessandro F. Garcia (0.097) and Carlos J. P. Lucena (0.083) stand out, presenting the two highest betweenness centrality values in the network. The relatively small difference between these two values suggests the presence of co-leading brokers rather than a single dominant intermediary. Both authors occupy positions compatible with important brokerage roles, linking multiple subcommunities within the network. Their positions indicate not only high visibility but also structural and strategic importance for the integration of the scientific community.

A noticeable gap separates the two leading authors from the third-ranked author, Tayana Conte (0.059), indicating a second tier of brokerage. Tayana Conte still occupies an important intermediary position, with a more localized brokerage position compared with the two leading authors. This pattern suggests a hierarchical brokerage structure, in which a small number of highly central actors appear to concentrate much of the inter-group connectivity.

Another relevant aspect is the presence of historically influential researchers such as Silvio Meira (0.056), Cláudia M. L. Werner (0.038), José Carlos Maldonado (0.037), and Guilherme H. Travassos (0.036). These authors are associated with different periods of the SBES history and likely contributed to connecting research groups across generations of the community. Their intermediary positions suggest a long-term structural role across different periods of the network.

Overall, the distribution of Betweenness Centrality indicates that the SBES coauthorship network exhibits a structured and partially centralized collaboration pattern. A relatively small set of authors appears to act as key intermediaries between otherwise weakly connected groups. This configuration is consistent with a mature community characterized by the coexistence of consolidated collaboration clusters and bridging actors that may contribute to knowledge circulation and network cohesion over time.

5.2.1 Longitudinal individual analysis. The longitudinal analysis of authors' weighted degree across the four analyzed periods reveals a dynamic pattern of scientific collaboration, with different authors occupying leading positions over time. Figure 3 shows the top five authors per period. Only three authors (Alessandro Garcia, Carlos Lucena, and José Maldonado) appear among the top contributors in multiple periods, while all other authors appear in the ranking of top collaborators in only a single period. Overall, there is a substantial increase in collaboration intensity in recent years, with a noticeable concentration of high values in the 2017–2025 period. This finding suggests both the persistence of a small group of highly influential researchers and the continuous renewal of prominent contributors within the SBES scientific community.

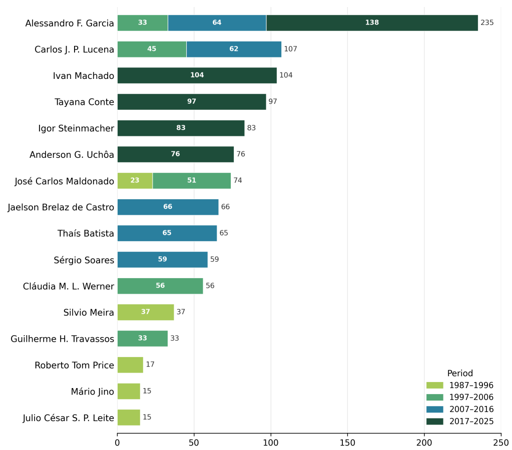

Figure 3: Authors' weighted degree by period

5.3 Collaboration Tie Strength

The presence of an edge in the network indicates that two authors collaborated, but it does not capture the intensity of that collaboration. The weight of each edge, defined by the number of joint publications, allows us to distinguish single-occurrence partnerships from more consolidated collaborations. The distribution of these weights reveals a pattern characteristic of scientific networks: many single-occurrence interactions and relatively few long-lasting relationships [19].

The results confirm the strong predominance of single-occurrence collaborations (Table 5). Of the 6,927 edges, 5,993 (86.51%) correspond to single-occurrence co-authorships between two authors, indicating that most partnerships occur only once. Collaborations with higher recurrence are significantly less frequent: 632 edges (9.12%) have weight 2 and 156 (2.26%) have weight 3. Edges with weight greater than 3, which characterize continuous and consolidated collaborations, account for only 2.11% of all connections, totaling 146 edges, with a maximum weight of 16. This pattern reveals a network largely characterized by sporadic interactions, in which sustained collaboration remains the exception rather than the rule.

Table 5: Distribution of edge weights in the co-authorship network.

