

Received August 20, 2021, accepted September 11, 2021, date of publication September 17, 2021, date of current version September 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3113637

Socio-Technical Congruence as an Emerging Concept in Software Development: A Scientometric Analysis and Critical Literature Review

BINISH RAZA¹, **RODINA AHMAD**¹,
MOHD HAIRUL NIZAM BIN MD NASIR¹, (Member, IEEE),
AND SHUKOR SANIM MOHD FAUZI²

¹Faculty of Computer Science and Information Technology, Universiti Malaya, Kuala Lumpur 50603, Malaysia

²Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perlis 02600, Malaysia

Corresponding author: Rodina Ahmad (rodina@um.edu.my)

ABSTRACT Ample evidence in the literature emphasizes using socio-technical congruence (STC) to address coordination issues in distributed software development. The recent decades have shown a progressive growth in STC, resulting in an increasing number of research studies in the scientific corpora. However, no existing study has systematically analyzed and illustrated the research patterns, latest trends, and evolution in STC. This study aims to explore the knowledge structure and create evolutionary trajectories from STC publications. To achieve this aim, a scientometric analysis is performed that combined a critical literature review (CLR) of STC-related published research in the Web of Science and Scopus databases from 2000 to 2020. The scientometric analysis is conducted through four scientometric techniques: 1) co-word network analysis; 2) co-author network analysis; 3) co-citation analysis; 4) document clustering with timeline analysis. The study outcomes will help understand and visualize STC's research status quo. CLR is objectively conducted to recognize the latest research topics, themes, and salient features of STC research in software development. A total of 306 bibliographic data are analyzed to generate study-related networks and density visualizations. The results reveal an evolution in the STC field from its conception to the recent developments of STC models and other related factors. This study primarily contributes to the literature by providing a systematic view related to STC research to assist software practitioners in identifying applications and key research areas. Moreover, the combination of scientometric analysis and CLR reveals key researchers, journals and conferences, institutions, prominent contributing countries, and six major research themes, including “community structure” and “socio-technical congruence” as the most prominent ones.

INDEX TERMS Socio-technical congruence, software development, scientometric analysis, critical literature review.

I. INTRODUCTION

Software development is a complex mechanism due to its inherently multidimensional nature [1]. Development efforts are often divided among teams or team members to cope with this complexity. The distributed organization of team members increases the need for coordination and collaboration

The associate editor coordinating the review of this manuscript and approving it for publication was Jolanta Mizera-Pietraszko¹.

among team members. However, dependencies among tasks change over time, given the dynamic nature of software development, which perturbs team coordination efficiency. Therefore, the coordination and collaboration mechanisms must align with the underlying project's organization to accommodate the dynamic changes in task dependencies.

A recognized approach, socio-technical congruence (STC), is used to overcome the challenges of appropriate communication and coordination by measuring the fit among

an organization's social and technical capabilities [2]. STC focuses on social and technical aspects of the software development process and a fit indicates the right fusion of social and technical abilities within a distributed team. STC helps measure the team coordination level, which helps an organization identify gaps that induce delays in work and results or overall project failure [3]. The literature reveals that a high degree of STC yields positive outcomes for project development concerning quick task resolution time [4], enhanced build quality [5], and improved project quality [6]. However, the misalignment of social and technical dependencies, known as the coordination gap, reduces productivity concerning a growing number of code changes [7], build failure, and low team performance [8].

STC is a complex and evolving research area in software development. However, the literature review demonstrates that the STC concept mainly concentrates on software development areas, such as industry research, distributed software development, open-source software (OSS), and global software development (GSD).

Numerous studies address the salient features and challenges associated with STC measurement. For instance, Cataldo *et al.* [9], [10] used a matrix-based STC model to measure the alignment between task dependencies and actual coordination. An improved weighted STC measurement model was presented in [11], [12] to overcome the limitations of Cataldo's model. The proposed model successfully detected the coordination gaps and suggested prioritizing key coordination tasks that must be managed for better performance. Chouhdhary *et al.* [13] presented another vein of STC, that is STC measurement for open source projects. The model relies on the analysis of units of bursts to compute the collaborative productivity. The results of the empirical investigation depict the positive influence of STC on team performance. Golzadeh *et al.* [14] proposed an empirical study to measure the relationship between STC and cargo package dependencies to show the advantage of STC on OSS ecosystems and community health. Kwan and Damian [11] enhanced the STC model by introducing the concept of awareness in STC measurement. Jiang *et al.* [15] proposed a novel three-dimensional STC measurement model that primarily focuses on measuring the congruence between task dependency and developer coordination. Zhang *et al.* [16] presented an improved STC framework to compute STC and the missing developer links (MDL) metrics at the file-level. The finding of the empirical study reveals the effective relationship between STC and MDL for software bug prediction. Portillo-Rodríguez *et al.* [17] introduced a multi-agent STC model for GSD by utilizing the concept of Kwan's model [12]. The researchers included several additional factors related to environmental needs.

Existing studies have conducted in-depth investigations of STC appositeness in particular areas of research. However, the application of STC is assorted with an anecdotal extent of intricacy. Therefore, an additional research endeavor is

needed to determine the range of STC's applicability and influence in interrelated areas of software development.

A. RESEARCH AIM, GAP, AND SIGNIFICANCE

Numerous researchers have conducted review studies on STC to summarize its different characteristics. For instance, Sierra *et al.* [18] summarized STC-related features, techniques, and tools. Inayat *et al.* [19] further presented a survey of the most relevant socio-technical aspects of requirements-driven collaboration among software development teams. Similarly, Suali *et al.* [20] conducted a systematic literature review (SLR) to identify the importance of coordination in different software development lifecycles. However, existing STC reviews merely focus on manual and qualitative methods, susceptible to literature subjectivity [21].

Over time, several developers and researchers introduced strategies to develop software at a large scale for different contexts. This increase in the number of research publications related to STC increases scientific corpora. An investigation of the knowledge structures of these publications can help in determining the research fronts and development trajectories in the intended field of research [22]. In existing studies, the main focus of researchers was to analyze the contents in STC-related publications, not the research trends. With an increase in the number of research publications in scientific corpora and focus towards software development, the scientometric analysis can help researchers to identify the research patterns, prominent publications, and publication characteristics (such as authors, publication sources, etc.) without going through the detailed study of each paper individually. Thereby, the mapping of STC knowledge facilitates researchers to analyze the existing scientific literature, evolution, trends, growth, and future directions of STC.

In the literature, various approaches have been identified to map the scientific data of intended research areas, such as traditional or narrative review [23], systematic mapping review [24], systematic reviews [25], critical reviews [26], content analysis [27], a bibliometric technique [28]–[31], latent semantic analysis [32], and scientometric analysis [33], [34]. However, scientometric analysis is considered one of the most widely used methodologies to map scientific knowledge efficiently. It helps examine and evaluate the development in the research field and salient frontiers of research using various mathematical techniques and visualizations [35]. Furthermore, it facilitates observing the performance of academics, institutes, faculties, and journals in the targeted research area [36].

Surprisingly, no research has been conducted on STC highlighting the research innovation, achievements, and struggles that embrace emerging trends in the intended research area. This scenario raises difficulties for researchers to integrate the existing knowledge and explore the research topics for future investigation. It is needed to discover the evolutions and development trajectories that can help scholars to better understand the existing STC knowledge and research trends.

Furthermore, as far as identified, no research has been conducted on STC outlining the working relationships and associations among clusters of STC research-related patterns to date, such as journals, researchers, institutions, and regions. Moreover, no study has explored the STC research corpus concerning different aspects, including clusters of co-words, co-authors, co-citations, and evaluations. To this end, we have exploited scientometric techniques to perform an in-depth analysis of STC-related knowledge. This scientometric analysis will help the research academic community to explore the research patterns and topics for the next stage of prospects.

Grounded on the research gap of STC reviews, this paper aims to investigate the knowledge structure and development growth in STC publications. The findings of this investigation facilitate software practitioners and the academic community to understand the latest trends and visualize new perspectives for future studies. To achieve this purpose, this paper performs an in-depth scientometric analysis, coupled with a critical literature review (CLR) of STC. The CLR is based on outcomes of detailed documents' co-citation cluster analysis while considering a timeline. The findings revealed the research themes and related challenges in need of future investigation. To the best of our knowledge, this is the first scientometric study that provides an insightful view of STC's latest trends and status quo in the domain of software development. Consequently, this study will help researchers and practitioners understand the research field and its influence on software development. Moreover, this study sidesteps the issue of subjectivity existing in published STC reviews by combining scientometric analysis (visualizing and analyzing bibliographic data) and CLR.

The rest of this paper is organized as follows. Section II presents the adopted research methodology with a brief overview of the steps followed in conducting this study smoothly. Section III discusses the mechanism of bibliographic data collection and analysis. A detailed analysis and numerous knowledge maps and networks (i.e., created from bibliographic data) are provided in Section IV. It further provides CLR on the timeline analysis of clusters that explore the research premises and subsequent arguments. Section V highlights the results and discusses the scientometric analysis and CLR findings. Finally, the last section concludes the paper by offering an overall summary of the results and providing suggestions for future studies.

II. RESEARCH METHODOLOGY

A review of the methodology is proposed to successfully achieve the research aim. It consists of three major stages (as shown in Fig.1): bibliographic data collection and analysis, scientometric analysis with CLR, and result interpretation. Each stage is supported with an illustration of numerous featured maps and diagrams.

The first stage, bibliographic data collection and analysis, was conducted by gathering academic publications from two data sources: Web of Science (WoS) and Scopus. Moreover, this stage involved a considerable research corpus consisting

of 306 journal articles and proceedings. This corpus was significantly larger than any existing reviews on STC. The bibliographic data collection and analysis helped analyze the written publications statistically. The statistical analysis indicated the number of publications, evaluations, and main trends in the intended research area.

In the second stage, scientometric analysis is conducted, which is considered the most widely used methodology for mapping scientific knowledge efficiently [37]. This approach examines and evaluates developments in certain research fields and presents salient frontiers of research using various mathematical techniques and visualizations. Additionally, scientometric analysis facilitates observing the performance of academic researchers, institutes, journals, and countries in the investigated area of research [38]. The academic publications (collected in the previous stage) were further analyzed via four scientometric techniques (section IV-B) to perform the scientometric analysis (i.e., the second stage). Employing four techniques identified the following results: i) determined a method to deduce the evolution of the STC area; ii) recognized significant researchers, countries, and institutions; iii) determined the key journals and conferences; iv) identified the salient and promising research directions; and v) deduced the origins of the researchers and numbers of publications in specific regions. The scientometric techniques were applied to the data collected using three powerful software tools: CiteSpace [39], VOSviewer [40], and NVivo [41]. In literature, these tools were popularly used for scientometric analysis. As such, the CiteSpace tool was considered valuable for mapping domains of knowledge and generating illustrative graphical maps [42]–[44]. The VOSviewer aids the construction and visualization of the bibliometric density network from the information (i.e., extracted from the scientific literature) [45]–[47]. NVivo is the most commonly used data analysis tool, which employs qualitative and mixed methods to generate bibliographic data results [45], [48]. Furthermore, CLR helps in discovering the prominent research themes with the major studies covered in each theme.

The final stage (i.e., result interpretation) highlighted the detailed discussion about the outcomes and key findings obtained from the previous stage. This stage facilitates identifying possible research directions and related challenges from the scientometric outcomes.

III. STAGE 1: BIBLIOGRAPHIC DATA COLLECTION AND ANALYSIS

A relevant list of publications and bibliographic data were gathered from two popular databases, Scopus and WoS, to establish a foundation for the scientometric analysis. This study intends to cover the maximum existing STC-related literature. Therefore, STC peer-reviewed articles published in the last 20 years were gathered through a precise searching approach (discussed in the subsequent section). The obtained data were further analyzed through a screening process that provided empirical evidence to facilitate the meta-scientific findings. The following sections present

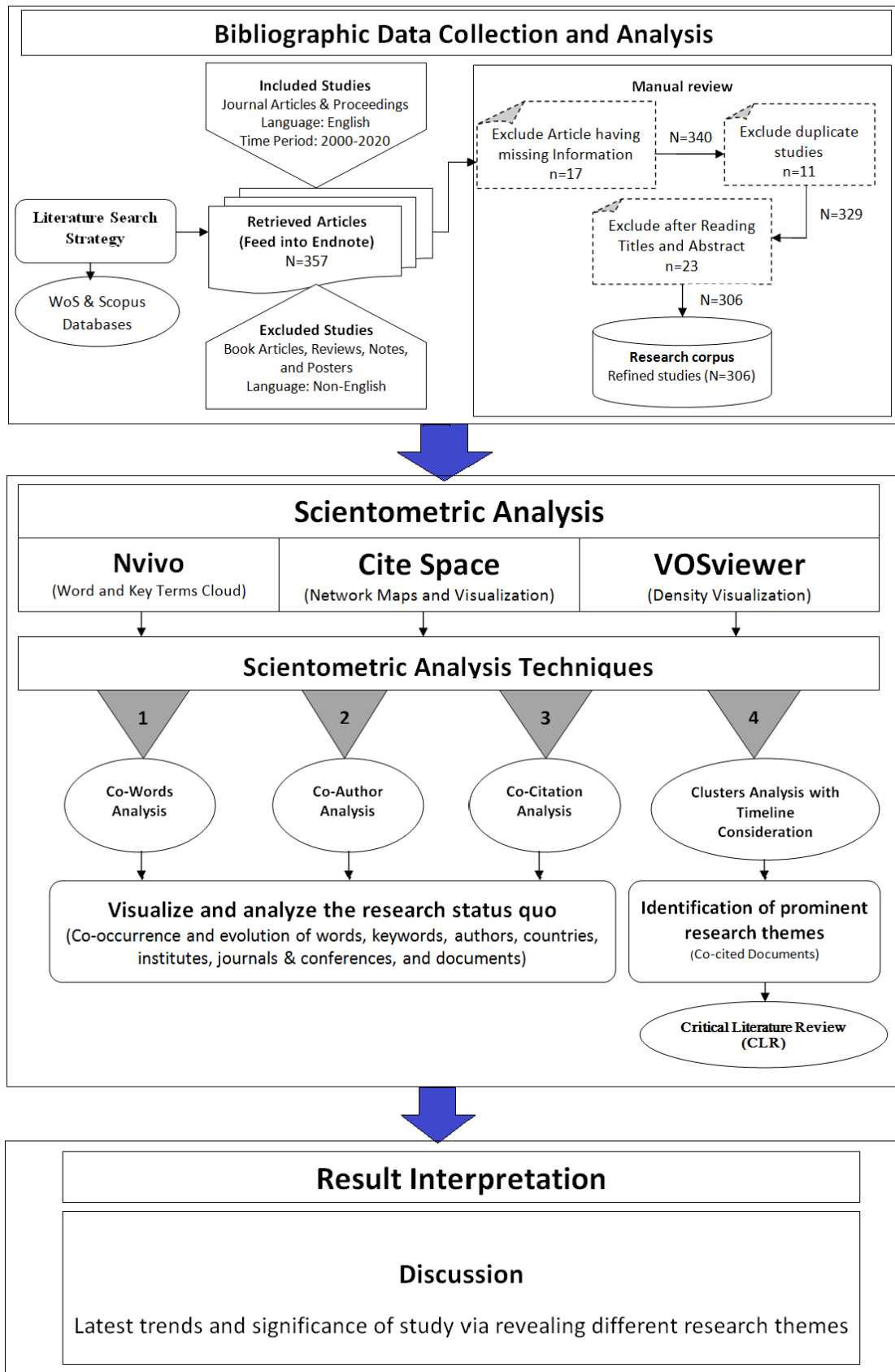


FIGURE 1. Research methodology.

the procedures of data searching, screening, and cleaning processes.

A. LITERATURE SEARCH STRATEGY

The data collection mechanism used in this paper is decisive and ranges from data sources selection to search methodology. Concerning data sources, WoS, Scopus, and Google Scholar are considered major databases for scientific publication. The authors in [29] performed a comparative analysis of these databases, highlighting their strengths and weaknesses. The purpose of the analysis was to help researchers in selecting relevant and significant databases. Several other databases containing core articles also exist. These databases are associated with different journal publication houses, such as IEEE Explore, Science Direct, Wiley Online Library, Elsevier, EBSCO, ASCE Library, ProQuest, Springer, Emerald, and Taylor & Francis. An in- depth investigation of the literature identified WoS and Scopus as the two key databases for significant and multidisciplinary studies [29], [34], [41], [49]. The present study included articles (published as journal articles, proceedings, and reviews) from WoS and Scopus to obtain comprehensive and high-quality data. Meanwhile, book articles, reviews, notes, and posters were excluded.

