

Bibliometric Analysis of the Tertiary Study on Agile Software Development using Social Network Analysis

Egemen Bayram

Department of Computer Engineering
Faculty of Technology
Marmara University
İstanbul, Türkiye
egemenbayram@yahoo.com

Buket Doğan

Department of Computer Engineering
Faculty of Technology
Marmara University
İstanbul, Türkiye
buketb@marmara.edu.tr

Volkan Tunali

Department of Software Engineering
Faculty of Engineering and Natural
Sciences, Maltepe University
İstanbul, Türkiye
volkan.tunali@gmail.com

Abstract— This study aims to examine the systematic literature reviews published on the Agile Software Development subject between 2013 and 2018 and to examine the citation relationships among the studies within the scope of the tertiary study with the help of social network analysis. In this study, the publications within the scope were visualized with VOSviewer and Gephi social network analysis tools, and the relations between publishing institutions, countries were revealed. In citation analysis; according to the total link strength, it is seen that UK, Spain and Slovenia are at the forefront at institution and country level. Brazil has the highest citation value and provides the link between the two large clusters obtained in the analysis. In the bibliographic coupling analysis, the five most active countries at the country level were Brazil, Germany, Finland, Malaysia and Pakistan. When the same analysis is made at the institution level, the top five institutions are in Brazil, Switzerland, Peru and Pakistan. The findings of the study indicate that developing countries have more studies on the subject and that the cited publications are mostly from developed countries; European countries seem to be more collaborative based on citation analysis yet developing countries such as Brazil and Malaysia have also relations with them; the number of publications is not directly proportional to the citations.

Keywords—*agile software development, tertiary study, social network analysis, systematic literature reviews, bibliometric analysis*

I. INTRODUCTION

After the declaration of the Agile Manifesto by a group of software development experts in 2001, many researchers examined the agile software development topic and produced various publications in this field. Agile software development applications such as Extreme Programming (XP) and Scrum are increasingly adopted to address the challenges of volatile business environments where markets and technologies are evolving rapidly and unexpected situations arise. As seen in Fig. 1, scientific studies on agile software development, which has a history of nearly two decades, are increasing steadily every year.

Most of the scientific research begins with a literature review. A systematic literature review is a tool for identifying, evaluating and interpreting all current research related to a particular research question or subject area and first appeared in the field of medicine [1]. Systematic literature reviews aim to provide an unbiased assessment of a research topic using a reliable, rigorous and auditable methodology [1].

Moreover, a systematic literature review provides a repeatable research method that provides sufficient detail to be reproduced by other researchers when properly

implemented. Besides, detailed documentation of the steps performed within the systematic literature review allows for an in-depth evaluation of the work carried out [2].

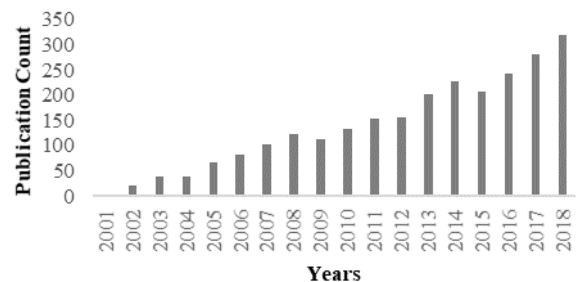


Fig. 1. Distribution of agile software development publications indexed in Scopus Database.

Researchers who want to carry out a systematic literature review should thoroughly examine all available literature on the subject they are dealing with. While individual studies contributing to a systematic review are called primary studies, a systematic literature review is considered as a secondary study. The study of systematic scans with the methodical literature review method is the tertiary study.

On the other hand, social network analysis is the process of examining social structures using networks and graph theory. It characterizes network structures in terms of nodes and relationships, edges or connections that connect them [3].

SNA (Social Network Analysis) in the social sciences began with the introduction of the sociogram concept in 1937 by Moreno [4]. Sociograms are considered as a representation, where interpersonal relationship patterns can be studied as larger social clusters. Sociograms are basic tools for investigating the texture of interpersonal relationships, as each node individual in the graph can represent relationships such as the exchange of information between these nodes [3].