Edge weight	Frequency	Percentage (%)
1 (single collaboration)	5,993	86.51
2	632	9.12
3	156	2.26
> 3	146	2.11
Total	**6,927**	**100.00**

To further examine recurrent collaborations based on edge weights, we filtered the data to include only edges with high weights (≥ 10). This procedure identified eight relationships within this range (spanning 10 to 16 coauthorships).

Table 6: Co-authorship ties with weight $\geq$ 10 and their relation to network centrality

Co-authorship	Weight	Centrality pattern
Alessandro F. Garcia and Carlos J. P. Lucena	16	Both central
José C. Maldonado and Mário Jino	13	One central
José C. Maldonado and Márcio E. Delamaro	12	One central
Auri M. R. Vincenzi and José C. Maldonado	11	One central
Ivan Machado and Larissa Rocha	11	One central
Márcio Ribeiro and Rohit Gheyi	11	Neither central
Angelo Perkusich and Mirko B. Perkusich	10	Neither central
Carlos J. P. Lucena and Uirá Kulesza	10	Both central

Table 6 presents the eight strongest coauthorship ties in the SBES collaboration network, considering edges with weight greater than or equal to 10. These ties involve 13 authors. Two pairs include authors who also rank among the top researchers by weighted degree (Table 3), while four ties involve one top-ranked author. The remaining two pairs involve authors outside this ranking, suggesting strong but more localized collaboration patterns. Overall, high-weight ties reveal both globally central partnerships and consolidated collaborations within specialized cores.

5.4 Characterization of collaboration communities

At the group level, we applied community detection to the giant component of the SBES coauthorship network to identify clusters of researchers who collaborate more intensively with each other than with the rest of the network. In this context, a community corresponds to a set of authors densely connected internally, with sparser connections toward other groups. These communities may reflect consolidated research groups, recurring thematic lines, or long-established institutional ties.

Community detection using the Louvain algorithm identified 30 communities in the giant component of the SBES coauthorship network. The largest community comprises 304 authors, representing 16.31% of the giant component. Its size is more than twice that of the second-largest community, which includes 141 authors. The eight largest communities together concentrate more than half of the authors in the giant component (57.8%), while the remaining twenty-two communities account for the other 42.9%.

This asymmetric distribution indicates that scientific collaboration within SBES is concentrated around a limited number of large collaboration communities, while a long tail of smaller and more specialized groups remains active. Such a pattern is commonly observed in scientific collaboration networks and suggests the coexistence of established research communities and more localized collaboration structures, reflecting both the consolidation and diversification of the Brazilian Software Engineering research community.

Table 7 presents the eight largest communities identified, ordered by number of authors, with the top-10 authors in each community ranked by weighted degree centrality, which indicates the volume of recurrent collaborations. For example, Uirá Kulesza (WDeg = 119), Alessandro F. Garcia (WDeg = 235), Christina Chavez (WDeg = 97), Tayana Conte (WDeg = 137), José Carlos Maldonado (WDeg = 153), Sabrina Marczak (WDeg = 54), Cláudia M. L. Werner (WDeg = 73), and Silvio Meira (WDeg = 77) occupy the top positions in their respective communities. The identified communities and their most prominent researchers can also be observed in Figure 2. While the rankings reveal the presence of highly active researchers in all eight communities, the concentration of collaborative activity varies considerably among them. Some communities exhibit a stronger dominance of a few researchers, whereas others display a more balanced distribution of weighted degree values among their leading members.

5.5 Knowledge areas within the Largest Community

To understand both the collaborative structure of the largest community (304 authors) and the research themes underlying these coauthorship relationships, we conducted a thematic classification of the 273 papers authored within this group, which represents the primary collaboration core within the SBES network.

Among these 273 papers, the initial classification assigned High confidence to 135 papers (49%), Medium confidence to 108 (40%), and Low confidence to 10 (4%), while 20 papers (7%) remained unclassified because their titles did not provide sufficient evidence for a reliable assignment. A third researcher independently reviewed the 118 papers that fell into the Medium or Low confidence categories. This review led to a change in the primary assigned area for 18 papers (15.3%), while the remaining 100 cases confirmed the generated classification. Overall, the final dataset maintained a disagreement rate of approximately 7% relative to the entire sample.