Grounded on the strong correlation between STC and software development, this study used search strings containing “Socio-technical congruence,” “Socio-technical dependency,” “Socio-technical coordination,” and “software development.” The initial search results (obtained via these key terms) presented various irrelevant papers that belong to other domains such as social sciences, psychology, and artificial intelligence, instead of software engineering. To retrieve the relevant studies, contextual terms were added in the search query, such as software development. Initially, several search term combinations were used to retrieve the relevant papers, and the results are compared with a preliminary collection of search outcomes. In the end, the search query was formulated by adding the key selected terms combined with Boolean “OR” and “AND” operators. To validate the search accuracy and cover all possible literature, search terms were adjusted in the search string.

Table 1 lists the search queries (according to the search formats of Scopus and WoS) used to seek and collect the relevant publications. Related papers were searched via the defined terms in different fields of articles (as given in the search query), such as the title, abstract, or keywords, to ensure that comprehensive data were obtained. However, utilizing the search criteria, such as “TITLE-ABS-KEY” (in the case of Scopus) and “Topic” (in the case of WoS), might have generated data that lacks a relationship to the intended research. Therefore, certain restrictions were applied to the search mechanisms to enhance the search accuracy and obtain high-quality, relevant results. For instance, the documents included were limited to English. Additionally, the publication sources were limited to journal articles, articles in press, conferences, and proceedings. To retrieve comprehensive information, the years 2000 to 2020 were chosen to select

TABLE 1. Search query strings.

No.	Source Database	Search Query
1	WoS	TOPIC: ((socio-technical OR socio technical OR social and technical)) AND TOPIC: ((dependenc* or congruen* or coordinati* or collaborativ*)) AND TOPIC: (*software development*)
2	Scopus	((TITLE-ABS-KEY(("socio-technical" OR "socio technical" OR "social and technical")) AND TITLE-ABS-KEY((dependenc* or congruen* or coordinati* or collaborativ*))AND TITLE-ABS-KEY(*software development*))

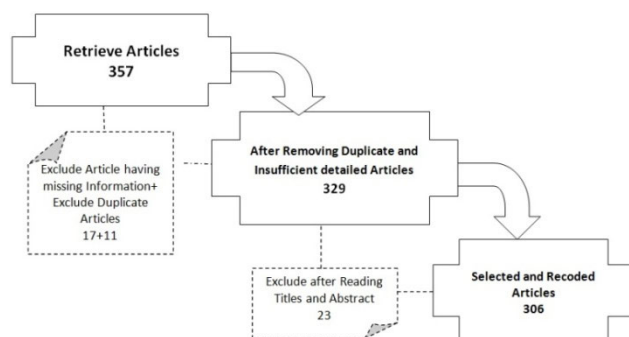


FIGURE 2. Data screening and cleaning process.

the data since the first STC framework was implemented in 2006 by Cataldo *et al.* [9].

B. DATA SCREENING AND CLEANING PROCESS

A complex screening method was needed to ensure that the final research corpus contained high-quality articles highlighting the trends and importance of STC in software development. After the searching mechanism, 357 articles were downloaded and fed into Endnote for further screening. A manual article screening mechanism was adopted to select the relevant publications from the downloaded records. First, articles with insufficient information (e.g., a missing author name or publication year) were excluded. Next, the duplicate publications were subsequently identified and removed from the research corpus. The articles’ relevance was determined by checking their titles, objectives, methods, and major findings to further refine the data. After the successful screening process (see Fig. 2), 306 articles were selected for further analysis.

Fig.3 presents an overall distribution of the bibliographic data (i.e., the 306 selected articles) collected from 2000 to 2020. 2012 and 2018 were the two peak years with the maximum publication numbers. The detailed analysis of the collected data revealed that the publications in 2012 focused on STC techniques and tools, analysis of various perspectives, and STC’s applicability in different phases of software development [15], [19], [50]–[57]. Conversely, most publications from 2018 focused on two main streams: i) the influence of STC in different fields [58], [59] and ii) the applications of STC in OSS development [60], [61]. The document

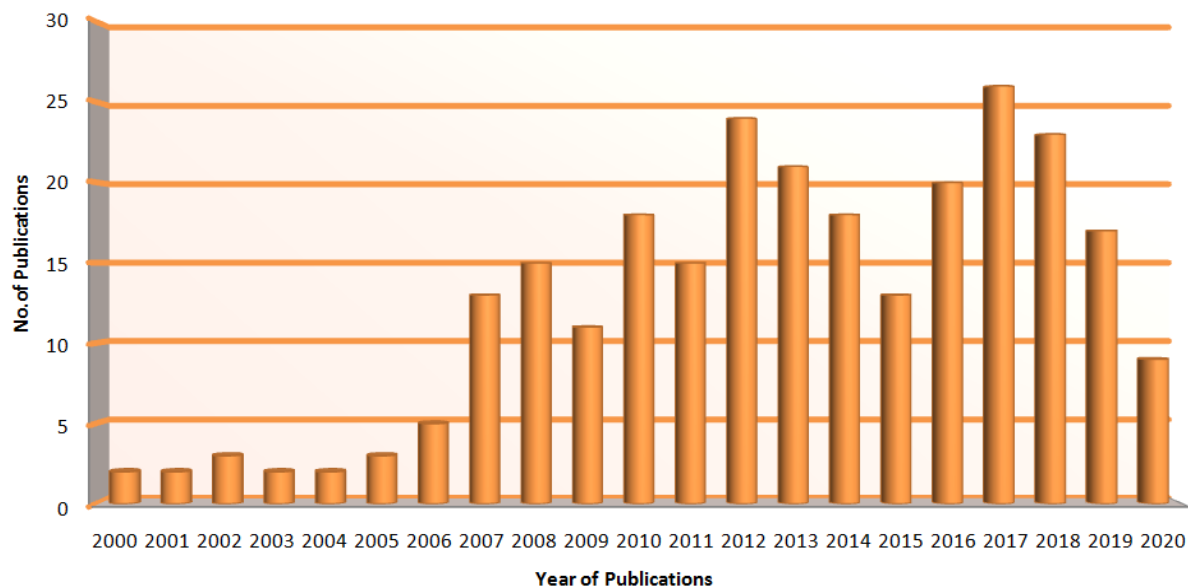


FIGURE 3. Number of articles published in WoS and scopus from 2000 to 2020.

citations from 2016 to 2020 indicate an increased interest of researchers and practitioners in STC.

IV. STAGE 2: SCIENTOMETRIC ANALYSIS

Manual reviewing is a satisfactory approach to present a thoughtful overview of the intended research area. However, it is susceptible to issues concerning biased author opinions and subjective interpretations [62]. A SLR also cannot sufficiently characterize the entire STC field due to its vast application range in different research domains [63].

A scientometric analysis helps find associations among concepts in the literature via numerous scientometric techniques. Therefore, it helps researchers identify new information in the literature that may be disregarded in traditional, manually conducted reviews [64]. Nalimov and Mul'chenko [65] introduced the concept of "scientometry" as the quantitative assessment of intended research to represent growth in the field of interest. The authors in [66] defined this concept as a technique to outline the knowledge corpus, improve the understandability of citation mechanisms, measure the research's influence, and highlight the evolutionary trends in a domain based on bibliographic data.

Additionally, Chen *et al.* [39] defined scientometry as analyzing published literature through different bibliometric techniques. This is done to outline the targeted domain's structure and evolution based on the high-quality scholarly collection. This study focuses on the scientometric methodology to present a holistic analysis of STC concerning software development activities. In the literature, different scientometric techniques are used to identify the prominent frontiers of research, such as co-word analysis [67], co-occurrence analysis [31], [68], co-author/collaborator analysis [69], co-citation analysis [70]–[72], bibliographic coupling [71], [72], and cluster analysis [73].

For STC scientometric analysis, four scientometric techniques were applied via different tools (discussed in subsection IV-B). This was undertaken to visualize the entire STC field and identify research patterns and global perspective trends. Table 2 provides a summary of the proposed scientometric analysis (i.e., the techniques and tools along with their outcomes).

A. SCIENTOMETRIC ANALYSIS TOOLS

This study selected three scientometric tools (CiteSpace, VOSviewer, and Nvivo) to perform the scientometric analysis of STC. An overview of each of these tools is provided in this section. The CiteSpace tool analyzes and visualizes the intended research field by creating numerous co-citation networks and graphs based on the scientific literature [39]. The co-citation networks and graphs help researchers understand existing studies and determine concealed hints within the data collection.

Additionally, CiteSpace computes two key metrics (collectively known as composite sigma) from the generated networks: betweenness centrality and citation bursts. These metrics determine the overall structural properties of a network that help identify STC's main points and evolutionary. The betweenness centrality metric is computed by using a ratio: the shortest path among the two nodes and the sum of all alike shortest paths (as shown in equation 1) [74]. The betweenness centrality depicts the structural holes that define the flow of information in the generated networks.

$$BC = \sum_{x \neq y \neq z \in V} \frac{\sigma_{xz}(y)}{\sigma_{xz}} \quad (1)$$

The citation burst metric is calculated using Kleinberg's algorithm [75]. This measure reveals the abrupt frequency changes in the citations over the short time interval within

the overall specified period. The strongest value of a burst denotes the topmost attractive work.

Additionally, CiteSpace also assisted in performing a cluster analysis to composite sigma and show the interconnection of nodes. The tool ensured that there were no overlaps in the clustering algorithm that would constrain a node to appear in only one cluster at a time. In this study, we focused on cluster analysis to perform a critical analysis of the published literature based on two clustering measures: 1) network modularity (denoted as Q) and 2) silhouette value (denoted as S). The Q value defines the extent of network decomposition, which determines the cluster's overall structure in a citation network. The value range for Q lies between 0 to 1. Any value above 0.3 indicates a well-structured knowledge network [76]. The S value provides an approach to measure the quality of cluster configuration. In other words, it indicates the estimation of uncertainty that may appear in the cluster's nature. The S value ranges from -1 to 1. Generally, any value above 0.5 depicts a reasonable view of the cluster, whereas a value of 1 represents a perfect separation of clusters [77].

VOSviewer, the second tool selected, was used to create a density view of items to visualize the main focus of the research patterns (i.e., keywords, authors, and documents). In density visualization, the items are represented by labels in the same manner as network visualization by CiteSpace. The densities of the items are identified via three colors: blue, green, and yellow.

The third tool selected, Nvivo, provides a cloud view of co-occurring keywords with respect to their frequency. This cloud view facilitates the identification of current trends in the intended field of research. Additionally, Nvivo helps in the deeper evaluation and exploration of particular themes provided by the CiteSpace tool.

B. SCIENTOMETRIC ANALYSIS TECHNIQUES

A scientometric analysis was performed on the data (i.e., gathered through the bibliographic analysis) via four scientometric techniques. This first technique used was a co-words analysis which comprised of three networks: co-occurring words, co-occurring terms or keywords, and keyword evolution. The second technique used was co-authors analysis, which focused on the collaboration network at three levels: author, institution, and country. The third technique used was co-citations analysis that determined the co-cited journals and co-cited articles. Finally, the fourth technique used was the analysis of co-cited document clusters considering the citation timeline.

1) CO-WORD ANALYSIS

Co-word analysis is a mechanism that counts and analyzes the number of keywords in an article related to a research field [78]. It explores the relationships among keywords in a research area. Furthermore, co-word analysis facilitates observing research trends and advancing the research topic [79].



FIGURE 4. Cloud of most frequent words.

A significant amount of information about the given research can be extracted from analyzing a structured text corpus (i.e., data collected from Scopus and WoS). Traditional text analysis locates documents that contain specific words or phrases. However, this type of analysis is tedious and time-consuming for researchers to spot new evolutionary events and trends in specified areas [80]. This study analyzed word frequency (via NVivo software) with an additional focus on the networks of co-occurring phrases and co-occurring keywords, grounded on the importance of text in documents.

C. WORD FREQUENCY ANALYSIS

NVivo software was used to analyze word frequency by applying the function “word frequency query” to all the retrieved publications in PDF format. The words were selected by setting the criteria to a minimum length of four letters and displaying the top 1,000 frequent words. The criterion of minimum four-letter long words was selected to prevent pronouns and adverbs in the retrieved data. Table 3 shows the top 20 most frequent words that appear in STC-related articles. A cloud view of the top 1,300 words in different font sizes based on the frequency of their occurrence is further shown in Fig.4.

D. NETWORK OF CO-OCCURRING PHRASE

A phrase is a group of words representing a unit of conceptualization. The analysis of the co-occurrence of phrases helps researchers explore the right direction of an investigation. In the current study, this analysis was conducted by generating a network of co-occurring phrases and considering the bursts detected from noun phrases and plain text. Fig.5 depicts a network of co-occurring phrases (developed through CiteSpace), consisting of 45 nodes and 29 links between these nodes.

The top seven co-occurring phrases with the highest frequency are listed in Table 4. These phrases include

TABLE 2. Summary of scientometric techniques with outcomes.

Sr#	Scientometric technique	Scientometric tool	Unit of analysis	Description	Outcomes	Future Directions
1	Co-word analysis	Citespace VOSviewer NVivo	<ul style="list-style-type: none"> • Words • Phrases • Keywords 	Identify words, phrases, and keywords that when appeared in the same titles, abstracts, or keywords list	Top 3 Frequent Words (%): Software (2.72), Technical (1.52), Systems(1.26) Top 3 key phrases (frequency): Software-engineering (15), Software-systems(11), “software-projects” (10) Top 3 Keywords (frequency): “software engineering” (48), “software design” (45), “Socio-technical” (42)	On the basis of content analysis the co-word analysis suggests more work for researchers in the domain of software engineering targeting socio-technical systems their technical information and software design.
2	Co-author analysis	Citespace VOSviewer	<ul style="list-style-type: none"> • Author • Country • Institutes 	Identify authors when they appeared as a co-author in the study	Top 3 Authors (Citations): Bird, C (116), Trainer, E (90), Valletto, G (45) Top 3 Countries (frequency): United States (98), Canada (27), Italy(21) Top 3 Institutions (frequency): Carnegie Mellon University, USA (13); the University of California, USA (5); Microsoft Research Cambridge, UK (3)	The identified key researchers and collaboration network related to the STC suggested focusing on recognized collaboration for comprehensive understanding and further research.
3	Co-citation analysis	Citespace VOSviewer	<ul style="list-style-type: none"> • Journal & Conference • Author • Document 	Identify journals and conferences, authors and documents on the basis of mutual appearances in the references list.	Top 3 Journals (IF): Communication ACM (0.715), IEEE Transactions On Software Engineering (3.331), IEEE Software (2.945) Top 3 Conferences (IF): Proceedings - International Conference on Software Engineering (0.44), Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW (0.31), Proceedings - Euromicro Conference on Real-Time Systems (0.59) Top 3 Authors (citation): Cataldo M (80), Herbsleb JD (51), Parnas DL (34) Top 3 Documents (citation): Cataldo et al. [9] (27), Brooks [96](19), Cataldo et al.[10](14) Clusters labels (members) Community structure(44), Socio-technical congruence(31), Current research(20)	Research can focus and published their research work in the recognized journals and conferences by enhancing the work of identified prominent authors and studies.
4	Cluster analysis	Citespace NVivo	Documents	Group out the documents on the basis of co-cited references.		Researchers can target structure of community to perform improvement in STC research.

“software-engineering” (frequency = 15), “software-systems” (frequency = 11), “software-projects” (frequency = 10), “socio-technical-congruence” (frequency = 9), “technical-congruence” (frequency = 9), “software components” (frequency = 6), and “design-principles” (frequency = 5). The frequency values of these phrases indicate their contribution to the advancement of STC research.

1) STRUCTURAL CHARACTERISTICS OF CO-OCCURRING PHRASES

In the context of structural network properties (i.e., betweenness centrality, citation burst, network modularity (Q), and silhouette (S) value, as discussed in section IV-A), the

co-occurring phrases network indicated the value of $Q = 0.81$. This showed the rational distribution of the phrases’ network in loosely coupled clusters. On the other hand, the value of $S = 0.33$ indicated high heterogeneity in the clustering mechanism.

Concerning betweenness centrality, three phrases signified high values of centrality: “software-projects” (centrality = 0.03), “socio-technical-congruence” (centrality = 0.02), and “different-role” (centrality = 0.02). The burst detection algorithm identified the six most cited term bursts: “software-systems” (burst = 7.02, 2015–2017), “software-engineering” (burst = 5.50, 2013–2014), “software-components” (burst = 3.56, 2007–2008), “software-projects”

TABLE 3. Top 20 most frequent words.