One of the biggest steps in the progress of social network analysis has been the development of social network analysis tools with the development of computer technology. UCINET, the most popular of these tools, dates back to the 1980s. In addition to UCINET, the development of many powerful social network analysis tools such as PAJEK, NodeXL, Gephi, and the method of social network analysis have become attractive for different disciplines such as engineering, primarily sociology, anthropology and political science, and are now widely used by researchers [5].

The purpose of this study is to determine the recent trends in agile software development using the tertiary study method, to identify gaps in the literature and to reveal the citation connections about countries and institutions between secondary studies through the social network analysis. Within the scope of this study, the publications to be analyzed were determined by using the tertiary study method, then VOSviewer software was used for analysis.

II. MATERIAL AND METHOD

In the first stage, a literature review was conducted for a tertiary study to identify and analyze existing Systematic Literature Reviews (SLR) related to Agile Software Development (ASD). The guidelines proposed by Kitchenham et al. [2] were applied for the tertiary study conducted within the scope of this study.

The research questions of tertiary study were determined by adapting the research questions Kitchenham et al. [6] to cover publication between 2013 and 2018. Accordingly, the research questions of our study are given below:

- RQ1: How many SLRs were published related with agile software development between 2013 and 2018?
- RQ2: Which institutions and countries are most active in agile software development SLRs between 2013 and 2018?
- RQ3: How is the social network structure of countries and institutions in SLRs between 2013 and 2018?

The data of an SLR in this context also served as the basis for social network analysis.

In the selection of the database, it is considered that the databases contain publications in the fields of Computer Engineering, Software Engineering and Computer Science. As a result, Scopus, Science Direct, IEEE and Web of Science databases were selected to search. As stated in the studies of SLRs published in previous years, these databases cover a significant part of the publications on agile software development.

While determining the search texts, the texts used in the publication of Kitchenham et al. [6] were taken into consideration and these texts were adapted to the scope of the study. Searches on determined texts were carried out on title, abstract and keywords as seen in SLRs published in previous years. Examples of the search strings specified are given below:

- “agile” AND “review of studies”
- “agile” AND “structured review”
- “agile” AND “systematic review”
- “agile” AND “literature review”
- “agile” AND “literature analysis”

As the search interval, January 2013 - December 2018 were determined and publications (research articles and conference papers) published in the last 6 years were researched. Searched databases were March 16, 2019 for Web of Science, March 18, 2019 for IEEE, and March 30, 2019 for Science Direct and Scopus [5].

896 publications found in the searched databases for each search string. Duplicate publications were eliminated and the

remaining 409 publications were subjected to the inclusion/exclusion criteria at the next stage. After this review, 281 publications were eliminated and the remaining 118 publications were examined according to the quality criteria in the next step. The quality criteria to be applied were chosen as the DARE quality criteria used in tertiary studies [7]. After the evaluation stage made according to the quality criteria, 19 publications under 2 points were eliminated. The bibliographic data of five publications that were not indexed in the Scopus database were not available, so the remaining 94 publications formed the dataset of this study for SNA.

A. Social Network Analysis Software

In this study, VOSviewer and Gephi were used for SNA. VOSviewer is a software tool designed specifically to create and visualize bibliometric maps and to create a graphic representation of maps of scientific papers [8, 9]. Developed by the Center for Science and Technology Research at Leiden University in the Netherlands. VOSviewer can create clusters of given node points. For the algorithm used for clustering, it gives the user the right to make changes in five parameters, two simple and three advanced. Simple parameters are resolution and minimum cluster size. The resolution parameter determines the level of detail of the clusters. The parameter cannot get a negative value.

Also, there are three advanced parameters for the clustering algorithm. These are random initialization, iterations and random seeds. The random initialization parameter shows how many times the optimization algorithm of the clustering technique will work. The default value is 10, the maximum value is 10,000. Iterations parameter shows how many iterations will take place within the clustering algorithm. Its default value is 10 and its maximum value is 1 million. Random seed is the value that makes the random number generation technique used by many programs repeatable. The equation used for clustering in VOSviewer is as follows:

$$V(c_1, \dots, c_n) = \sum_{i < j} \delta(c_i, c_j)(s_{ij} - \gamma) \quad (1)$$

In (1), nodes are assigned to clusters to maximize the equation. c_i represents the cluster to which node i is assigned. $\delta(c_i, c_j)$ specifies a function with 1 if $c = c_j$ and 0 if not. γ specifies a resolution parameter that determines the resolution level of the cluster.