Figure 4 presents the number of studies per SWEBOK area. The Software Engineering Models and Methods area concentrated the highest number of papers (n=38), which prompted additional scrutiny during our analysis. We observed that papers in this category contained terms associated with empirical research methodology, such as threats to validity, grey literature, and human-oriented experiments, rather than software engineering methods themselves. Consequently, we revised these classifications or left the area blank according to our agreed-upon criteria. The remaining prominent areas included Software Requirements (n=30), Software Testing (n=27), Software Construction (n=25), and Software Engineering Professional Practice (n=24).

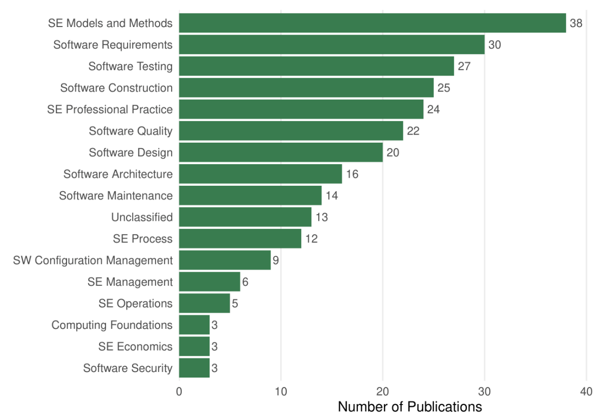

Figure 4: Distribution of publications by SWEBOK area in the largest community of the coauthorship network.

Figure 5 shows the thematic evolution of the largest coauthorship community across decades, measured by the number of publications per SWEBOK area. The temporal analysis reveals a clear shift in the community's research focus over time.

In the 1980s and 1990s, production concentrated on *SE Models & Methods* (n=16 in the 1990s), reflecting a period marked by formal specification, modeling, and system refinement. In the 2000s, the community diversified significantly, with *SW Construction* (n=19), *SW Requirements* (n=18), and *SW Design* (n=14) taking center stage, a period that coincides with the consolidation of aspect-oriented programming (AOP) as the community's central paradigm. In the 2010s, *SE Models & Methods* reclaimed the lead (n=18), accompanied by the growth of *SW Testing* (n=14), signaling a methodological maturation with greater emphasis on verification and quality.

The most pronounced shift occurs in the 2020s, when *SE Professional Practice* emerges as the dominant area (n=22), surpassing all others. This growth may indicate a reorientation toward topics such as software engineering education, developer productivity, professional practices, and human factors in software development. Concurrently, *SW Quality* (n=9) and *SW Configuration Management* (n=7) grow in prominence, while *SW Construction* and *SW Design* become far less frequent. Taken together, these shifts suggest

Table 7: Top-10 authors by weighted degree in the eight largest communities detected in the giant component of the SBES coauthorship network.

Com. 14 — Uirá Kulesza (304 nodes)				Com. 3 — Alessandro F. Garcia (141 nodes)				Com. 6 — Christina Chavez (136 nodes)				Com. 16 — Tayana Conte (119 nodes)			
R	Author	WD	D	R	Author	WD	D	R	Author	WD	D	R	Author	WD	D
1	Uirá Kulesza	119	68	1	Alessandro F. Garcia	235	120	1	Christina Chavez	97	65	1	Tayana Conte	137	85
2	Jaelson B. de Castro	113	80	2	Carlos J. P. Lucena	156	89	2	Awdren de Lima Fontão	67	49	2	Igor Steinmacher	99	60
3	Paulo César Masiero	97	61	3	Anderson G. Uchôa	76	51	3	Davi Viana	66	46	3	Guilherme H. Travassos	70	48
4	Sérgio Soares	95	77	4	Márcio Ribeiro	75	49	4	Rodrigo P. dos Santos	59	45	4	Bruno Gadelha	51	35
5	Thaís Batista	84	60	5	Rohit Gheyi	69	44	5	Luis Rivero	41	31	5	Igor Wiese	48	36
6	Arndt von Staa	75	48	6	Juliana Alves Pereira	64	44	6	Paulo Meirelles	34	27	6	Marco Aurélio Gerosa	41	25
7	Paulo Borba	75	58	7	Baldoino Fonseca	46	30	7	Fabio Kon	24	19	7	Ana Carolina Oran	27	20
8	Carla T. L. L. Silva	70	51	8	Rafael M. de Mello	44	29	8	Maria Istela Cagnin	22	18	8	Williamson Silva	26	21
9	Eduardo Figueiredo	70	60	9	Daniel Coutinho	43	31	9	Fernando Ayach	17	13	9	Mairieli S. Wessel	23	19
10	Rosângela A. D. Penteado	65	44	10	Bruno B. P. Cafeo	42	30	10	Patrícia Matsubara	16	15	10	Maria A. C. Meireles	23	13