Word	Count	Weighted (%)
Software	1206	2.72
Technical	671	1.52
Systems	557	1.26
Socio	514	1.16
Development	493	1.11
Social	456	1.03
Design	379	0.86
Engineering	311	0.70
Coordination	291	0.66
Research	282	0.64
Information	263	0.59
Open	262	0.59
Collaborative	252	0.57
Source	234	0.53
System	234	0.53
Management	221	0.50
Project	216	0.49
Based	214	0.48
Data	213	0.48
Study	203	0.46
Analysis	197	0.44

TABLE 5. Top 10 most frequent co-occurring keywords.

Keyword	Links	Link Strength	Freq
Software engineering	132	333	48
Software design	140	370	45
Sociotechnical	116	296	42
Coordination	90	212	36
Open source software	99	239	29
Computer software	107	194	26
Collaboration	86	155	24
Sociotechnical system	100	171	22
Open system	76	167	20

(burst = 3.30, 2013–2014) in phrases network. This outcome revealed that researchers have significantly focused on these phrases in the STC field over the topical years.

E. NETWORK OF KEYWORD CO-OCCURRENCE

Keywords are considered the most significant and descriptive words used to understand the basic concepts and key findings of research publications. A network of co-occurring keywords helps identify the hot topics and topical research trends over a particular time. It also describes the advancements in the intended research area over a particular period [34]. In this study, the publications obtained from WoS contained two types of keywords: i) keywords provided by the authors and ii) keywords based on the journal’s research classification. However, the data retrieved from Scopus merely contained author keywords. of co-occurring keywords (Fig. 6a) was generated through CiteSpace, which utilized keywords and merged similar ones. The network contained 297 nodes with 1,892 links. The node size within the network represented the frequency of keywords in the bibliographic data. Furthermore, the keywords with a considerable role in the STC domain were portrayed via density visualizations (Fig. 6b) of the co-occurring keyword network. This visualization generated through VOSviewer helped highlight the key areas of the STC research domain.

In the network of co-occurring keywords, the top 10 most frequently appearing keywords are given in Table 5. They are “software engineering” (frequency = 48), “software design” (frequency = 45), “Socio-technical” (frequency = 42), “coordination” (frequency = 36), “open-source software” (frequency = 29), “computer software” (frequency = 29), “collaboration” (frequency = 24), “Socio-technical system” (frequency = 22), “open system” (frequency = 20), and “Socio-technical system” (frequency = 18). This result revealed that “software engineering” has the highest count in the bibliographic data. Therefore, it is a hot research topic and area connected to the STC field.

An insightful analysis of the publications gathered also evidenced the role of STC in software engineering. For instance, Kwan et al. [5] defined STC as a technique used to measure the coordination among software teams. Paasivaara et al. [81] also defined the concept

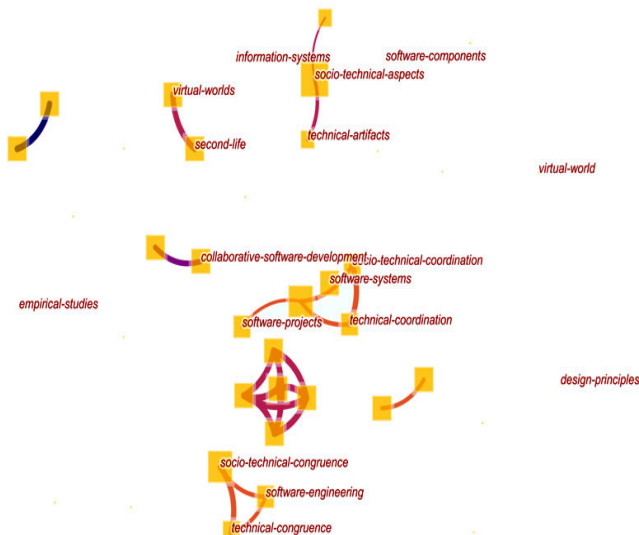


FIGURE 5. Network of co-occurring phrase.

TABLE 4. Top seven co-occurring phrases.

Phrases	Frequency	Burst
Software-engineering	15	5.5
Software-systems	11	7.02
Software-projects	10	3.34
Socio-technical-congruence	9	3.3
Technical-congruence	9	3.3
Software-components	6	3.56
Design-principles	5	3.5

(burst = 3.34, 2012–2013), “socio-technical-congruence” (burst = 3.30, 2013–2014), and “technical-congruence”

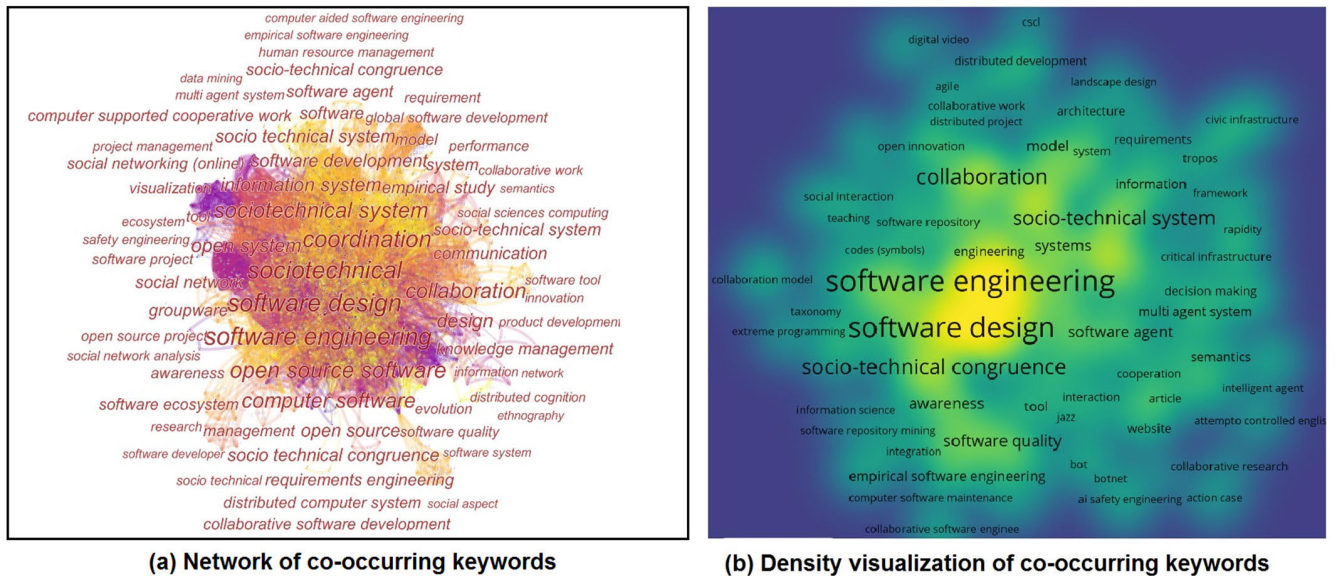


FIGURE 6. Visualization maps of co-occurring keyword.

concerning a global software engineering project and showed that STC could improve industrial practices. Similarly, Marczak *et al.* [82] discussed the significance of STC in requirement engineering. Their findings demonstrated the efficient identification of coordination needs through STC.

The network of keywords is a static illustration source of a particular area; however, it fails to depict changes in the STC field concerning time. Therefore, the time factor was added to the network of co-occurring keywords to describe the progress of STC-based keywords persistent in the intended period (2000–2020). Moreover, co-occurring links helped recognize the periods of the keywords. The lines were colored according to the connection establishment time. Fig. 7 reveals that “global software development” and “collaborative software development” co-occurred with “Socio-technical system” and “Socio-technical congruence,” respectively, for each period from 2000 to 2020. However, notably, Fig. 7 reveals that the area of “collaborative software development” has gained more attention and a longer period of attention (2002–2020). Next, “global software development” (GSD) has been a hot area of focus for researchers (from 2001 to 2018). The analysis of literature, such as [9], [13], [16], [19], [50], [51], [54], [61], [82]–[87], also supports this finding.

1) STRUCTURAL CHARACTERISTICS OF CO-OCCURRING KEYWORD NETWORK

Concerning the co-occurring keyword network’s structural characteristics, the modularity of the overall network was measured as $Q = 0.88$ (higher than the threshold of 0.3). This indicated a network distribution of loosely coupled clusters. However, the mean silhouette value of $S = 0.95$ showed high heterogeneity in network clustering.

Concerning betweenness centrality, several keywords scored high. The “empirical study” keyword with centrality equaling 0.25 was at the top. In contrast, the least scored keyword was “global software development” with centrality equaling 0.13. These keywords’ centrality values indicated their considerable influence on the growth of the STC research and numerous associated research areas.

The citation burst detection algorithm identified seven keywords (see Fig. 8) with strong burst values in different periods. All the extracted keywords with high burst strength represented prominent areas and topics in STC research. The list depicted an extensive number of works dedicated to the identified research areas connected to STC. This finding also revealed the progress in the STC field overtime. For instance, STC was initially utilized for collaborative software development due to its advantages in tackling social and technical perspectives. Thereafter, other factors were subsequently added to STC when software models and technologies began growing worldwide. For instance, Tamburri *et al.* [88] added the community smell concept in STC concerning OSS development. The improved STC methodology helped in identifying coordination issues related to different communities. Moreover, Zhang *et al.* [60] presented the influence of STC through the element of continuous defect prediction. The study highlighted the positive effects of STC on the software project’s outcome. In contrast, a lack of STC depicted a negative influence concerning software failure.

a: CO-AUTHOR ANALYSIS

An author is considered a key carrier of knowledge in academic exchanges and communication [89]. Bibliographic records contain information about the authors and other potential aspects of research publications. In a scientometric

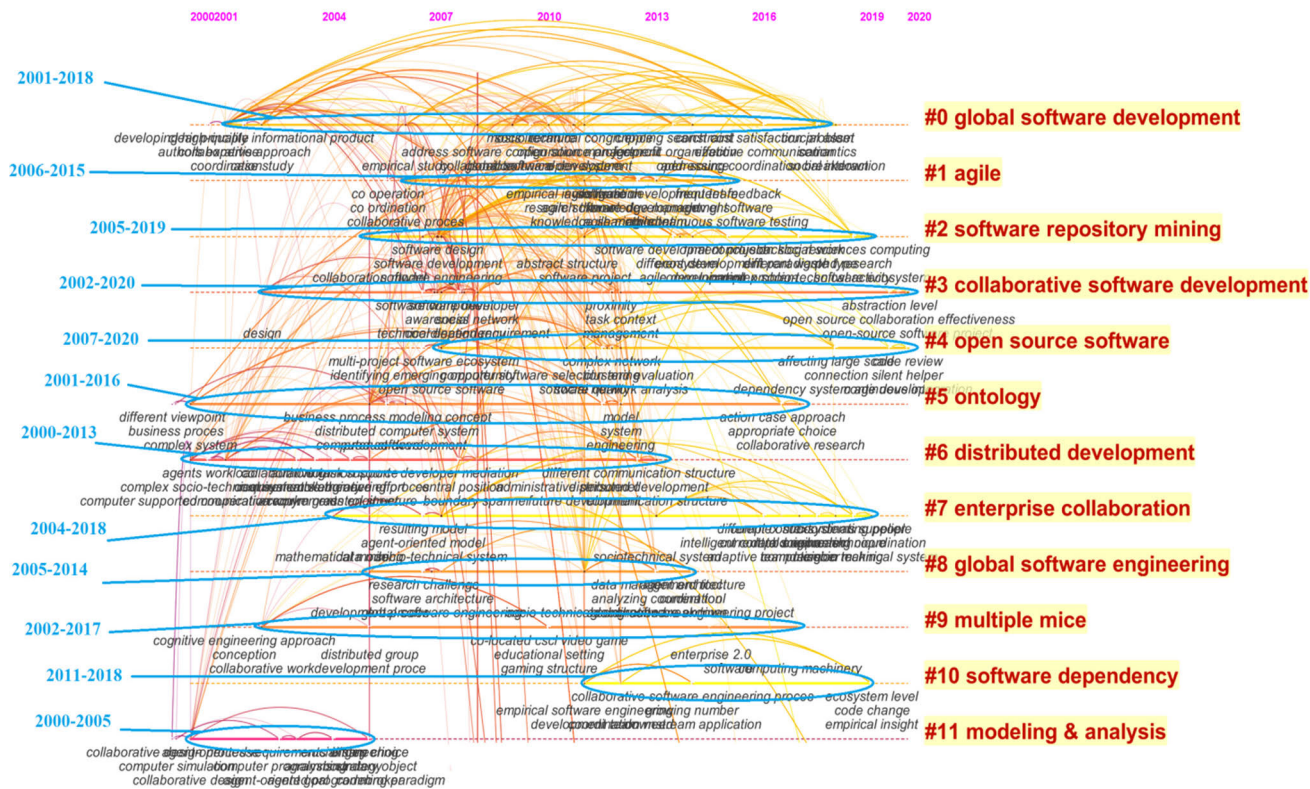


FIGURE 7. Timeline representation of co-occurring keywords.

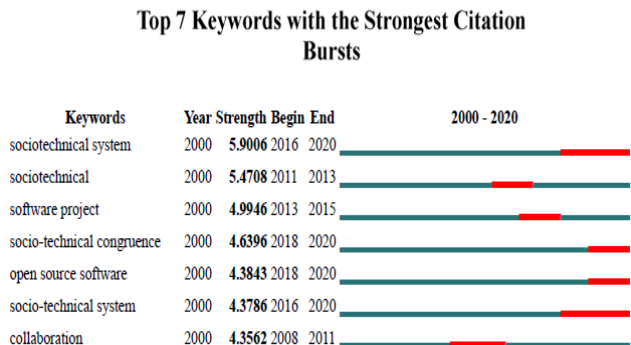


FIGURE 8. Burst strength of co-occurring keywords.

study, co-occurrence relations can be analyzed via a co-authorship network. Furthermore, the co-authorship network nodes also depict authors, countries, and institutes that share a study’s authorship [90]. Equation (2) is commonly used to compute the co-authorship among authors, countries, or institutes [91].

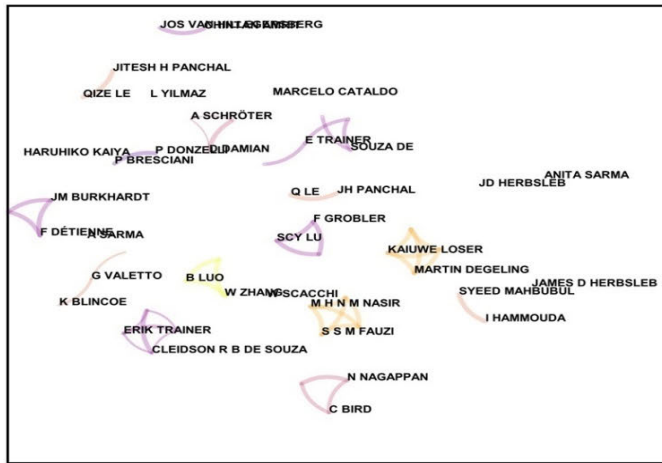
$$N_{pub} = (I_n)^r \tag{2}$$

where N_{pub} denotes the number of publications, I_n describes the number of items (authors, countries, or institutes) that share at least one publication, and r shows the co-authorship pattern, which is calculated by using the ratio of the natural log of N_{pub} and I_n .

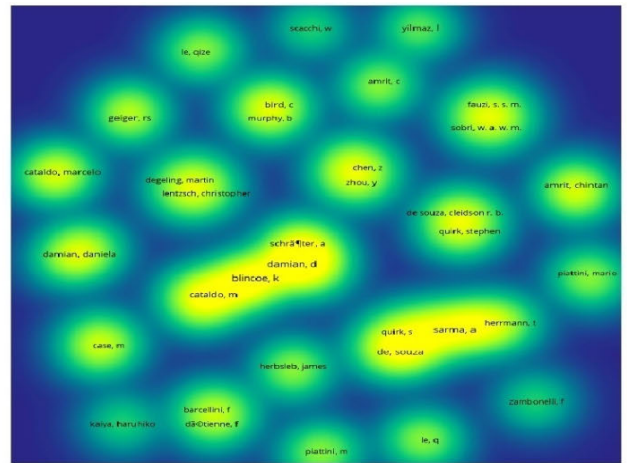
In this study, a co-author network was generated through CiteSpace using bibliographic records (i.e., it was gathered after processing and screening). CiteSpace can systematically outline the domains of knowledge via numerous innovative graphs [39]. Thus, it was also utilized for developing and analyzing other co-authors networks, such as the co-authorship network, co-occurring country/region network, and co-occurring institution network. For each co-author network, bursts were detected using Kleinberg’s algorithm [75]. Additionally, a density visualization of each network was generated via VOSviewer.