In this study; random initialization, iteration, and random seed parameters were set to 10,000, 20, and 0 respectively.

Gephi is an open-source network analysis and visualization software written in Java on the NetBeans platform. It uses a 3D rendering engine to view large networks in real-time and speed up research [10].

B. Preprocessing Stage

Citation analysis determines network relationships according to the number of publications the authors cite each other. The citation analysis does not only take place on the basis of publication, the relations can also be visualized on the basis of the author, the institutions where the authors work, the countries, and the source of the publication (scientific journal, book or conference).

Citation analysis shows the effects of publications, authors, sources, institutions or countries on the subject through the number of citations [11]. It provides rapid publication of important publications in the field. Newer publications are disadvantageous in this analysis compared to older publications, since earlier publications will have more citations than newer publications.

Bibliographic coupling is introduced in the literature by Meyer Mike Kessler in 1963 [12]. This analysis matches the publications that share common references. It is a useful approach that can be used in all fields of research, as it helps the researcher find related publications in the past.

The bibliographic data of 94 selected secondary studies were downloaded from the Scopus database. This data were checked and edited during the preprocessing phase.

Authors with multiple names can use a single name in some publications, while some authors may use different or multiple names together. Also, the reference style can be different in journals or conferences, the same publication may appear different as a reference in the data downloaded from Scopus. Since social network analysis tools cannot detect this situation, the preprocessing process on the data has been completed carefully. The number of publications in our dataset, which was 4,914, was reduced to 3,697 after preprocessing stage.

III. RESULTS

A. Citation Analysis

Our dataset contains 270 researchers and 133 institutions. In our dataset, it can be seen in Table I that seven institutions publish a total of three articles, which is the largest total article value published at the institution level.

TABLE I. THE MOST ACTIVE INSTITUTIONS

Institution (Country)	Citation	Publication
Blekinge Institute of Technology (Sweden)	73	3
Federal University of Campina Grande (Brazil)	9	3
Federal Institute of Paraiba (Brazil)	9	3
Pontifical Catholic University of Peru (Peru)	20	3
Pontifical Catholic University of Rio Grande do Sul (Brazil)	139	3
University of Brazil (Brazil)	1	3
University of Oulu (Finland)	83	3

TABLE II. THE MOST ACTIVE INSTITUTIONS OF CITATION ANALYSIS BY TLS (TLS-TOTAL LINK STRENGTH, THE-TIMES HIGHER EDUCATION RANKING)

Institution (Country)	TLS	Link Count	Citation Count	THE
University of York (UK)	10	8	53	100-150
University of Seville (Spain)	6	6	71	800-1000
Andalusian Agency of Cultural Institutions (Spain)	5	5	19	-
University of Maribor (Slovenia)	4	4	6	800-1000
MSG Life Odateam D.O.O. (Slovenia)	4	4	6	-

TABLE III. THE MOST ACTIVE COUNTRIES OF CITATION ANALYSIS BY TLS

Country	TLS	Link Count	Citation Count	Publication Count
United Kingdom	10	6	69	4
Spain	7	4	82	5
Slovenia	4	4	6	1
Brazil	2	2	421	21
Chile	1	1	7	2
Finland	2	2	236	9
Netherlands	2	2	146	4
Peru	2	2	20	3

TABLE IV. CITATION CLUSTERS BY COUNTRY

Cluster	Countries
1	Chile, Peru, Spain, UK
2	Iran, Malaysia, Slovenia
3	Finland, Switzerland
4	Denmark, Netherlands

The cited publications in SLRs related to agile software development published between 2013 and 2018 were examined based on the authors' institutions. According to the total link strength, the first five of the 133 institutions in the dataset are shown in Table II.

Our dataset was also analyzed on the basis of the country of the authors. Accordingly, a total of 33 countries have been identified.