Com. 1 — José Carlos Maldonado (106 nodes)				Com. 23 — Sabrina Marczak (91 nodes)				Com. 5 — Cláudia M. L. Werner (85 nodes)				Com. 13 — Silvio Meira (82 nodes)			
R	Author	WD	D	R	Author	WD	D	R	Author	WD	D	R	Author	WD	D
1	José Carlos Maldonado	153	66	1	Sabrina Marczak	54	43	1	Cláudia M. L. Werner	73	45	1	Silvio Meira	77	63
2	Márcio E. Delamaro	68	44	2	Edson A. de Oliveira Jr.	44	38	2	Leonardo G. P. Murta	37	29	2	Eduardo S. de Almeida	32	24
3	Rafael S. Durelli	48	34	3	Rafael Prikladnicki	43	34	3	Regina Braga	31	23	3	Daniel Lucrédio	26	23
4	Auri M. R. Vincenzi	45	21	4	Cleidson de Souza	33	31	4	Márcio de O. Barros	27	19	4	Alexandre M. L. Vasconcelos	18	13
5	Mário Jino	33	10	5	Anna B. S. Marques	23	22	5	Marcos R. S. Borges	24	19	5	Sandro R. B. Oliveira	16	11
6	Vinicius H. S. Durelli	27	21	6	Rafael Chanin	19	14	6	José Maria N. David	19	14	6	Augusto C. A. Sampaio	13	12
7	Simone R. S. de Souza	24	16	7	Caio Steglich	19	14	7	Marta Mattoso	18	14	7	Ana Cristina Rouiller	13	9
8	Adenilso da S. Simão	24	13	8	Gislaine C. L. Leal	18	16	8	Victor Ströele	16	12	8	Carlos E. A. da Cunha	12	10
9	Sandra C. P. F. Fabbri	20	13	9	Itana M. S. Gimenes	15	13	9	Robson Pinheiro de Souza	15	13	9	Iuri Santos Souza	12	8
10	Marcos Lordello Chaim	16	8	10	Cássio A. W. Trindade	14	9	10	Catarina de Souza Costa	14	11	10	Gecynalda S. S. Gomes	9	7

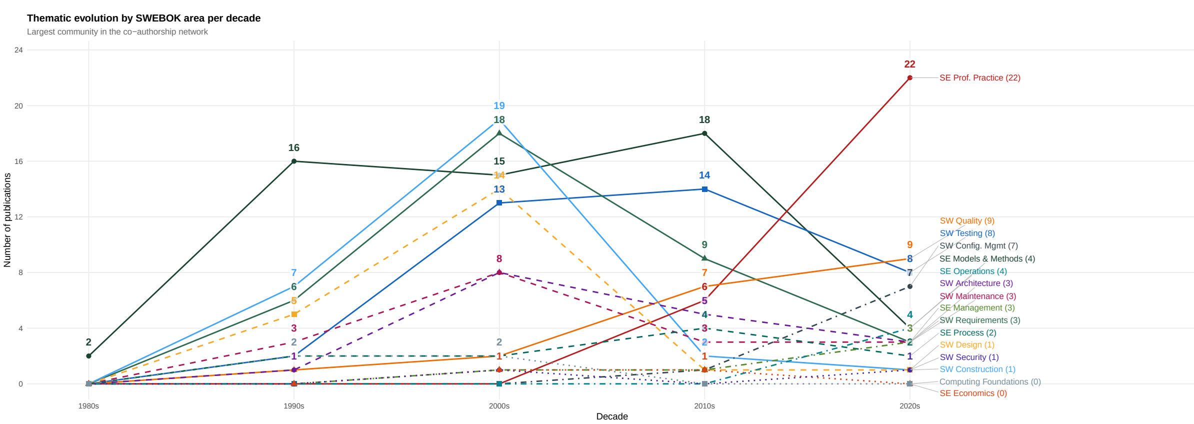

Figure 5: Thematic evolution by SWEBOK area per decade in the largest community of the coauthorship network.

that the community transitioned from a technical and constructive focus, prevalent in the 2000s, toward a people-, process-, and quality-oriented agenda in the 2020s.