F. CO-AUTHORSHIP NETWORK

The co-authorship network facilitated in identifying the most significant researchers in the STC field and the volume of collaboration among different researchers. The co-authorship network generated via CiteSpace facilitated the illustration and analysis of the scientific information, logical connection, and well-structured knowledge about co-authors. Fig. 9a shows the co-authorship network representing the authors and their collaboration activities. Pathfinder, recommended in an existing study [92], was applied to remove the excessive links among different nodes and optimize network visualization. Overall, the network consisted of 23 clusters with 53 nodes and 49 links. In the co-authorship network, the node size symbolized the number of publications, whereas the link thickness indicated the strength of the collaborative relationships



(a) Co-authorship network



(b) Density visualization of co-authorship network

FIGURE 9. Visualization maps of co-authorshi.

among the nodes in each publication year. The links were displayed via a range of colors (i.e., orange, purple, brown, yellow, and red). Each color link denoted a different year from 2000 to 2020. Moreover, the color transitions (from cool to warm tones) represented the publication years.

Additionally, density visualization (see Fig. 9b) highlighted numerous research communities consisting of the number of authors with strong collaboration. Two prominent identified communities were: 1) Cataldo, Blincoe, Damian, and Schralter; and 2) Souza, Quirk, Sarma, and Herrmann. The central authors were also recognized within these communities based on higher numbers of collaboration activities than other authors. For instance, in the first community, Damian was identified as the central author among Cataldo, Blincoe, and Schralter. Similarly, Souza was recognized as the central author in the latter community among Quirk, Sarma, and Herrmann.

From the co-authorship network, the 10 most significant researchers based on the frequency of journal publications are listed in Table 6. Among these identified researchers, Valetto (IBM), Bird (Microsoft Research), Blincoe (Drexel University), and Kwan (University of Victoria) are at the top positions, in that order.

1) STRUCTURAL CHARACTERISTICS OF THE CO-AUTHORSHIP NETWORK

The co-authorship network yielded a value for modularity of $Q = 0.93$ (higher than the threshold of 0.3), indicating a rational division of the network via loosely tied clusters. The mean silhouette S score was 0.65, which signified the existence of heterogeneous clustering in the network. The third characteristic (betweenness centrality) indicated the influence of researchers and was computed based on the links among authors. The nodes with high betweenness centrality values were recognized as core hubs in the network that might serve as mediators for connecting different research groups. CiteSpace displayed the betweenness centrality

Top 6 Authors with the Strongest Citation Bursts

Authors	Year	Strength	Begin	End	2000 - 2020
A SARMA	2000	1.8159	2008	2011	
SCYLU	2000	1.7588	2000	2006	
E TRAINER	2000	1.728	2007	2008	
W SCACCHI	2000	1.7218	2005	2007	
A SCHRÖTER	2000	1.7156	2010	2011	
C BIRD	2000	1.5553	2009	2011	

FIGURE 10. Burst strength of co-author.

(i.e., higher than 0.1) as a purple ring. The entire network (see Fig. 9a) revealed zero betweenness centrality among the nodes. This finding suggested that future collaboration among different communities should be reinforced. Therefore, academic researchers should increase their communication to strengthen development in the STC field.

The author burst value (the fourth characteristic) represented an enhancement in the number of citations within a short period. In the co-authorship network, the top three bursts (as shown in Fig. 10) included (i) A Sarma with a burst strength of 1.8159 from 2008–2011, ii) Scylu with a burst strength of 1.7588 from 2000–2006, and iii) E Trainer with a burst strength of 1.728 from 2007–2008. The value of the burst strength denotes that a significant level of attention was gained in 2011. However, the burst detection algorithm indicates no burst within the preceding 9 years. In recent years, STC has substantially attracted verity worldwide. This is why a specific author has hardly obtained a high number of citations within a short span.

G. NETWORK OF COUNTRIES

The country-wide network of publications was generated through CiteSpace to investigate the structural distribution of STC publications. The network (Fig. 11a) consisted

TABLE 6. Top 10 significant researchers.

Name	Researcher Affiliation	Country	Citation	Count	Year
Valleto, G	IBM T.J. Watson Research Center.	New York	45	3	2012
Bird, C	University of California, Davis, United States.	United States	116	3	2009
Blincoe, K	College of Information, Science and Technology, Drexel University, Philadelphia, PA, United States.	United States	21	3	2012
Kwan, I	Software Engineering Global InterAction Lab. - SEGAL, University of Victoria, VIC, Canada.	Canada	6	3	2009
Souza, De	Fed Univ Para, BR-66075110 Belem, Para, Brazil.	Brazil	24	3	2007
Sarma, A	Carnegie Mellon Univ, Inst Software Res, Pittsburgh, PA 15213 USA	USA	41	3	2009
Damian, D	Software Engineering Global InterAction Lab. - SEGAL, University of Victoria, VIC, Canada.	Canada	6	3	2009
Trainer, E	Universidade Federal Do Par�, Bel�m, PA, Brazil.	Brazil	90	3	2007
Schr�-Ter, A	Univeristy of Victoria, Canada.	Canada	4	3	2010
Barcellini, F	INRIA, Eiffel Research Group, Domaine de Voluceau, Rocquencourt, 78153, Le Chesnay, France.	France.	45	2	2006

TABLE 7. Top contributing countries.

Country	Freq	Centrality	Year
United States	98	0.67	2006
Canada	27	0.18	2009
Italy	21	0.47	2003
United Kingdom	13	0.25	2001
France	12	0.02	2006

of 29 nodes connected via 45 links. The size of the node symbolized the number of publications in a country from 2000 to 2020. Moreover, the density visualization (as shown in Fig. 11b) of the network of countries displayed the most prominent countries with the highest numbers of STC publications.

Table 7 lists the countries that made significant contributions in the STC domain over the intended period. The United States is at the top with 98 publications, Canada with 27 publications, Italy with 21 articles, the United Kingdom with 13 publications, and France with 12 publications. A substantial number of publications in these countries denote the advancement in STC studies related to software development. Furthermore, concerning international collaboration, the United States appeared to participate in any collaborative activities with researchers from Canada, Italy, and the United Kingdom.

1) STRUCTURAL CHARACTERISTICS OF THE CO-OCCURRING COUNTRIES NETWORK

The co-occurring countries’ network (as depicted in Fig. 11a) depicted modularity at $Q = 0.69$. Therefore, a rational division is observed in the network of loosely tied clusters. The value of the mean silhouette (0.49) indicated heterogeneity in the network clustering.

The betweenness centrality denotes the contribution of a country in developing the STC field. The core countries most significantly exhibit high betweenness centrality values, displayed as purple trims in the network. The noted top contributors (in terms of centrality) are Italy (0.51), United Kingdom (0.46), Canada (0.44), United States (0.42), and Switzerland (0.25). On the other hand, the strength of citation bursts calculated by CiteSpace indicates abrupt changes in the citation frequency over a short period. The burst detection algorithm identified the following four countries (Fig.12) with strong burst strength: Brazil, Finland, Australia, and Canada. These countries generated a substantial number of publications along with high numbers of citations over the intended years. However, in the last 5 years (2016–2020), no citation burst was detected.

H. NETWORK OF INSTITUTIONS

A network of institutions was generated to explore their publication contribution in developing the STC domain. The generated network (via CiteSpace) consisted of 37 nodes with 11 links (as shown in Fig.13a). The analysis of the institution network and density visualization (Fig. 13b) indicated that STC research has grown considerably in numerous universities. The top four institutions are Carnegie Mellon University, USA (13 articles), the University of California, USA (5 articles), Microsoft Research Cambridge, UK (3 articles), and the Federal University of Para, Brazil (3 articles). A considerable number of publications from these institutions denote their important role in advancing STC research.

1) STRUCTURAL CHARACTERISTICS OF THE INSTITUTION NETWORK

Concerning structural network properties, the value of network modularity of $Q = 0.74$ was obtained, which is higher than the threshold of 0.3. Thus, a rational distribution was evident in the network of loosely coupled clusters. The mean

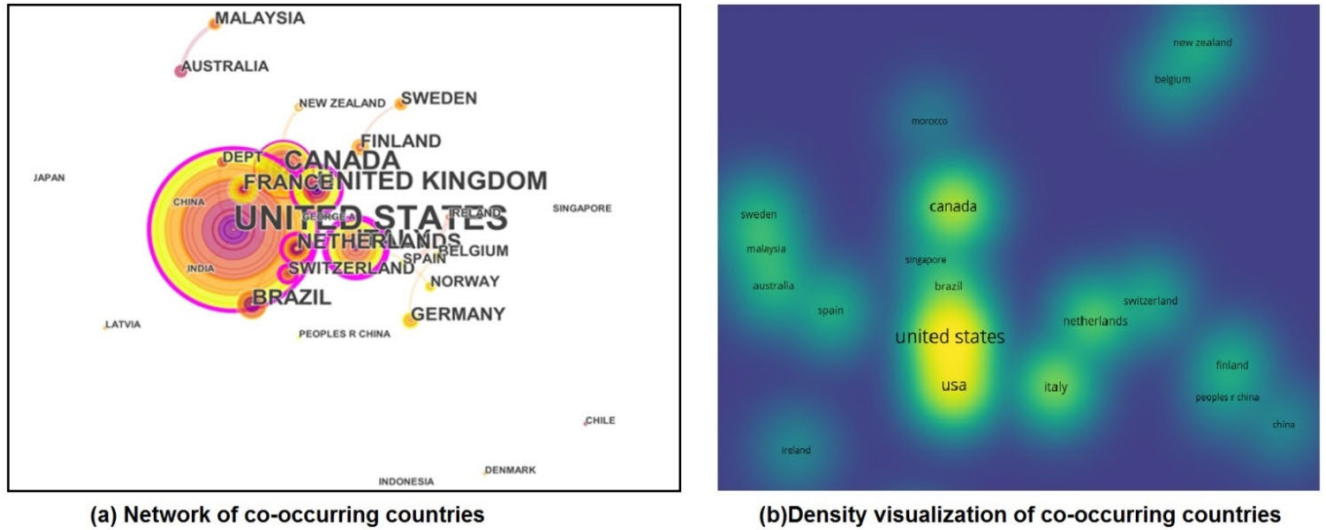


FIGURE 11. Visualization maps of co-occurring countries.

Top 4 Countries with the Strongest Citation Bursts

Countries	Year	Strength	Begin	End	2000 - 2020
BRAZIL	2000	3.4422	2005	2008	
FINLAND	2000	2.5967	2013	2015	
AUSTRALIA	2000	2.5831	2009	2010	
CANADA	2000	2.2229	2009	2011	

FIGURE 12. Burst strength of co-occurring countries.

silhouette value of $S = 0.19$ indicated substantial heterogeneity in the clustering mechanism.

Concerning betweenness centrality, Carnegie Mellon University (centrality = 0.02) and the University of California (centrality = 0.01) exhibited more connections and were deemed core collaborators when compared to other institutions. Concerning citation bursts (see Fig. 14), the University of California exhibited the highest burst strength of 2.5815 from 200–2009, the Federal University of Para achieved a burst strength of 1.7775 from 200–2008, and the Microsoft Research Center had a burst strength of 1.5375 from 200–2011. The institutional citation burst analysis indicates no burst in the last 9 years (2012–2020).

a: CO-CITATION ANALYSIS

The co-citation analysis is considered as a measure of semantic similarity for articles that utilize associations among the cited documents [40]. This analysis measures the connections and relationships among publication instances, such as authors, journals, or documents. The co-citation analysis determines how often two instances are cited by a third one [93]. The co-citation calculation can be better understood

through equation (3), as mentioned in [94].

$$N_{p_1 p_2} = \sum_{I=1}^n L_{I p_1} L_{I p_2} \tag{3}$$

where $N_{p_1 p_2}$ represents the number of items (journals, authors, or documents) that cite p_1 and p_2 in pairing, I denotes the item citing p_1 and p_2 , and L reflects the array of citations for p_1 and p_2 .

The present study investigates three networks: i) journal and conference co-citation, ii) author co-citation, and iii) document co-citation.

I. JOURNAL AND CONFERENCE CO-CITATION NETWORK

The network of journal and conference co-citation (generated through CiteSpace) consists of 405 nodes and 2,714 links (as shown in Fig. 15a). The most highly cited journals and conferences are represented by large nodes in the generated network that reveal their influence on STC research in software development. Additionally, Fig. 15b illustrates the density visualization of the cited journal and conference network, indicating the focus of researchers on journals and conferences on STC.

In software development, a total of 87 journals and conferences related to STC were identified from the research corpus (i.e., 306 articles). Among these, five journals and conferences published a considerable number of articles. Tables 8 and 9 list the top five journals and conferences concerning publication frequency, respectively. The Communications of the ACM journal published 43 articles (constituting 20.19%) related to STC research, ranking first on the list. This journal exhibits considerable authority in STC and popularity with the researchers. IEEE Software Engineering, Proceedings in Conference of Software, IEEE Software, and Datamation were the four additional journals and conferences substantially contributing to STC research publications.

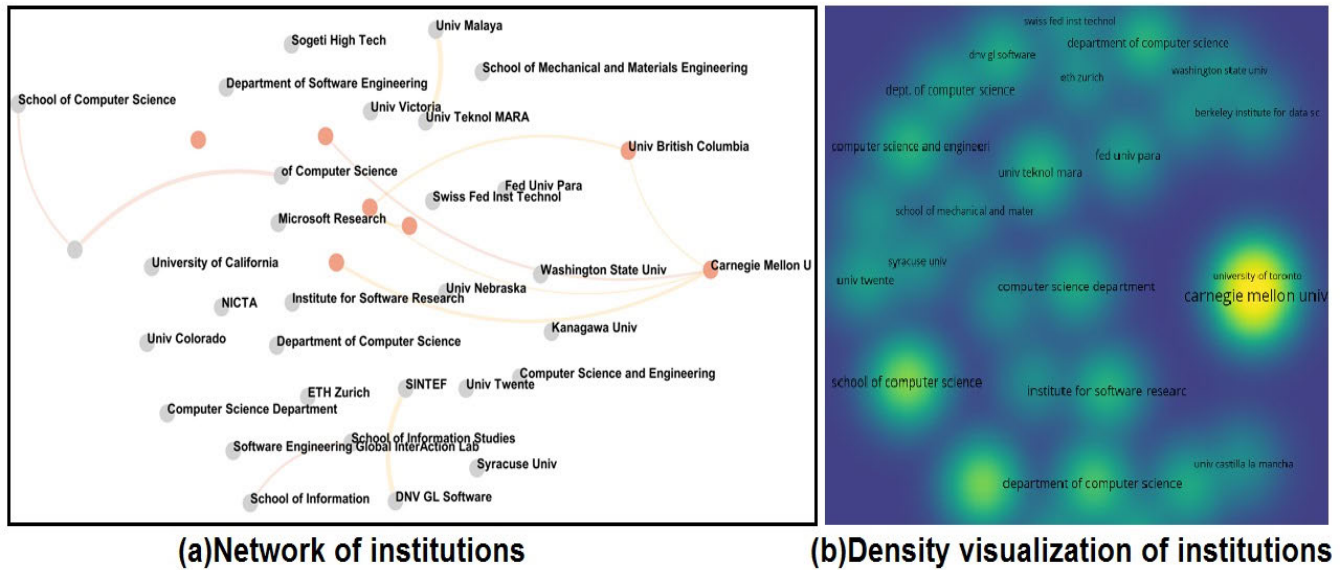


FIGURE 13. Visualization maps of institutions.

TABLE 8. Top five journals.

No.	Journal	Host Country	Publisher	Documents	IF
1	Communication ACM	United States	Association for Computing Machinery	43	0.715
2	IEEE Transactions On Software Engineering	United States	IEEE Computer Society	42	3.331
3	IEEE Software	United States	Institute of Electrical and Electronics Engineers	31	2.945
4	Datamation	United States	EarthWeb	21	0.150
5	Human-Computer Interaction	United States	Lawrence Erlbaum Associates Inc.	19	2.006

Top 3 Institutions with the Strongest Citation Bursts

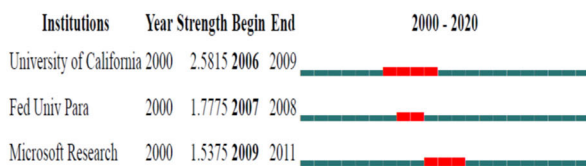


FIGURE 14. Burst strength of institutions.