Table III shows the most active countries of citation analysis with Gephi ordered by TLS. While the largest value in TLS and Citation Number is Brazil, it is seen that the UK has the largest value in Number of Publication and Numbers of Citation.

VOSviewer analysis results about country-level citation analysis can be seen in Fig. 2. The size of the nodes in Fig. 2 is shown by the number of citations. Table IV shows the country citation analysis cluster results from VOSviewer. The contents of the clusters are shown in Fig. 2 with arrows. It is seen that 7 out of 11 countries that provide a connection between clusters are from Europe, which shows that the authors refer to each other more in Europe. It can be seen that Brazil does not belong to any cluster and connects the two large clusters.

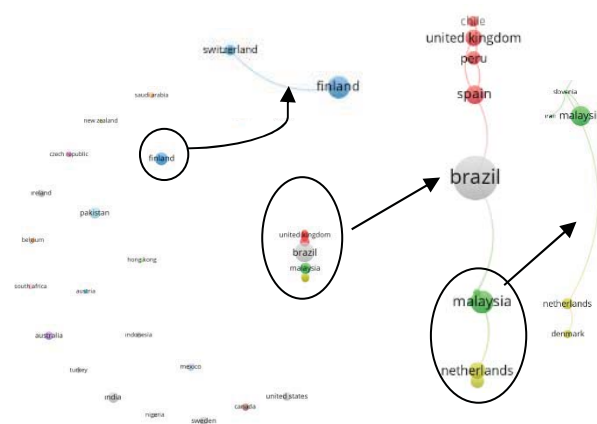


Fig. 2. The most active countries of country-level citation analysis.

B. Bibliographic Coupling Analysis

Bibliographic coupling is a measure of similarity to establish a similarity relationship between publications. It occurs when two studies refer to a common third work in their bibliographies. For 33 countries in the database, the result of bibliometric coupling with VOSviewer can be seen in Fig. 3, the size of the nodes indicates the number of citations. According to the total link strength of the bibliographic coupling analysis, the five most active countries are Brazil, Germany, Finland, Malaysia and Pakistan, respectively. Brazil is at the center of the graph, as in citation analysis. Brazil has two links and has a special location that merges two biggest clusters. Excluding Brazil, there are a total of six countries with more than one link. It can be seen that these countries are the United Kingdom, Spain, Slovenia, Netherlands, Finland and Peru. It was observed that five of the six most central countries after Brazil were in Europe.

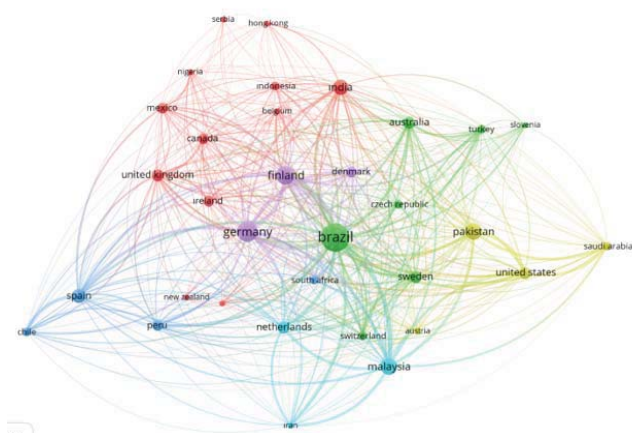


Fig. 3. The most active countries of bibliographic coupling analysis.

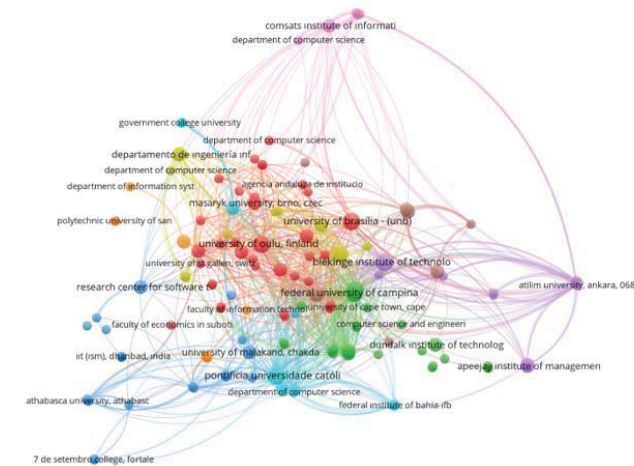


Fig. 4. The most active institutions of bibliographic coupling analysis.