6 Discussion

The results provide a comprehensive picture of how the Brazilian Software Engineering research community evolved through scientific collaboration over nearly four decades of SBES. Five main findings emerge from the analysis, from which we derive a set of implications and recommendations for the SBES community.

Growth and Integration of the Community. The evolution of the coauthorship network reveals a continuous expansion of the SBES community, reflected in the increasing number of authors, collaborations, and publications over time. The presence of a giant component encompassing more than 80% of all authors suggests that the community has become progressively integrated, allowing researchers from different institutions and collaboration groups to connect through scientific partnerships. This pattern is consistent with the maturation process commonly observed in scientific communities, where initially fragmented groups gradually become part of a larger and more cohesive collaboration structure [17]. Particularly noteworthy is the sharp increase in the number of authors during the most recent period (2017–2025), which may reflect the expansion of Software Engineering research in Brazil, the consolidation of graduate programs, and the incorporation of new researchers and institutions into the community.

Emergence of a Small-World Collaboration Structure. The small-world analysis indicates that the SBES collaboration network combines strong local cohesion with efficient global connectivity. Researchers tend to collaborate within tightly connected groups while remaining only a few collaboration steps away from most other members of the community. As noted by Newman [17], such a structure may favor knowledge circulation, the exchange of ideas, and opportunities for new collaborations.

Persistence and Renewal of Scientific Leadership. The longitudinal analysis of centrality measures reveals the coexistence of continuity and renewal among the most prominent researchers. While a small number of authors, such as Alessandro Garcia, Carlos Lucena, and José Maldonado, maintained central positions across multiple periods, most highly connected researchers appeared only in specific decades. This finding suggests that the SBES community has continuously incorporated new influential actors while preserving the contributions of established leaders. Such a balance between persistence and renewal may be an important factor in sustaining the vitality and long-term development of the community.

Collaboration Communities and Research Specialization. The identification of multiple collaboration communities within the giant component indicates the existence of well-defined collaboration clusters. These communities likely reflect a combination of institutional collaborations, advisor–student relationships, thematic affinities, and long-term research partnerships. At the same time, the presence of researchers occupying high betweenness positions highlights the importance of cross-community connections in maintaining network cohesion. The unequal distribution of community sizes, characterized by a small number of large communities and numerous smaller ones, further suggests the coexistence of consolidated collaboration structures and more specialized research groups within the SBES ecosystem.

Strong and Persistent Collaboration Relationships. The edge-level analysis revealed a highly heterogeneous pattern of scientific collaboration. While most coauthorship ties were characterized by a single or a small number of joint publications, a limited set of author pairs maintained strong and recurrent collaborations over extended periods. This finding is consistent with patterns commonly observed in scientific collaboration networks, where numerous weak ties coexist with a smaller number of strong and persistent partnerships [12]. Such enduring collaborations likely contribute to the stability of research communities, the continuity of research agendas, and the formation of cohesive collaboration cores that sustain long-term scientific productivity.

Recommendations for the SBES Community. The structural properties observed in the SBES collaboration network suggest that preserving—and, where possible, further expanding—the integration of the scientific community should remain an important objective for future editions of the symposium. In particular, maintaining a large giant component and the favorable balance between local cohesion and global connectivity may help sustain an environment that supports knowledge exchange and the formation of new scientific collaborations.

To this end, SBES could continue promoting opportunities for interaction among researchers from different collaboration communities while encouraging the participation of newcomers in the main collaboration network. Possible initiatives include networking events, collaborative workshops, interdisciplinary sessions, mentoring activities, and other community-building actions that facilitate interactions between experienced participants and first-time authors or researchers from previously disconnected collaboration groups. These initiatives may represent potential mechanisms for strengthening the long-term integration of the SBES research community.