1) STRUCTURAL CHARACTERISTICS OF JOURNAL AND CONFERENCE NETWORK

The network’s structural properties indicated that the overall network modularity was $Q = 0.72$, depicting the loose coupling of network clusters. The mean silhouette was $S = 0.25$, indicating the diversity in network clustering. The centrality measures in the co-citation network define the quantity of distinctive interconnected multidisciplinary journals [39].

Fig.15a presents high values of betweenness centrality for several journals and conferences, including the Computer Journal (0.21), IEEE Software (0.19), Information Management (0.17), Communications of the ACM (0.14), Proceedings in Conference of Software (0.14), ACM Computing Surveys (0.13), and IEEE Software Engineering (0.13). The fourth metric (the burst detection algorithm) detected three journal co-citation bursts (see Fig.16): ACM Computing Surveys (burst strength = 3.6605, 2012–2017), Empirical Software Engineering (burst strength = 4.3301, 2017–2020), and IEEE Software Engineering (burst strength = 4.3301, 2017–2020). The findings indicate the strong co-citations of these journals over the corresponding period.

J. AUTHOR CO-CITATION NETWORK

The relationships between distinct authors could be analyzed through the author co-citation network. This could be generated based on the authors whose works were displayed together in the same publication’s cited references [30].

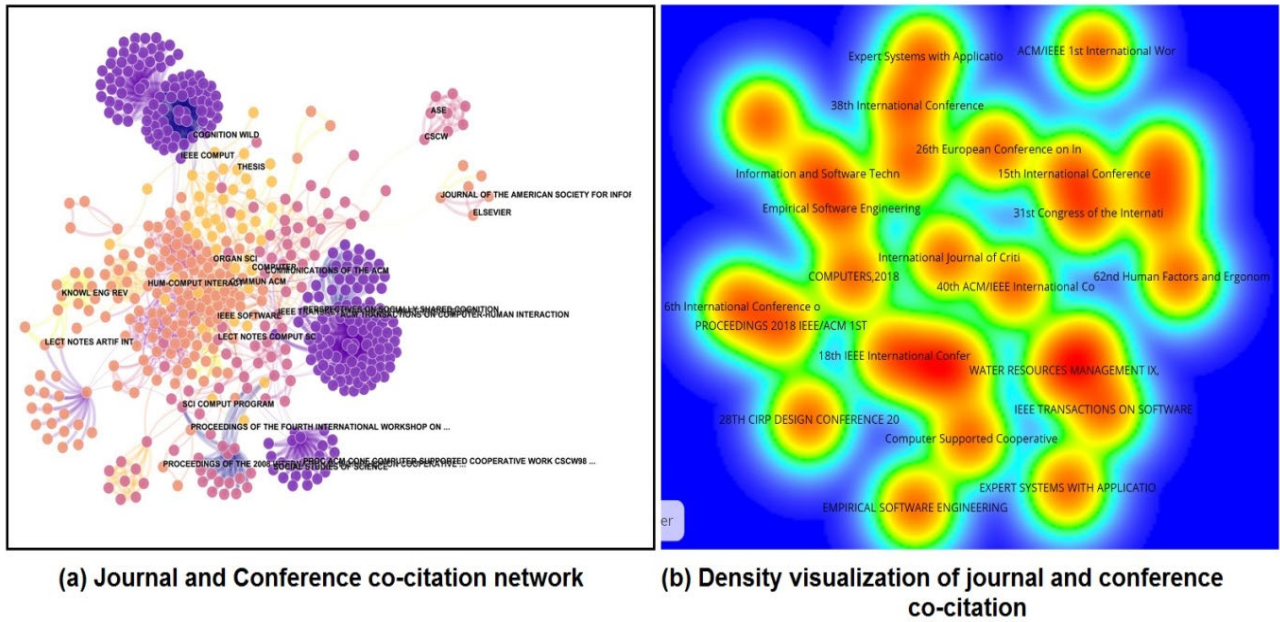


FIGURE 15. Visualization maps of co-cited journal and conference.

TABLE 9. Top five conference or proceedings.

No.	Conference or Proceedings	Host Country	Number of Publications	Impact Factor
1	Proceedings - International Conference on Software Engineering	United States	34	0.44
2	Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW	United States	8	0.31
3	Proceedings - Euromicro Conference on Real-Time Systems	United States	3	0.59
4	Proceedings - 2016 16th IEEE International Conference on Computer and Information Technology, CIT 2016, 2016 6th International Symposium on Cloud and Service Computing, IEEE SC2 2016 and 2016 International Symposium on Security and Privacy in Social Networks and Big Data, SocialSec 2016	United States	2	0.16
5	Software Engineering Education Conference, Proceedings	United States	2	0.14

Top 3 Cited Journals with the Strongest Citation Bursts



FIGURE 16. Burst strength of co-cited journal and conference.

This research developed a network of author co-citations via CiteSpace using the bibliographic data, where the network contained 449 nodes and 1732 links (see Fig. 17a). The node size indicated the frequency of the authors co-cited, whereas the links referred to the indirect supportive associations among authors based on the number of co-citations. Another

view of the highly co-cited authors was better represented through density visualization (as highlighted in Fig. 17b).

The results (co-citation network and density visualization) revealed that the most highly cited authors included Cataldo (frequency = 80, United States), Herbsleb (frequency = 51, United States), Parnas (frequency = 34, Canada), Kwan (frequency = 33, Canada), Bird (frequency = 33, United States), Conway (frequency = 33, United States), Sarma A (frequency = 31, United States), Ehrlich (frequency = 28, United States), Crowston (frequency = 26, United States), Mockus (frequency = 25, United States), and Valetto (frequency = 23, United States).

Table 10 lists the top 10 most significant cited researchers with their affiliated country and publication year. As depicted in Table 10, Cataldo exhibited a greater contribution towards developing STC models and their application in distributed software development, specifically from the

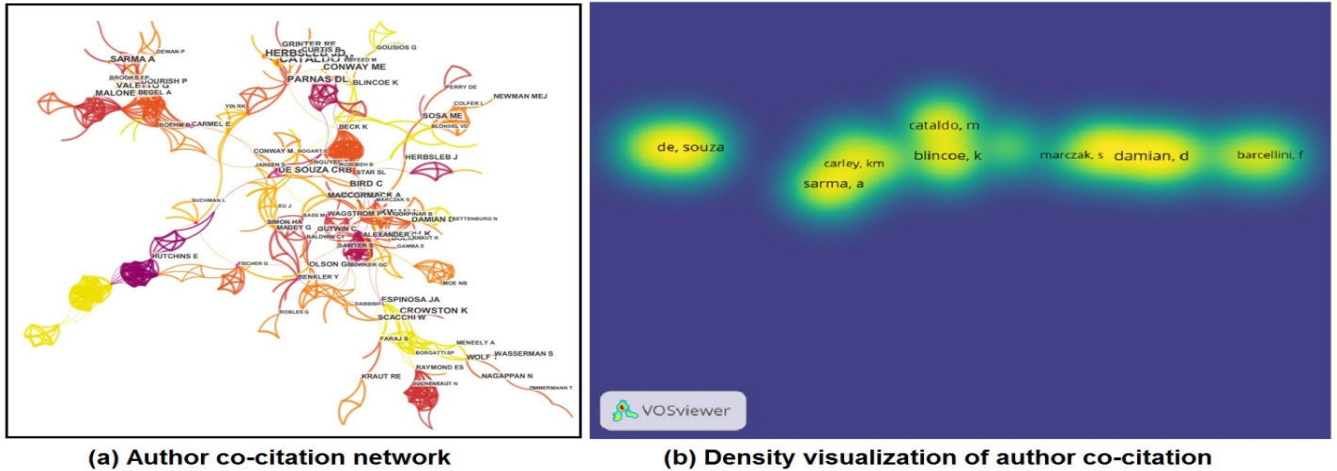


FIGURE 17. Visualization maps of author co-citation.

TABLE 10. Top 10 cited researchers.

Author	Country	Freq	Sigma	Year
Cataldo M	United States	80	1.16	2007
Herbsleb JD	United States	51	1	2006
Pamas DL	Canada	34	1.08	2007
Kwan I	Canada	33	1.16	2011
Bird C	United State	33	1.06	2009
Conway ME	United States	33	1	2007
Sarma A	Pennsylvania	31	1.21	2007
Ehrlich K	United State	28	1.03	2009
Crowston K	United States	26	1.06	2009
Mockus A	United State	25	1.29	2006

industrial perspective. The highly cited authors are from diverse locations, highlighting that STC research is conducted globally.

1) STRUCTURAL CHARACTERISTICS OF AUTHOR CO-CITATION NETWORK

The overall network modularity was measured as $Q = 0.84 (> 0.3)$, which showed loosely coupled clusters within the network. The mean silhouette ($S = 0.32$) showed heterogeneity in network clustering. Regarding the betweenness centrality metric, the top three authors were identified as Dourish (centrality = 0.29), Faraj (centrality = 0.24), and Espinosa (centrality = 0.23).

The fourth metric, the burst detection algorithm, determined various potential researchers by identifying hasty enhancements in citation counts over a diminutive period. Fig.8 provides the list of authors with the strongest citation bursts. Thus, these authors have a strong impact on the STC field research regarding software development in a particular period.

Top 6 Cited Authors with the Strongest Citation Bursts



FIGURE 18. Burst strength of co-cited authors.

K. DOCUMENT CO-CITATION NETWORK

The analysis of the cited references in the selected publications demonstrates the overall scientific knowledge of the respective publications [49]. The basic purpose of document co-citation analysis (DCA) is to study the network of co-cited references. A network of document co-citation was used to insightfully investigate the STC research domain. CiteSpace was used to draw a network of document co-citations based on the collected bibliography to highlight the relationships between citations at the author level.

Fig. 19a illustrates the constructed network and comprises of 211 nodes and 669 links. Each node is labeled by the author name and publication year, and the node size denotes the co-citation frequency of the documents. The aforementioned documents' nodes were obtained from the list of cited references in the selected publications (the corpus of 306 retrieved articles). Fig. 19b illustrates the density map view of the document co-citation, outlining co-cited references' strong relationships (generated using VOSviewer).

The document co-citation network revealed the top five most cited documents and their publication years and authors, which are listed in Table 11. The high citation values of these articles indicate their global popularity and significant contribution to STC research in software development.

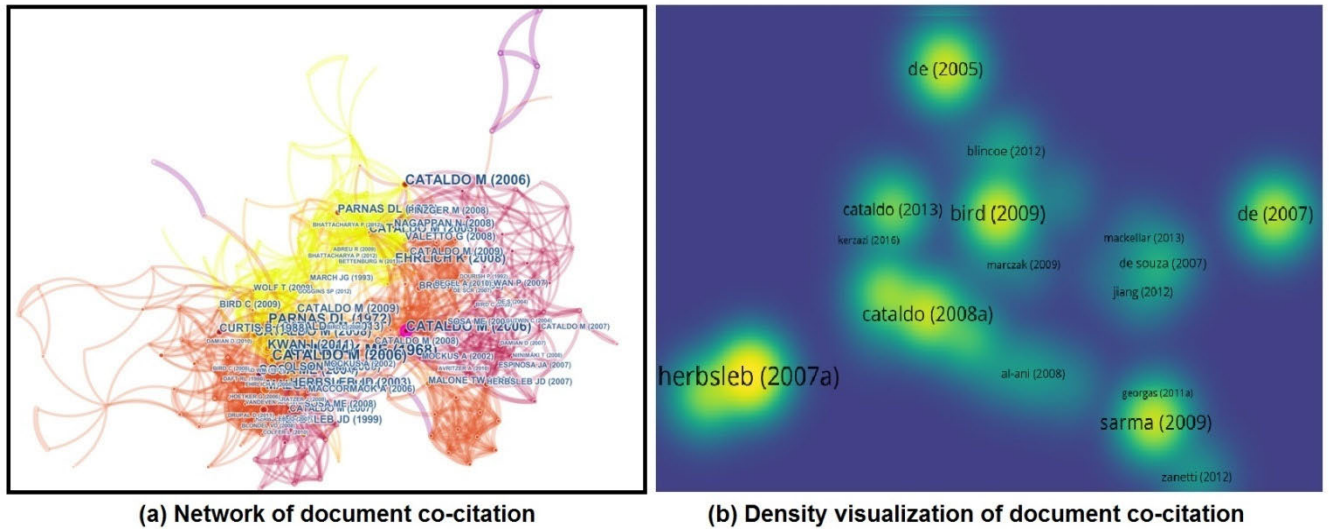


FIGURE 19. Visualization maps of document co-citation network.

TABLE 11. Top cited document.

No.	Number of Citation	Article	Year	Author
1	27	Identification of coordination requirements: Implications for the Design of collaboration and awareness tools	2006	Cataldo et al. [9]
2	19	The mythical man-month	1995	Brooks [95]
3	14	Socio-Technical Congruence: A Framework for Assessing the Impact of Technical and Work Dependencies on Software Development Productivity.	2008	Cataldo et al.[10]
4	11	Extending socio-technical congruence with awareness relationships	2011	Kwan et al.[11]
5	11	The Misalignment of Product Architecture and Organizational Structure in Complex Product Development	2004	Sosa et al.[96]

1) STRUCTURAL CHARACTERISTICS OF DOCUMENT CO-CITATION NETWORK

The document co-citation network analysis displayed the overall network modularity as $Q = 0.76$ and the mean silhouette as $S = 0.3019$, demonstrating the loose coupling and diversity in network clusters, respectively. Concerning betweenness centrality, certain documents exhibited high centrality values. For instance, Kwan et al. [11] had the highest centrality (0.24) among all the documents, whereas Cataldo et al. [9] and Sosa et al [96] achieved 0.01 and 0.09 centrality, respectively. The results denote that these articles serve as foundations for STC research and exhibit a significant effect on STC growth concerning software development. Concerning citation burst detection, the only document with a citation burst is the work of Cataldo and Herbsleb [97] (see Fig.20). This also indicated that Cataldo and Herbsleb’s work was cited frequently over a short time.

a: CO-CITED DOCUMENTS CLUSTERING WITH TIMELINE ANALYSIS

Cluster analysis is a data mining technique commonly used to discover hidden knowledge and semantic concepts from

Top 1 References with the Strongest Citation Bursts

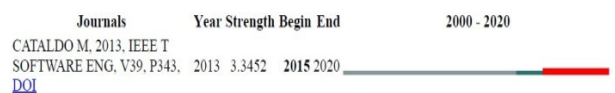


FIGURE 20. Burst strength of co-cited documents.

text [92]. This method can be used to classify the collections of research data into different classes according to the correlations among various terms. The present study applied cluster analysis to identify different themes, salient features, trends, and interdisciplinary relationships in the intended research area. Among various types of cluster labeling algorithms (Latent Semantic Indexing, Log-likelihood Ratio (LLR), and Mutual Information (MI)), LLR is most widely used in literature as it produces high-quality results in relation to convergence and uniqueness [98]. The current study utilized this algorithm to label the generated clusters anchored on the efficiency of LLR.

The documents were classified in clusters according to the index terms defined in the cited references of the 306 collected publications. This cluster classification was

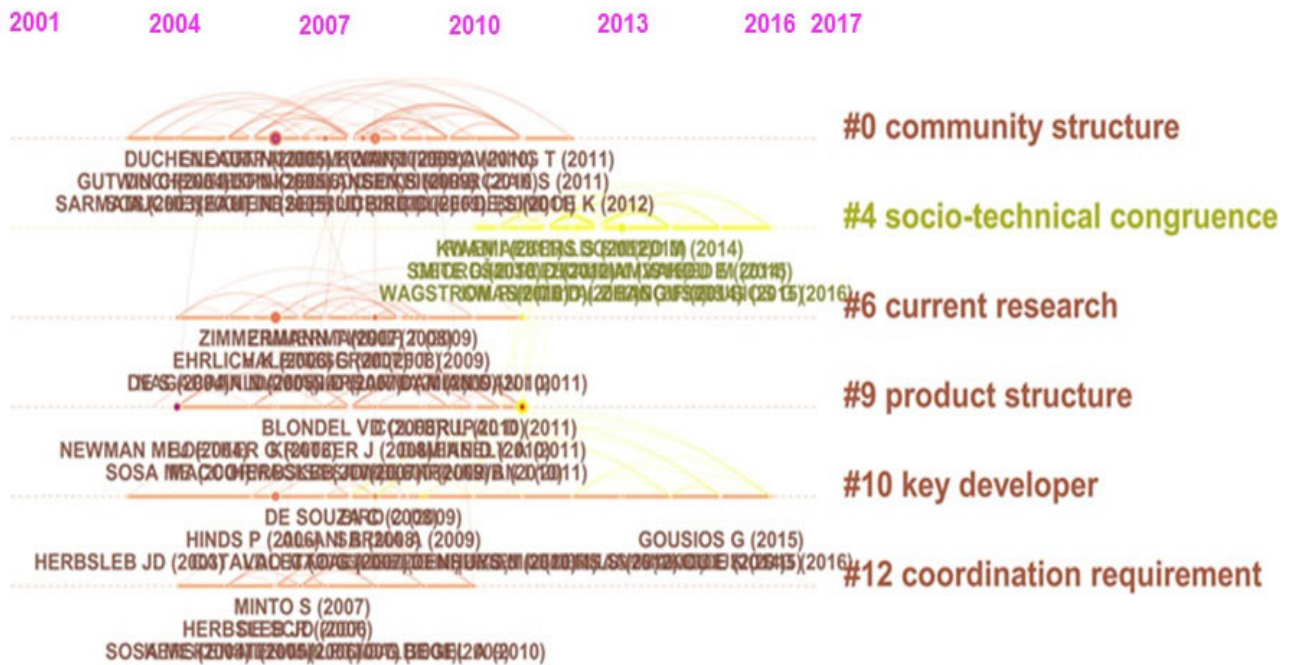


FIGURE 21. Timeline view of cited document cluster.

generated through the LLR algorithm to denote STC-related research’s semantic structures and themes. The generated document co-citation network contained six co-citation clusters (as shown in Fig.21). The two biggest clusters were cluster number 0 with 44 members (labeled as “community structure”) and cluster number 4 with 31 members (labeled as “Socio-technical congruence”). The cluster labeled as “coordination requirement” (cluster number 5) was the smallest cluster with 10 members.