Fig. 4 shows the bibliographic mapping analysis results performed for the 133 institutions in the dataset at the institution level of the authors. According to the total link strength in this analysis with Gephi, the five most active institutions are Pontifical Catholic U. of Rio Grande do Sul from Brazil, Blekinge Institute of Technology from Sweden and University of Sao Paulo from Brazil, Pontifical Catholic University of Peru from Peru and COMSATS Institute of Inf. Technology from Pakistan.

IV. CONCLUSIONS AND FUTURE WORKS

Agile software development has gained a great deal of interest worldwide and has increased its reputation and effective role in the industry after its first announcement, and even after two decades, still growing steadily. In this study, the literature reviews on agile software development have been selected in the concept of systematic literature review and analyzed with social network analysis. In the study, year span of 2013-2018 was selected to observe the trend in recent years. In the selected publications, social network analysis was applied based on the citation due to its academic reputation, and citation analysis and bibliographic coupling analyzes were applied. Relations based on institution and country were mapped and detailed information about the most active roles in terms of relationships was presented within the scope of the study.

The increasing trend of agile software development literature reviews has been seen in recent years and with the help of social network analysis, it is aimed to be a guide in the literature reviews that researchers will make prior to their studies. With the age of communication removing the boundaries between researchers, it is easier for researchers to conduct collaborative work. Countries and institutions focusing on agile software development can also be identified thanks to this type of work, and can be a guide for the professional lives of individuals.

As future work, not only all-time agile software development literature reviews, but also specific empirical topics of agile software development will be analyzed with social network analysis. Hidden patterns on articles' relationships will provide researchers a broad perspective, save a great deal of time and guide their academic studies.

REFERENCES

- [1] B. Kitchenham, "Procedures for performing systematic reviews," Keele, UK, Keele University, vol. 33, no. 2004, pp. 1-26, 2004.
- [2] B. Kitchenham, S. Charters, D. Budgen, P. Brereton, and M. Turner, "Guidelines for performing systematic literature reviews in software engineering," Technical report, Ver. 2.3 EBSE Technical Report. EBSE2007.
- [3] V. Tunalı, Sosyal Ağ Analizine Giriş, Ankara: Nobel Yayıncılık, 2016.
- [4] P. Chinowsky and J. E. Taylor, "Networks in engineering: an emerging approach to project organization studies," Engineering Project Organization Journal, vol. 2, no. 1-2, pp. 15-26, 2012.
- [5] E. Bayram, "Çevik Yazılım Geliştirme Üzerine Bir Sistematik Literatür Taraması: Sosyal Ağ Analizi İle Bir Üçüncül Çalışma," Yüksek Lisans Tezi Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Marmara Üniversitesi, 2020.
- [6] B. Kitchenham et al., "Systematic literature reviews in software engineering—a tertiary study," Information and software technology, vol. 52, no. 8, pp. 792-805, 2010.
- [7] E. Matters, "The Database of Abstracts of reviews of Effects (DARE)," United Kingdom: The University of York, 2002.
- [8] N. J. Van Eck and L. Waltman, "Visualizing bibliometric networks," in Measuring scholarly impact: Springer, 2014, pp. 285-320.
- [9] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: an open source software for exploring and manipulating networks," in Third international AAAI conference on weblogs and social media, 2009.
- [10] (2020). Gephi - The Open Graph Viz Platform. Available: <https://gephi.org/>
- [11] A. Pilkington and J. Meredith, "The evolution of the intellectual structure of operations management—1980–2006: A citation/co-citation analysis," Journal of operations management, vol. 27, no. 3, pp. 185-202, 2009.
- [12] M. M. Kessler, "Bibliographic coupling between scientific papers," American documentation, vol. 14, no. 1, pp. 10-25, 1963.