7 Threats to Validity

This study is subject to some limitations that should be considered when interpreting the results. **Data Collection.** The coauthorship network was constructed from bibliographic metadata collected from SBC OpenLib (SOL) and DBLP. Errors or omissions in the original records may affect dataset completeness. **Data Preparation.** Author name disambiguation and thematic classification may introduce inaccuracies. To mitigate these threats, author names were normalized and manually validated, while SWEBOK classifications were reviewed by the researchers. **Network Analysis.** The study adopts coauthorship as a proxy for scientific collaboration, which may not capture all forms of interaction among researchers. In addition, community structures were identified using the Louvain algorithm, whose results may vary slightly under different parameter settings. **Interpretation.** The findings are limited to collaborations represented in SBES publications and may not generalize to the entire Brazilian Software Engineering community.

8 Final Remarks

This paper presented a longitudinal analysis of the SBES coauthorship network covering all 39 editions between 1987 and 2025. Using Social Network Analysis measures, we investigated structural, individual, edge-level, and group-level aspects of scientific collaboration within the Brazilian Software Engineering community. The results revealed a continuous growth of the network, the emergence of a highly integrated giant component, and the presence of small-world characteristics that combine local cohesion with efficient global connectivity.

As future work, we intend to extend the thematic analysis beyond the largest collaboration community by investigating the research topics represented in the remaining communities. Such an analysis may provide a more comprehensive understanding of the thematic diversity, specialization patterns, and knowledge distribution. Furthermore, we plan to complement the collaboration network analysis with bibliometric indicators related to scientific productivity and impact, including publication output and citation-based measures. Combining several bibliometric perspectives may provide a richer characterization of the evolution, influence, and scientific contributions of the SBES community over time.

Artifact Availability

The artifacts are publicly available on Zenodo [4]. The repository includes the raw bibliographic dataset, the curated coauthorship data and edge list, the metadata-collection scripts, the coauthorship network and Gephi metrics files, and the thematic-classification dataset for the largest collaboration community. The artifacts are distributed under the Creative Commons Attribution 4.0.

Acknowledgments

Generative Artificial Intelligence (AI) tools were used to support the preparation of this manuscript. Specifically, ChatGPT (OpenAI) was employed to assist with language refinement and the drafting of selected passages in the Introduction and Background. All AI-generated content was critically reviewed and validated by the authors, who assume full responsibility for the accuracy, integrity, and final content of the manuscript.

References

[1] Albert-László Barabási, Hawoong Jeong, Zoltán Néda, Erzsébet Ravasz, András Schubert, and Tamás Vicsek. 2002. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications* 311, 3–4 (2002), 590–614. doi:10.1016/S0378-4371(02)00736-7

[2] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 3. 361–362. doi:10.1609/icwsm.v3i1.13937

[3] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008, 10 (2008), P10008.

[4] Andrea Sabedra Bordin, Luiza Carolina Miranda Garcia, and Rafael Parizi. 2026. *SBES (1987–2025): Bibliographic Metadata, Coauthorship Network, and Data Collection Scripts*. doi:10.5281/zenodo.21710814

[5] Luiz Paulo Carvalho, Silas Lima Filho, Michele A. Brandão, Mônica Ferreira Da Silva, Jonice Oliveira, and Maria Da Graça Pimentel. 2023. Thirteen Years of WebMedia (2010 to 2022) - A Comprehensive Descriptive and Meta-Scientific Analysis. In *Proceedings of the 29th Brazilian Symposium on Multimedia and the Web* (Ribeirão Preto, Brazil) *(WebMedia '23)*. Association for Computing Machinery, New York, NY, USA, 184–192. doi:10.1145/3617023.3617048

[6] Lucas Henrique Costa de Lima, Guilherme Penha, Leonardo Magalhães de Alencar Rocha, et al. 2017. The Collaboration Network of the Brazilian Symposium on Databases. *Journal of the Brazilian Computer Society* 23, 1 (2017), 10. doi:10.1186/s13173-017-0059-6

[7] Paulo Anselmo da Mota Silveira Neto, Joás Sousa Gomes, Eduardo Santana de Almeida, Jair Cavalcanti Leite, Thais Vasconcelos Batista, and Larissa Leite. 2013. 25 years of software engineering in Brazil: Beyond an insider's view. *Journal of Systems and Software* 86, 4 (2013), 872–889. doi:10.1016/j.jss.2012.10.041 SI : Software Engineering in Brazil: Retrospective and Prospective Views.