A time representation of the clusters was generated (through CiteSpace) using the document co-citation network to analyze each cluster’s development concerning dynamic changes in the research area. Fig.21 shows the clusters and period of each cluster. Notably, the high coverage period is shown in cluster 10 (13 years from 2003–2016). In contrast, cluster 4 and cluster 12 indicate a shorter period (6 years from 2004–2010) of citation.

L. CRITICAL LITERATURE REVIEW (CLR) OF CO-CITED DOCUMENTS CLUSTERS

The prominent STC features were deduced based on the cluster’s results from the document co-citation cluster analysis. However, each node’s label represents its intellectual concepts based on the articles cited in the selected 306 documents, which may generate vague knowledge on the confronting challenges in the specified investigated area. For instance, the label of cluster 6 revealed that the majority of the works depict current research development.

However, the coverage period (2004–2011, as highlighted in Table 12) shows literature older than the last 5 years. Therefore, this study combined the cluster analysis (as discussed in section IV-4) of bibliographic data with a CLR to better understand STC research concepts and cover all relevant research themes. A CLR of the collected information reduces the ambiguity in research interpretation and helps identify the concerned research themes and challenges intended for future focus.

Bibliographic data were insightfully analyzed by focusing on the research themes (recognized from the co-cited document clusters) chronologically to perform the CLR. Furthermore, Nvivo was used to determine the distribution of the cited publications over time. The top-ranked terms were in the focused cluster, highlighting the latest research trends in the intended area. The in-depth detail of each cluster is described in the subsequent subsection.

1) CLUSTER 0: COMMUNITY STRUCTURE

The largest cluster was cluster number 0, with 44 members and a 0.80 silhouette value, signifying high reliability. The ranked terms (as shown in Fig. 22) indicated that the cluster members represent the STC aspects related to the community (team) and its organization, such as socio-technical relationships, development behavior, team communication design, and structure. The graph in Fig. 22 presents the entire coverage of cluster 0, including the distribution of the cited studies from 2003–2012.

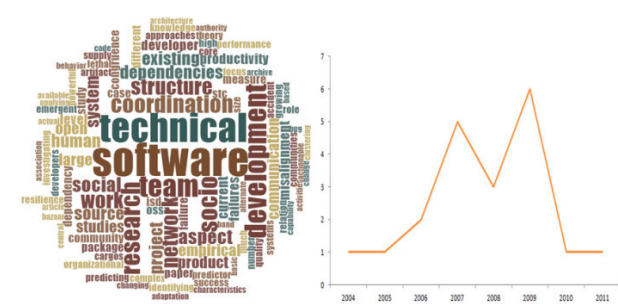


FIGURE 24. Cluster 6 with a cloud of top-ranked terms and coverage period with a number of studies.

explored that these challenges could be reduced by utilizing efficient STC measurements. Wang *et al.* [58] presented the concept of transgressive incongruence and found the negative effect of excessive congruence (i.e., unnecessary communication among developers) on software quality. Kilamo *et al.* [104] explored the role of developer communication in STC measurement and discussed the influence of different software development strategies on developers' communication patterns.

Numerous other studies (in cluster 4) proposed models and techniques for measuring STC. For instance, Zhang *et al.* [60] proposed a building-level STC model to measure the fit between coordination needs and actual coordination activities. The applicability of the proposed model was investigated through the continuous prediction of software defects. Zhang *et al.* [16] introduced a new STC measurement technique for the file level and investigated the relationship between STC and bug proneness in OSS projects. The study primarily contributed by providing a coordination breakdown required to reduce coordination issues in OSS development. Concerning information technology, Landegren *et al.* [59] presented an IT network as a Socio-technical system. The system's resilience was assessed using simulation-based methods, whereas the system managers were considered decision-makers. Similarly, Sobri *et al.* [105] presented a novel method to determine the relationship between STC and team performance using an incremental software development model.

3) CLUSTER 6: CURRENT RESEARCH

The third main research theme identified, "current research" (cluster 6), had 20 members and a silhouette value of 0.92. As depicted by the theme name and cloud of keywords (see Fig. 24), the publications included were related to STC development and the modern era (i.e., evolution in STC-related constraints, techniques, tools, and applications). Fig. 24 also indicates that the studies classified in this cluster are mainly published from 2004–2011.

This cluster's main representative article is Schroter [100], which highlighted STC's evolution related to industrial development. The study explored and listed the prominent human aspect that must be considered in STC measurement

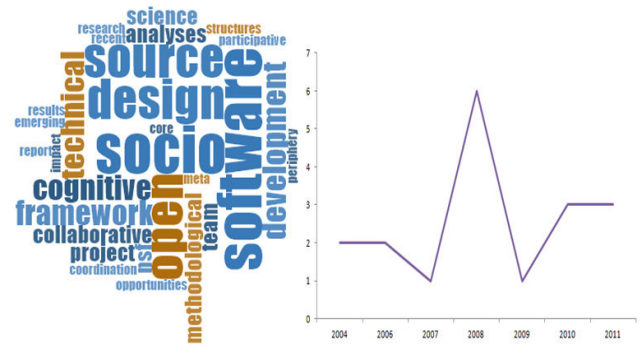


FIGURE 25. Cluster 9 with a cloud of top-ranked terms and coverage period with a number of studies.

concerning software development. Various studies in cluster 6 focused on using automated tools to facilitate STC measurement. For instance, Syeed *et al.* [87] proposed an STC-based automated tool that analyses and visualizes OSS development data through different data services. MacKellar [106] proposed a tool based on STC to suggest and support coordination activities during software engineering project development. Damian *et al.* [107] presented the role of distributed domain knowledge and multi-communication structure in STC. The authors identified that the accessibility to distributed domain knowledge affects coordination activities.

4) CLUSTER 9: PRODUCT STRUCTURE

Cluster number 9 (labeled as "product structure") consisted of 18 members and showed a silhouette value of 0.96, demonstrating its high reliability. Fig. 25 displays the term cloud and coverage graph of cluster 9. The term cloud indicates the main theme components (i.e., product structure), whereas the graph represents the number of studies cited from 2004–2011. The peak of the cited publications was in 2008.

The illustrative article for cluster 9 was by Rytsareva *et al.* [84]. This article investigated the relationship between organizational (product structure) and communication structure in OSS communities. The study revealed that a product structure reflects the community communication structure. Therefore, a project can be successful if these structures match. However, many studies in this cluster focused on a communication structure that enhanced product quality. Le and H [103] explored the influence of product and communication structures on STC through different dependency modeling techniques. Dependency modeling was utilized to illustrate the structure and evolution of products. The results revealed that a good product and communication structure implies significant product quality. Similarly, Zanetti [50] highlighted the importance of STC by referring to a need for a unified product framework that considers various factors, such as technical dependencies, human factors, and social aspects. Additionally, Amrit and Hillegersberg [85] showed the influence of a social network consisting of core peripheral developers on the structure of OSS projects. The study revealed that a communication tool helps identify coordination problems that affect OSS projects.

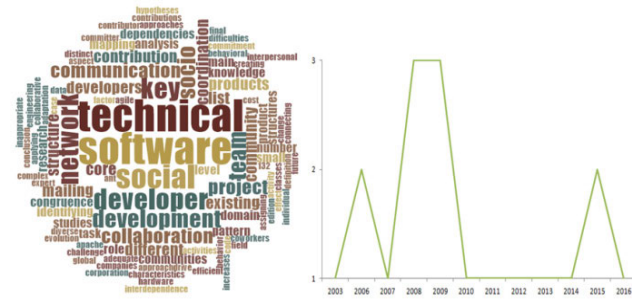


FIGURE 26. Cluster 10 with a cloud of top-ranked terms and coverage period with a number of studies.

5) CLUSTER 10: KEY DEVELOPER

The research theme “key developer” (i.e., cluster 10) consisted of 18 members and achieved a silhouette value of 0.025, showing the cluster’s reliability. Fig.26 presents the term cloud and shows the ingredients of the cluster’s research theme. Meanwhile, the graph shows the number of studies published from 2003–2016. The period 2008–2009 was identified as the peak period of citation. Mens [101] identified the prominent and active citer of this cluster and highlighted the role of developers in different phases of software development, such as design, analysis, maintenance, and evolution. Other researchers focused on developing tools that could map teamwork according to product structure. For instance, Georgas and Sarma [108] developed Stcml, a tool for STC modeling using extensible XML-based language. This tool’s main purpose was to identify the core structural components and key developers to measure STC efficiently. Oliva *et al* [109] proposed an STC tool to identify the core product components and key developers. The developers were classified based on contribution and coordination activities. Two categories of developers were identified for STC modeling: developers acting as a bridge and rarely coordinating developers. Kerzazi and El Asri [110] proposed a method to identify the core team members in virtual OSS communities. The core members were deemed as the most significant people and were related to the code review activity. Thus, identifying core members affects the social and technical dimensions of STC measurement directly. Palyart *et al* [61] facilitated the identification of developers who frequently interact in projects with component dependency to achieve coherence with product structure.

6) CLUSTER 12: COORDINATION REQUIREMENT

The last cluster, cluster 12 (labeled as “coordination requirement”), was the smallest with 10 members. The cluster showed high reliability by achieving a 0.977 silhouette value. The coordination requirement was an aspect of STC based on different types of artifacts (i.e., technical or social artifacts).

Cluster 12 contained articles that mainly discussed various types of coordination requirements, methods, and tools (as illustrated in terms cloud in Fig. 27). The articles were mostly cited in 2007. The results showed that Blincoe *et al.*’s [55]

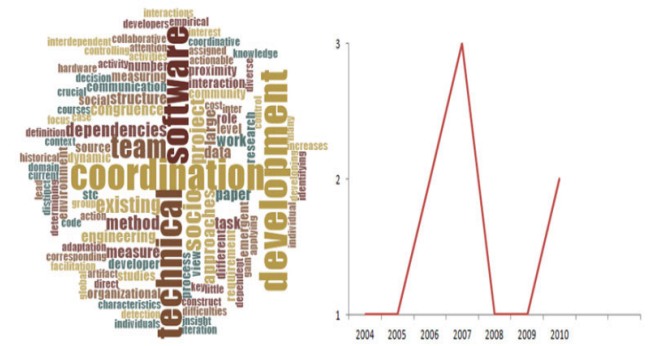


FIGURE 27. Cluster 12 with a cloud of top-ranked terms and coverage period with a number of studies.

study is the representative article in cluster 12. The article proposed an STC tool to quantify the developer’s coordination requirements based on live database analysis. Similarly, Blincoe [57] suggested a method to timely and efficiently identify developers’ coordination requirements using a proximity tool presented in [55]. Fauzi *et al.* [111] also explained various methods for measuring the coordination requirements in the mechanism of software configuration management, particularly in GSD. Moreover, Bettenburg [112] presented a method of extracting coordination requirements from software repositories with different data mining techniques.

V. STAGE 3: RESULT INTERPRETATION

This scientometric study utilized four scientometric techniques: co-word analysis, co-author analysis, co-citation analysis, and cluster analysis. This study was undertaken to report STC’s publication trend and evolution in software development. The result of each technique is briefly summarized in this section.

The present study identified two key research areas in the STC domain based on co-word analysis, which included a network analysis and density visualization of co-occurring keywords. The first area involved studies related to the application of STC in software engineering. The second area included publications on STC about software design basics, development, and challenges.

Subsequently, the significant associations among researchers from various institutions and countries were determined through network analysis and the mapping of relationships among authors (i.e., co-author analysis). The results revealed evolution in the STC field from its conception, as described by Cataldo *et al* [9], to recent developments in STC models and factors, such as awareness, community smell, technical dependencies, human factors, and social aspects. Furthermore, the major countries and groups that lead STC research and exhibit important roles in this area’s progress were also identified. Concerning countries’ contributions, 40% of the publications originated in the United States and Canada.

Concerning co-citation analysis (the third scientometric technique), the obtained network maps were not as strong as

TABLE 13. Summary of co-cited document cluster analysis with prominent keywords.

Cluster#	Theme	Representative Citations	Top 10 Keywords (frequency)	Peak citation year
0	Community structure	Kwan, et al.[5], Chintan et al.[85], Schröter, A et al.[100], Fauzi[111],Kwan et al.[11], Le et al.[103], Bird, C[86], Rytsareva et al.[84],Syeed [87],Betz[8]	Software system (25); product structure (22); conway law (21); open-source software project (20); communication information (19); program logic (18); collaborative activities (16); development repositories (16); actual encoding (15); collaborative effort (15)	2008
4	Socio-technical congruence	Suali et al.[99],Sierra[18],Mens et al.[101], Wang, X et al.[58], Kilamo et al.[104], Zhang et al.[60], Zhang et al.[16] , Landegren et al.[59], Sobri et al.[105]	Work team (96); social interaction (50); coordination level (48); stc control (48); coordination iteration (48); particular interest (46); standard measure (45); controlling stc (42); identifying gap (40); organizational structure (32)	2013
6	Current research	Schröter et al. [113],Schröter et al.[100], Damian et al.[5], Bettenburg [112], Bird[114], Damian[107], Kuhn et al. [115], MacKellar et al.[106]	Predicting failure (40); software team (33); investigating human aspect (30); technical relation (29); developer productivity (26); social dependencies (24); socio-technical congruence (24); coordination requirement (20); socio-technical coordination (20); products quality (19)	2009
9	Product structure	Rytsareva et al.[84],Kerzazi[110] Zanetti[50],Damian[107]	Informal communication (48); collaborative research (38); open source project (37); coordination requirement (35); information technology (27); pattern theory (25); socio-technical structure (24); social network concept (22); core-periphery structure (22); socio-technical challenge (21)	2008
10	Key developer	Mens [101], Georgas and Sarma [108],Kilamo[104], Fauzi et al.[111], Oliva et al.[109], Kerzazi and El Asri [110], Palyart et al.[61]	Social network (53); social skill (40); technical core (37); core team (30); socio-technical perspective (27); complex collaborative process (25); interpersonal relationship (21); productive collaboration (21); mailing list (19); distinct behavior (18)	2008 &2009
12	Coordination requirement	Blincoe [57], Blincoe et al. [55], Fauzi et al.,[111], Bettenburg[112]	Work dependencies (39); software development organization (29); developer activity (23); historical data (21); proactive detection (20); collaborative software development (19); coordination need (17); dependent task (15); direct inter-personal communication (15); source code (6)	2007

in the previous two networks (i.e., co-word and co-author). The main reason for the weak co-citation network maps was the small number of publications in the last few years, revealing that STC is a relatively emerging field. However, the analysis of co-citation networks helped identify prominent journals and conferences that published dominant figures of STC-related articles. In the past 5years, numerous papers have appeared in recognized journals and conferences emphasizing STC in software development. Concerning publication sources, the journal “Communication ACM” and conference “International Conference on Software Engineering” showed many studies.

Table 13 summarizes the outcomes of the documents’ co-citation cluster analysis, listing the major research themes, top-cited documents (representative citations), significant research focuses (via prominent keywords), and peak cited years. The in-depth cluster analysis revealed that the identified research themes were named based on highly cited references within the intended theme. Additionally, we analyzed these themes by investigating the main studies of each theme. Additional details for each theme are as follows:

- In the first theme identified, the phrase “community structure” was obtained via prominent keywords (the top 10 keywords are listed in Table 13). The highlighted keywords depicted that studies in this cluster focused on the effect of team structure and organization on development outcomes. The prominent studies of this cluster utilized most of the references from 2008.
- In the second theme identified (“Socio-technical congruence”), the keywords obtained represented that the studies paid more attention to different aspects and issues of STC measurement techniques and models. For example, work team outcomes and social interaction among team members. The peak referred citation year of this cluster is 2013.
- The studies in the third theme (“current research”) focused on the challenges and under-investigated areas of STC (as evident from the keywords collected). The phrases “predicting failure,” “software team,” and “investigating human aspect” represented the prominent areas of STC. Furthermore, most citations were identified from 2009.

- The fourth theme (“product structure”) represented researchers’ focus on the relationship between STC and the composition of software products. From the collection of keywords, “informal communication” achieved the highest frequency (48), whereas “collaborative research” obtained the minimum frequency (21). The peak identified citation year in this cluster is 2008.
- In the fifth theme identified (“key developer”), the prominent studies covered different characteristics of software developers, techniques, and methods to recognize core developers in software development. The years 2008 and identified as peak citation years for this research theme.
- Studies in the sixth theme (“coordination requirement”) presented techniques to compute one of the STC components. These included “work dependencies,” “software development organization,” “developer activity,” and other aspects (as mentioned in Table 13). The year 2007 appeared as the peak cited year of this cluster.

In summary, the cluster analysis results showed that significant STC-related articles were published from 2007–2013. The year 2008 was considered the most significant year as it had the most cited works. Concerning the evolution of major research themes’ keywords, STC has gained increased attention in software development in recent years. Nonetheless, it can be envisaged that this research area will become central for future software developers and communities by analyzing the number of publications.

This scientometric study was the first to report the publication trends and research patterns related to STC to the best of our knowledge. This study attempted to collect and analyze all relevant data comprehensively. However, some limitations still exist. First, the study aimed to utilize all possible keywords related to the intended topic. However, some false-negative and false-positive results may have existed in the study as it could have included imprecise bibliographic data. Second, the data may contain researcher and institution names with different spellings (extracted from the WoS and Scopus databases). Therefore, multiple profiles for the same author may exist.

VI. CONCLUSION

STC has started gaining the attention of researchers due to the development of large-scale, high-quality software for numerous fields. Although various literature reviews on STC have previously been conducted, they may be prone to subjectivity. Indeed, the literature lacks research on certain STC aspects. More specifically, limited research has analyzed STC’s implementation applicability and issues in software engineering strategies for OSS development. An enhanced awareness through insightful and focused attention on STC research may promote commercialized software support. Therefore, the present study conducted a scientometric analysis based on bibliographic data to understand STC’s status quo, latest trends, and themes. Additionally, a CLR was performed through a detailed analysis of co-cited document

clusters to provide insight into STC-related concepts. This study provided an insightful perspective and visualization maps of STC related literature that benefit researchers and software practitioners to understand the research field, its key applications and significance in software development.

The main contributions of the current study can be presented in four-folds. Firstly, a novel and better understanding of the STC concept based on the historical point of view and quantitative analysis is presented in this study, which provides a roadmap to the researcher to determine the evolutionary trajectories of the STC domain. Secondly, the applicability of four scientometric techniques covers every aspect of STC research such as research patterns, significant collaborators, countries, institutions, journals and conferences, and promising research directions. Thirdly, the cluster analysis with consideration of timeline helps to discover the key research topics and trending themes in STC. This can provide an approach to deeply understand the research topics and development growth related to STC. Lastly, this study provides a paradigm for future studies to explore the evolution in information related to STC research.

This study provided the first scientometric analysis on STC using 306 studies collected from WoS and Scopus from 2000 to 2020. Four scientometric techniques were applied: co-word network analysis, co-author network analysis, and co-citation network analysis. This was undertaken to identify the core researchers, publications, institutions, countries, and research sources in the STC domain. The analysis exhibited that most works in the STC field were performed in isolation with respect to researchers. Therefore, the findings suggest that researchers should collaborate to improve the coordination, conversation, and exchange of diverse intellectual ideas. This scientometric analysis of bibliographic data clarified the results and findings on software development-related STC research by using the VOSviewer, CiteSpace, and Nvivo scientometric tools.

Despite the deliverables of the current study, the results are prone to few limitations. For instance, an STC-related article search was performed based on the initially selected keywords, constraining the latest literature’s boundaries. The study did not focus on the exact mechanism of research employed as it is beyond the scope of this scientometric study. In the future, the research community should attempt to focus on improving the identified aspects of STC. They should also seek solutions to combat challenges in STC research related to software development.

REFERENCES

- [1] A. E. Akgün, “Team wisdom in software development projects and its impact on project performance,” *Int. J. Inf. Manage.*, vol. 50, pp. 228–243, Feb. 2020.
- [2] N. M. Nor, “Modelling lower-level technical dependencies to improve coordination in software engineering projects: A conceptualization,” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 3, pp. 3695–3699, Jun. 2020.
- [3] M. A. Akbar, W. Naveed, S. Mahmood, S. Rafi, A. Alsanad, A. Abdul-Aziz Alsanad, A. Gumaei, and A. Alothaim, “Prioritization of global software requirements’ engineering barriers: An analytical hierarchy process,” *IET Softw.*, vol. 15, no. 4, pp. 277–291, Aug. 2021.

- [4] S. Datta, R. Roychoudhuri, and S. Majumder, "Understanding the relation between repeat developer interactions and bug resolution times in large open source ecosystems: A multisystem study," *J. Softw., Evol. Process.*, vol. 33, no. 4, Apr. 2021, Art. no. e02317.
- [5] I. Kwan, A. Schroter, and D. Damian, "Does socio-technical congruence have an effect on software build success? A study of coordination in a software project," *IEEE Trans. Softw. Eng.*, vol. 37, no. 3, pp. 307–324, May/Jun. 2011.
- [6] D. A. Tamburri, F. Palomba, and R. Kazman, "Success and failure in software engineering: A followup systematic literature review," *IEEE Trans. Eng. Manag.*, vol. 68, no. 2, pp. 599–611, Apr. 2021.
- [7] A. Barrak, E. E. Eghan, B. Adams, and F. Khomh, "Why do builds fail?—A conceptual replication study," *J. Syst. Softw.*, vol. 177, Jul. 2021, Art. no. 110939.
- [8] S. Betz, D. Mite, S. Fricker, A. Moss, W. Afzal, M. Svahnberg, C. Wohlin, J. Borstler, and T. Gorschek, "An evolutionary perspective on socio-technical congruence: The rubber band effect," in *Proc. 3rd Int. Workshop Replication Empirical Softw. Eng. Res.*, Oct. 2013, pp. 15–24.
- [9] M. Cataldo, P. A. Wagstrom, J. D. Herbsleb, and K. M. Carley, "Identification of coordination requirements: Implications for the design of collaboration and awareness tools," presented at the 20th Anniversary Conf. Comput. Supported Cooperat. Work, Banff, AB, Canada, 2006.
- [10] M. Cataldo, J. D. Herbsleb, and K. M. Carley, "Socio-technical congruence: A framework for assessing the impact of technical and work dependencies on software development productivity," presented at the 2nd ACM-IEEE Int. Symp. Empirical Softw. Eng. Meas., Kaiserslautern, Germany, 2008.
- [11] I. Kwan and D. Damian, "Extending socio-technical congruence with awareness relationships," in *Proc. 4th Int. workshop Social Softw. Eng. (SSE)*, Szeged, Hungary, 2011, pp. 23–30.
- [12] I. Kwan, A. Schröter, and D. Damian, "A weighted congruence measure," in *Proc. Workshop SocioTechnical Congruence*, 2009, pp. 1–4.
- [13] S. Choudhary, C. Bogart, C. Rose, and J. Herbsleb, "Using productive collaboration bursts to analyze open source collaboration effectiveness," in *Proc. IEEE 27th Int. Conf. Softw. Anal., Evol. Reengineering (SANER)*, Feb. 2020, pp. 400–410.
- [14] M. Golzadeh, "Analysing socio-technical congruence in the package dependency network of Cargo," presented at the 27th ACM Joint Meeting Eur. Softw. Eng. Conf. Symp. Found. Softw. Eng., Tallinn, Estonia, 2019.
- [15] L. Jiang, K. M. Carley, and A. Eberlein, "Assessing team performance from a socio-technical congruence perspective," presented at the Int. Conf. Softw. Syst. Process, Zurich, Switzerland, 2012.
- [16] W. Zhang, S.-C. Cheung, Z. Chen, Y. Zhou, and B. Luo, "File-level socio-technical congruence and its relationship with bug proneness in OSS projects," *J. Syst. Softw.*, vol. 156, pp. 21–40, Oct. 2019.
- [17] J. Portillo-Rodríguez, A. Vizcaíno, M. Piattini, and S. Beecham, "Using agents to manage socio-technical congruence in a global software engineering project," *Inf. Sci.*, vol. 264, pp. 230–259, Apr. 2014.
- [18] J. M. Sierra, A. Vizcaíno, M. Genero, and M. Piattini, "A systematic mapping study about socio-technical congruence," *Inf. Softw. Technol.*, vol. 94, pp. 111–129, Feb. 2018.
- [19] I. Inayat, S. S. Salim, S. Marczak, and Z. Mk, "Identifying and reviewing the most relevant socio-technical aspects of requirements-driven collaboration in agile teams," presented at the Conf., Int. Conf. Adv. Eng. Technol. (ICAET), Singapore, 2014.
- [20] A. J. Suali, S. Fauzi, W. Sobri, and M. H. N. M. Nasir, "Developers' coordination issues and its impact on software quality: A systematic review," presented at the 3rd Int. Conf. Sci. Inf. Technol. (ICSITech), Bandung, Indonesia, 2017.
- [21] M. Hammersley, "On 'systematic' reviews of research literatures: A 'narrative' response to Evans & Benefield," *Brit. Educ. Res. J.*, vol. 27, no. 5, pp. 543–554, Dec. 2001.
- [22] D. Yu and T. Pan, "Tracing knowledge diffusion of TOPSIS: A historical perspective from citation network," *Expert Syst. Appl.*, vol. 168, Apr. 2021, Art. no. 114238.
- [23] P. Dasgupta, J. VanSwearingen, A. Godfrey, M. Redfern, M. Montero-Odasso, and E. Sejdic, "Acceleration gait measures as proxies for motor skill of walking: A narrative review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 249–261, 2021.
- [24] H. Nina, J. A. Pow-Sang, and M. Villavicencio, "Systematic mapping of the literature on secure software development," *IEEE Access*, vol. 9, pp. 36852–36867, 2021.
- [25] M. R. Belgaum, S. Musa, M. M. Alam, and M. M. Su'ud, "A systematic review of load balancing techniques in software-defined networking," *IEEE Access*, vol. 8, pp. 98612–98636, 2020.
- [26] Y. Pan and L. Zhang, "Roles of artificial intelligence in construction engineering and management: A critical review and future trends," *Autom. Construct.*, vol. 122, Feb. 2021, Art. no. 103517.
- [27] A. Saka and M. Igami, "Mapping modern science using co-citation analysis," in *Proc. 11th Int. Conf. Inf. Visualizat. (IV)*, Jul. 2007, pp. 453–458.
- [28] O. J. de Oliveira, F. F. da Silva, F. Juliani, L. C. F. M. Barbosa, and T. V. Nunes, "Bibliometric method for mapping the state-of-the-art and identifying research gaps and trends in literature: An essential instrument to support the development of scientific projects," in *Scientometrics Recent Advances*. Rijeka, Croatia: InTechOpen, 2019.
- [29] N. J. van Eck, "Methodological advances in bibliometric mapping of science," in *Cancer Research*. Jan. 2011.
- [30] O. Ellegaard and J. A. Wallin, "The bibliometric analysis of scholarly production: How great is the impact?" *Scientometrics*, vol. 105, no. 3, pp. 1809–1831, Dec. 2015.
- [31] D. Yu, Z. Xu, and W. Pedrycz, "Bibliometric analysis of rough sets research," *Appl. Soft Comput.*, vol. 94, Sep. 2020, Art. no. 106467.
- [32] T. Magerman, B. Van Looy, B. Baesens, and K. Debackere, "Assessment of latent semantic analysis (LSA) text mining algorithms for large scale mapping of patent and scientific publication documents," FBE, Res. Rep. MSI_1114, 2011.
- [33] M. C. Kim and Y. Zhu, "Scientometrics of scientometrics: Mapping historical footprint and emerging technologies in scientometrics," in *Scientometrics*. 2018, p. 9.
- [34] V. Venkatraman, N. Arunkumar, A. Chantre-Astaiza, A. I. Muñoz-Mazon, L. Fuentes-Moraleda, and M. S. Khan, "Mapping the structure and evolution of heavy vehicle research: A scientometric analysis and visualisation," *Int. J. Heavy Vehicle Syst.*, vol. 25, nos. 3–4, pp. 344–368, 2018.
- [35] E. Basak and F. Calisir, "Publication trends in Facebook: A scientometric study," presented at the department nameInternational Conf. Trends Econ., Humanities Manage., Singapore, 2015.
- [36] O. Konur, "The evaluation of the global research on the education: A scientometric approach," *Procedia-Social Behav. Sci.*, vol. 47, pp. 1363–1367, Jan. 2012.
- [37] J. L. Schaefer, J. C. M. Siluk, I. C. Baierle, and E. O. B. Nara, "A scientometric approach to analyze scientific development on renewable energy sources," *J. Data Inf. Sci.*, vol. 6, no. 3, Nov. 2020, Art. no. 000010247820210009.
- [38] P. Savov, A. Jatowt, and R. Nielek, "Identifying breakthrough scientific papers," *Inf. Process. Manage.*, vol. 57, no. 2, Mar. 2020, Art. no. 102168.
- [39] C. Chen, "CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 57, no. 3, pp. 359–377, 2006.
- [40] N. J. van Eck and L. Waltman, *VOSviewer Manual*, vol. 1. Leiden, The Netherlands: Univeriteit Leiden, 2013.
- [41] A. Castleberry, "NVivo 10 [software program]. version 10. QSR international; 2012," *Amer. J. pharmaceutical Educ.*, vol. 78, 2014.
- [42] F. Qin, Y. Zhu, T. Ao, and T. Chen, "The development trend and research frontiers of distributed hydrological models—Visual bibliometric analysis based on citespace," *Water*, vol. 13, no. 2, p. 174, 2021.
- [43] B. Chen and S. Shin, "Bibliometric analysis on research trend of accidental falls in older adults by using citespace—Focused on web of science core collection (2010–2020)," *Int. J. Environ. Res. Public Health*, vol. 18, no. 4, p. 1663, 2021.
- [44] V.-M. Muñoz-Sánchez and A.-M. Pérez-Flores, "The connections between ecological values and organic food: Bibliometric analysis and systematic review at the start of the 21st century," *Sustainability*, vol. 13, no. 7, p. 3616, Mar. 2021.
- [45] I. Moya-Clemente, G. Ribes-Giner, and J. C. Chaves-Vargas, "Sustainable entrepreneurship: An approach from bibliometric analysis," *J. Bus. Econ. Manage.*, vol. 22, no. 2, pp. 297–319, Feb. 2021.
- [46] D. N. Effendi, W. Anggraini, A. Jatmiko, H. Rahmayanti, and I. Z. Ichsan, "Bibliometric analysis of scientific literacy using VOS viewer: Analysis of science education," in *Proc. J. Phys., Conf.*, vol. 1796, Jan. 2021, Art. no. 012096.
- [47] W. Y. Song Yahan and F. Daoyang, "Knowledge mapping of three-dimensional printing in biomedical field based on VOSviewer," *Chin. J. Tissue Eng. Res.*, vol. 25, pp. 2385–2393, May 2021.
- [48] D. E. Prastyana and A. Nurmandi, "A bibliometric analysis of E-democracy on government research," *Jurnal Ilmiah Mimbar Demokrasi*, vol. 20, no. 2, pp. 71–80, Apr. 2021.