[8] Santo et al. Fortunato. 2018. Science of science. *Science* 359, 6379 (2018).

[9] Linton C Freeman. 1978. Centrality in social networks conceptual clarification. *Social networks* 1, 3 (1978), 215–239.

[10] Mark S. Granovetter. 1973. The Strength of Weak Ties. *Amer. J. Sociology* 78, 6 (1973), 1360–1380. http://www.jstor.org/stable/2776392

[11] Mark D. Humphries and Kevin Gurney. 2008. Network 'small-world-ness': a quantitative method for determining canonical network equivalence. *PLoS ONE* 3, 4 (2008), e0002051. doi:10.1371/journal.pone.0002051

[12] Qing Ke and Yong-Yeol Ahn. 2014. Tie Strength Distribution in Scientific Collaboration Networks. *Physical Review E* 90, 3 (2014), 032804. doi:10.1103/PhysRevE.90.032804

[13] Xiaoming Liu, Johan Bollen, Michael L. Nelson, and Herbert Van de Sompel. 2005. Co-authorship Networks in the Digital Library Research Community. *Information Processing & Management* 41, 6 (2005), 1462–1480. doi:10.1016/j.ipm.2005.03.012

[14] Guilherme Maia, Pedro O. S. Vaz de Melo, Daniel L. Guidoni, Fernanda S. H. Souza, Thiago H. Silva, Jussara M. Almeida, and Antonio A. F. Loureiro. 2015. On the Analysis of the Collaboration Network of the Brazilian Symposium on Computer Networks and Distributed Systems: 30 Editions of History. *Journal of the Brazilian Computer Society* 21, 1 (2015), 3. doi:10.1186/s13173-015-0028-9

[15] Jesús Pascual Mena-Chalco, Luciano Antonio Digiampietri, Fabrício Martins Lopes, and Roberto Marcondes Cesar Junior. 2014. Brazilian Bibliometric Coauthorship Networks. *Journal of the Association for Information Science and Technology* 65, 7 (2014), 1424–1445. doi:10.1002/asi.23010

[16] Mario A. Nascimento, Jörg Sander, and Jeffrey Pound. 2003. Analysis of SIGMOD's co-authorship graph. *SIGMOD Rec.* 32, 3 (Sept. 2003), 8–10. doi:10.1145/945721.945722

[17] Mark E. J. Newman. 2001. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences* 98, 2 (2001), 404–409.

[18] Mark E. J. Newman. 2003. The Structure and Function of Complex Networks. *SIAM Rev.* 45, 2 (2003), 167–256.

[19] Mark E. J. Newman. 2004. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences* 101, suppl_1 (2004), 5200–5205. doi:10.1073/pnas.0307545100

[20] Mark E. J. Newman and Michelle Girvan. 2004. Finding and evaluating community structure in networks. *Physical Review E* 69, 2 (2004), 026113.

[21] Derek J. de Solla Price. 1963. *Little Science, Big Science*. Columbia University Press.

[22] Alan Pritchard. 1969. Statistical bibliography or bibliometrics? *Journal of Documentation* 25, 4 (1969), 348–349.

[23] Alan F. Smeaton, Gary Keogh, Cathal Gurrin, Kieran McDonald, and Tom Sødring. 2003. Analysis of papers from twenty-five years of SIGIR conferences: what have we been doing for the last quarter of a century? *SIGIR Forum* 37, 1 (April 2003), 49–53. doi:10.1145/945546.945550

[24] Hironori Washizaki (Ed.). 2024. *Guide to the Software Engineering Body of Knowledge (SWEBOK Guide), Version 4.0.* IEEE Computer Society, Los Alamitos, CA, USA.

[25] Stanley Wasserman and Katherine Faust. 1994. *Social network analysis: Methods and applications*. Vol. 8. Cambridge university press.

[26] Duncan J. Watts and Steven H. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature* 393, 6684 (1998), 440–442. doi:10.1038/30918