- [49] P. Martinez, M. Al-Hussein, and R. Ahmad, "A scientometric analysis and critical review of computer vision applications for construction," *Autom. Construct.*, vol. 107, Nov. 2019, Art. no. 102947.
- [50] M. S. Zanetti, "The co-evolution of socio-technical structures in sustainable software development: Lessons from the open source software communities," presented at the 34th Int. Conf. Softw. Eng., Zurich, Switzerland, 2012.
- [51] K. Harrison-Broninski and J. J. Korhonen, "Collaboration infrastructure for the learning organization," in *Proc. Int. Conf. Bus. Inf. Syst.* Berlin, Germany: Springer, 2012, pp. 120–131.
- [52] T. Kocaturk, R. Balbo, B. Medjdoub, and A. Veliz Reyes, "An innovative approach to technology mediated architectural design education," presented at the eCAADe: 30th Int. Conf. Educ. Res. Comput. Aided Architectural Design Eur., Prague, Czech Republic, 2012.
- [53] K. Blincoe, F. Harrison, N. Kaur, and D. Damian, "Reference coupling: An exploration of inter-project technical dependencies and their characteristics within large software ecosystems," *Inf. Softw. Technol.*, vol. 110, pp. 174–189, Jun. 2019.
- [54] I. Inayat, S. S. Salim, and Z. M. Kasirun, "Socio-technical aspects of requirements-driven collaboration (RDC) in agile software development methods," in *Proc. IEEE Conf. Open Syst.*, Oct. 2012, pp. 1–6.
- [55] K. Blincoe, G. Valetto, and S. Goggins, "Proximity: A measure to quantify the need for developers' coordination," presented at the ACM conference Comput. Supported Cooperat. Work, Seattle, WA, USA, 2012.
- [56] J. McAuley, A. O'Connor, and D. Lewis, "Exploring reflection in online communities," presented at the 2nd Int. Conf. Learn. Anal. Knowl., Vancouver, BC, Canada, 2012.
- [57] G. Valetto, K. Blincoe, and S. P. Goggins, "Actionable identification of emergent teams in software development virtual organizations," presented at the 3rd Int. Workshop Recommendation Syst. Softw. Eng., Zurich, Switzerland, 2012.
- [58] X. Wang, L. Xiao, Y. Yang, X. Xu, and Y. Jiang, "Identifying TraIn: A neglected form of socio-technical incongruence," in *Proc. 40th Int. Conf. Softw. Eng.: Companion (ICSE-Companion)*, May 2018, pp. 358–359.
- [59] F. Landegren, M. Höst, and P. Möller, "Simulation based assessment of resilience of two large-scale socio-technical IT networks," *Int. J. Crit. Infrastruct. Protection*, vol. 23, pp. 112–125, Dec. 2018.
- [60] W. Zhang, Z. Chen, and B. Luo, "Does socio-technical congruence have an effect on continuous integration build failures? An empirical study on 10 GitHub projects," presented at the IEEE Int. Conf. Softw. Qual., Rel. Secur. (QRS), Lisbon, Portugal, 2018.
- [61] M. Palyart, G. C. Murphy, and V. Masrani, "A study of social interactions in open source component use," *IEEE Trans. Softw. Eng.*, vol. 44, no. 12, pp. 1132–1145, Dec. 2018.
- [62] J. Pollack and D. Adler, "Emergent trends and passing fads in project management research: A scientometric analysis of changes in the field," *Int. J. Project Manage.*, vol. 33, no. 1, pp. 236–248, Jan. 2015.
- [63] C. Chen and M. Song, "Visualizing a field of research: A methodology of systematic scientometric reviews," *PLoS ONE*, vol. 14, no. 10, Oct. 2019, Art. no. e0223994.
- [64] H.-N. Su and P. C. Lee, "Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in technology foresight," *Scientometrics*, vol. 85, no. 1, pp. 65–79, 2010.
- [65] V. V. E. Nalimov and Z. M. Mul'chenko, "Measurement of science. Study of the development of science as an information process," *Foreign Technol. Div Wright-Patterson AFB, Tech. Rep.* OHIO1971.
- [66] K. Börner, C. Chen, and K. W. Boyack, "Visualizing knowledge domains," *Annu. Rev. Inf. Sci. Technol.*, vol. 37, no. 1, pp. 179–255, Jan. 2005.
- [67] N. Yang, L. Sun, and J. Tan, "Theme trends and knowledge structure of assisted reproductive technology and birth defects: A quantitative and co-word analysis," *J. Obstetrics Gynaecol. Res.*, vol. 47, no. 5, pp. 1780–1788, 2021.
- [68] Y. Yano, T. Nakamura, S. Ishitsuka, and A. Maruyama, "Consumer attitudes toward vertically farmed produce in Russia: A study using ordered logit and co-occurrence network analysis," *Foods*, vol. 10, no. 3, p. 638, Mar. 2021.
- [69] D. Yu, G. Kou, Z. Xu, and S. Shi, "Analysis of collaboration evolution in AHP research: 1982–2018," *Int. J. Inf. Technol. Decis. Making*, vol. 20, no. 1, pp. 7–36, 2021.
- [70] D. Yu and T. Pan, "Tracing the main path of interdisciplinary research considering citation preference: A case from blockchain domain," *J. Informetrics*, vol. 15, no. 2, May 2021, Art. no. 101136.
- [71] A. Mas-Tur, N. Roig-Tierno, S. Sarin, C. Haon, T. Segó, M. Belkhouja, A. Porter, and J. M. Merigó, "Co-citation, bibliographic coupling and leading authors, institutions and countries in the 50 years of technological forecasting and social change," *Technol. Forecasting Social Change*, vol. 165, Apr. 2021, Art. no. 120487.
- [72] L. Phan Tan, "Mapping the social entrepreneurship research: Bibliographic coupling, co-citation and co-word analyses," *Cogent Bus. Manage.*, vol. 8, no. 1, Jan. 2021, Art. no. 1896885.
- [73] Y. Sun, C. Li, Y. Zhao, and J. Sun, "Trends and developments in oral health literacy: A scientometric research study (1991–2020)," *BDJ Open*, vol. 7, p. 13, Mar. 2021.
- [74] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, Mar. 1977.
- [75] J. M. Kleinberg, "Bursty and hierarchical structure in streams," *Data Mining Knowl. Discovery*, vol. 7, no. 4, pp. 373–397, 2003.
- [76] Y. Takeda and Y. Kajikawa, "Tracking modularity in citation networks," *Scientometrics*, vol. 83, no. 3, pp. 783–792, Jun. 2010.
- [77] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. 1, pp. 53–65, 1987.
- [78] A. de la Hoz-Correa, F. Muñoz-Leiva, and M. Bakucz, "Past themes and future trends in medical tourism research: A co-word analysis," *Tourism Manage.*, vol. 65, pp. 200–211, Apr. 2018.
- [79] X. Y. Leung, J. Sun, and B. Bai, "Bibliometrics of social media research: A co-citation and co-word analysis," *Int. J. Hospitality Manage.*, vol. 66, pp. 35–45, Sep. 2017.
- [80] W. Zhao, J. Mao, and K. Lu, "Ranking themes on co-word networks: Exploring the relationships among different metrics," *Inf. Process. Manage.*, vol. 54, no. 2, pp. 203–218, Mar. 2018.
- [81] M. Paasivaara, K. Blincoe, C. Lassenius, D. Damian, J. Sheoran, and F. Harrison, "Learning global agile software engineering using same-site and cross-site teams," presented at the IEEE/ACM 37th IEEE Int. Conf. Softw. Eng., Florence, Italy, 2015.
- [82] S. Marczak, I. Kwan, and D. Damian, "Investigating collaboration driven by requirements in cross-functional software teams," in *Proc. Collaboration Intercultural Issues Requirements: Commun., Understand. Softskills*, Aug. 2009, pp. 15–22.
- [83] C. Hunsen, J. Siegmund, and S. Apel, "On the fulfillment of coordination requirements in open-source software projects: An exploratory study," *Empirical Softw. Eng.*, vol. 25, no. 6, pp. 4379–4426, Nov. 2020.
- [84] I. Rytasareva, L. Qize, E. Conner, A. Kalyanaraman, and J. H. Panchal, "Evaluating socio-technical coordination in open-source communities: A cluster-based approach," in *ASME Comput. Inf. Eng.*, Chicago, IL, USA, 2012.
- [85] C. Amrit and J. Hillegersberg, "Exploring the impact of socio-technical core-periphery structures in open source software development," *J. Inf. Technol.*, vol. 25, no. 2, pp. 216–229, Jun. 2010.
- [86] C. Bird, "Sociotechnical coordination and collaboration in open source software," presented at the 27th IEEE Int. Conf. Softw. Maintenance, 2011.
- [87] M. M. M. Syeed, I. Hammouda, and C. Berko, "Exploring socio-technical dependencies in open source software projects: Towards an automated data-driven approach," in *Proc. Int. Conf. Making Sense Converging Media*, Tampere, Finland, 2013, pp. 273–280.
- [88] D. A. Tamburri, F. Palomba, and R. Kazman, "Exploring community smells in open-source: An automated approach," *IEEE Trans. Softw. Eng.*, vol. 47, no. 3, pp. 630–652, Mar. 2021.
- [89] Y. Bu, B. Wang, Z. Chinchilla-Rodríguez, C. R. Sugimoto, Y. Huang, and W.-B. Huang, "Considering author sequence in all-author co-citation analysis," *Inf. Process. Manage.*, vol. 57, no. 6, Nov. 2020, Art. no. 102300.
- [90] M. E. J. Newman, "Coauthorship networks and patterns of scientific collaboration," *Proc. Nat. Acad. Sci. USA*, vol. 101, pp. 5200–5205, Apr. 2004.
- [91] J. P. A. Ioannidis, "Measuring co-authorship and networking-adjusted scientific impact," *PLoS ONE*, vol. 3, no. 7, p. e2778, Jul. 2008.
- [92] C. Chen and S. Morris, "Visualizing evolving networks: Minimum spanning trees versus pathfinder networks," in *Proc. IEEE Symp. Inf. Vis.*, Oct. 2003, pp. 67–74.
- [93] P. Benckendorff and A. Zehrer, "A network analysis of tourism research," *Ann. Tourism Res.*, vol. 43, pp. 121–149, Oct. 2013.
- [94] G. Surwase, B. Kademani, and K. Bhanumurthy, "Mapping the structure and development of science using co-citation analysis," *Tech. Rep.*, 2012.

- [95] F. P. Brooks, Jr., *The Mythical Man-Month, Anniversary*. Reading, MA, USA: Addison-Wesley, 1995.
- [96] M. E. Sosa, S. D. Eppinger, and C. M. Rowles, "The misalignment of product architecture and organizational structure in complex product development," *Manage. Sci.*, vol. 50, no. 12, pp. 1674–1689, Dec. 2004.
- [97] M. Cataldo and J. D. Herbsleb, "Coordination breakdowns and their impact on development productivity and software failures," *IEEE Trans. Softw. Eng.*, vol. 39, no. 3, pp. 343–360, Mar. 2013.
- [98] Y. Zhu, M. C. Kim, and C. Chen, "An investigation of the intellectual structure of opinion mining research," *Inf. Res.*, vol. 22, p. 739, Mar. 2017.
- [99] A. Suali, S. Fauzi, M. Nasir, and W. Sobri, "A brief method to assess the association of socio and technical dependencies on software quality," *J. Fundam. Appl. Sci.*, vol. 9, no. 5S, pp. 101–114, 2017.
- [100] A. Schröter, "Failure preventing recommendations," in *Proc. 32nd ACM/IEEE Int. Conf. Softw. Eng. (ICSE)*, May 2010, pp. 397–400.
- [101] T. Mens, "An ecosystemic and socio-technical view on software maintenance and evolution," presented at the IEEE Int. Conf. Softw. Maintenance Evol. (ICSME), Raleigh, NC, USA, 2016.
- [102] G. Valetto, M. Helander, K. Ehrlich, S. Chulani, M. Wegman, and C. Williams, "Using software repositories to investigate socio-technical congruence in development projects," in *Proc. 4th Int. Workshop Mining Softw. Repositories (MSR:ICSE Workshops)*, May 2007, p. 25.
- [103] Q. Le and J. H. Panchal, "Analysis of the interdependent co-evolution of product structures and community structures using dependency modelling techniques," *J. Eng. Des.*, vol. 23, nos. 10–11, pp. 807–828, Nov. 2012.
- [104] T. Kilamo, M. Leppänen, and T. Mikkonen, "The social developer: Now, then, and tomorrow," presented at the 7th Int. Workshop Social Softw. Eng., Bergamo, Italy, 2015.
- [105] W. Sobri, S. Fauzi, M. Nasir, R. Ahmad, and A. Suali, "A mechanism to assess the relationship between socio-technical congruence and project performance in incremental model," *J. Fundam. Appl. Sci.*, vol. 9, no. 5s, pp. 58–74, 2017.
- [106] B. Mackellar, "Analyzing coordination among students in a software engineering project course," presented at the 26th Int. Conf. Softw. Eng. Educ. Training (CSEE&T), San Francisco, CA, USA, 2013.
- [107] D. Damian, R. Helms, I. Kwan, S. Marczak, and B. Koelewijn, "The role of domain knowledge and cross-functional communication in socio-technical coordination," presented at the 2013 Int. Conf. Softw. Eng., San Francisco, CA, USA, 2013.
- [108] J. C. Georgas and A. Sarma, "STCML: An extensible XML-based language for socio-technical modeling," in *Proc. 4th Int. Workshop Cooperat. Hum. Aspects Softw. Eng. (ICSE)*, 2011, pp. 61–64.
- [109] G. Oliva, T. da Silva, M. Gerosa, F. W. S. Santana, M. L. Werner, C. R. B. de Souza, and K. C. M. de Oliveira, "Evolving the system's core: A case study on the identification and characterization of key developers in Apache Ant," *Comput. Informat.*, vol. 34, no. 3, pp. 678–724, Jan. 2015.
- [110] N. Kerzazi and I. El Asri, "Who can help to review this piece of code?" in *17th Work. Conf. Virtual Enterprises (PRO-VE)*, Porto, Portugal, 2016, pp. 289–301.
- [111] S. S. M. Fauzi, P. L. Bannerman, and M. Staples, "Software configuration management in global software development: A systematic map," presented at the Asia Pacific Softw. Eng. Conf., 2010.
- [112] N. Bettenburg, "Mining development repositories to study the impact of collaboration on software systems," presented at the 19th ACM SIGSOFT Symp. 13th Eur. Conf. Found. Softw. Eng., Szeged, Hungary, 2011.
- [113] A. Schröter, "Predicting build outcome with developer interaction in Jazz," presented at the 32nd ACM/IEEE Int. Conf. Softw. Eng., vol. 2, Cape Town, South Africa, 2010.
- [114] C. Bird, B. Murphy, N. Nagappan, and T. Zimmermann, "Empirical software engineering at Microsoft research," presented at the ACM 2011 Conf. Comput. Supported Cooperat. Work, Hangzhou, China, 2011.
- [115] A. Kuhn, "Immediate search in the IDE as an example of socio-technical congruence in search-driven development," presented at the ICSE Workshop Search-Driven Develop.: Users, Infrastruct., Tools Eval., Cape Town, South Africa, 2010.



BINISH RAZA received the bachelor's degree in computer engineering from Pakistan, in 2012, and the master's degree from the University of Engineering and Technology Lahore (UET), Lahore, Pakistan, in 2013. She is currently pursuing the Ph.D. degree in software engineering with the Faculty of Computer Science and Information Technology, Universiti Malaya. Her research interests include software engineering, data mining, empirical software engineering, mining software repository,

social network analysis, socio-technical congruence, and software modeling.



RODINA AHMAD received the bachelor's degree from USA, in 1984, the master's degree in computer science, in 1989, and the Ph.D. degree in information systems management from the National University of Malaysia, in 2006. She started teaching at the Department of Computer Science, University Technology Malaysia, in 1991. She moved to Universiti Malaya to continue her teaching career, in 1993. Her research interests include requirements engineering and organizational analysis and modeling. She received the Petronas Scholarship for her bachelor's and master's degrees.



MOHD HAIRUL NIZAM BIN MD NASIR (Member, IEEE) received the bachelor's and master's degrees in computer science from Universiti Malaya, Malaysia, and the Ph.D. degree in computer science from Universiti Teknologi Malaysia. He is currently a Senior Lecturer at the Faculty of Computer Science and Information Technology, Universiti Malaya. His research interests include project management, software quality, empirical software engineering, software process improvement, and software process models.



SHUKOR SANIM MOHD FAUZI received the Diploma degree in computer science and the B.Sc. degree (Hons.) in information systems engineering from the Universiti Teknologi MARA, Malaysia, the M.Sc. degree from the Centre for Advanced Software Engineering, Universiti Teknologi Malaysia, and the Ph.D. degree in software engineering from the University of New South Wales (UNSW). He is currently a Senior Lecturer with the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA. His research interests include software engineering, empirical software engineering, mining software repository, social network analysis, socio-technical congruence, computer supportive collaborative work, and software process.

...