A bibliometric analysis of research status on copula from 2000 to 2019

Lukun Zheng*[1] and Richard, Adjei-Boateng †[2]

[1]Department of Mathematics, Western Kentucky University, Bowling Green, KY 42101, USA

Abstract

Copula(s), together with its applications, have received world-wide attention. Research about copula have flourished in recent years. This paper aims to present a detailed bibliometric analysis of the research status in the domain of copula, using all copula-related publications obtained from the Web of Science database from the years 2000 to 2019 as the research sample. General characteristics about document type, language, year of publication, country/territory of publication, keywords, institution and authors are presented. Collaboration and dominance situations among authors, institutes, and countries are also studied. Citation analysis are conducted to help researchers quickly understand the copula research area and its development trajectory. In addition, we proposed and concluded the most concerned topics using visualization networks and clustering analysis based on high-frequency keywords. The results of this paper will help researchers grasp the current research status in copula studies and provide a supporting role for its future development.

1 Introduction

A copula is a multivariate function of the marginal distributions which restores the joint distribution among random variables. It is used to describe the dependence between random variables, which is critical in multivariate studies. The idea of a copula can be dated back to the 19th

*lukun.zheng@wku.edu
†richard.adjei-boateng053@topper.wku.edu

century in modeling multivariate non-Gaussian distributions. Modern theories and applications about copula was introduced by Sklar in 1959, which states that an n-dimensional joint distribution can be decomposed into its n univariate marginal distributions and an n-dimensional copula [67].

Theorem 1 (Sklar, 1959)**.** *Suppose that H is a distribution function on R^n with marginal distributions $F_1,\ \ldots,\ F_n$, then there is a copula C such that*

$$H(x_1,\ldots,x_k) = P\left[F_1(X_1) \leq F(x_1),\ldots,F_n(X_n) \leq F_n(x_n)\right] = C\left[F_1(x_1),\ldots,F_k(x_n)\right] \qquad (1)$$

If H is continuous, then C is unique and is given by

$$C(u_1,\ldots,u_n) = H\left[F_1^{-1}(u_1),\ldots,F_n^{-1}(u_n)\right]$$

for $\mathbf{u} = (u_1,\ldots,u_n)^T \in [0,1]^n$ where $\mathbf{x}^T$ is the transpose of the vector of $\mathbf{x}$ and $F_i^{-1}(u_i) = \inf\{x : F_i(x) \geq u\},\ i = 1,\ldots,n$. Conversely, if C is a copula on $[0,1]^n$ and $F_1,\ldots,F_n$ are distribution functions on R, the the distribution function defined in (1) is a distribution function on R^n with marginal distributions $F_1,\ \ldots,\ F_n$.

Copula became increasing popular at the end of 1990s. Two books ([68] and [69]) about copula were published and became the standard references in literature. Copula modeling has been used in various applied fields, especially in finance, see [66, 59, 60, 61, 62, 63, 64, 65]. The advent of copulas in finance originated numbers of investigations, and especially, applications of copulas. For instances, in hydrology, Favre, et al. [58] used copula approach to study the peak flows from the watershed of Peribonka in Quebec, Canada and the dependence structure among peak flows and volumes in the watershed of the Rimouski River in Quebec, Canada. . Chebana and Ouarda [57] used copula to specify a joint distribution for the multivariate quantiles of peak flow and volume. Kang et al. [56] apply a time varying copula as to model flood peak and 7-day flood volume at e Yichang station in the Yangtze River, China. In climate and weather research, Schoelzel and

Friederichs [55] provides an brief overview of copulas for application in meteorology and climate research with a experimental study based on observations of daily precipitation and temperature at two stations in German. Zhou et al. [54] developed and evaluated a coupled dynamical-copula downscaling model using data on historical climatology of the Canadian Prairies and generated future climate projections over three time slices. In civil engineering, Kilgore and Thompso [53] introduced a copula strategy to estimate joint probabilities of design coincident flows at stream confluences often encountered at highway drainage structures. Zhang et al. [52] proposed a copula-based feature sensitive indicator for the bridge structural health monitoring to remove the operational and environmental conditions.

Copula modeling develops rapidly in both theory and applications. So, it's necessary for researchers to review and summaries the development of copula. Some previous studies focused on the review of copula in detailed areas, e.g., review on copula in economics and finance [51, 50] , copula in hydrology and water resources [49], copula in engineering [48], etc. However, few studies explored the issues of copula in a comprehensive and quantitative manner. In contrast to previous studies, we try to have an overall review of the research status in copula modeling by bibliometric methods, which may present a relatively systematic and quantitative review for current and future researchers. The paper is organized as follows. In the method section, we described the retrieve of the sampling articles, the method of collaboration and dominance analysis, the method of citation analysis, and keywords co-occurrence analysis. In the result and discussion section, we mainly presents five parts, i.e., the basic descriptive characteristics of studies in copula modeling, the collaboration and dominance analysis, the citation and co-citation analysis, and the trends and hot spots in copula modeling. We conclude in the last section.

2 Methods

Bibliometrics is a well-established method in the field of information and intelligence science. It utilizes technologies in different areas such as computing, social networks, and statistics to analyze

the bibliographic and content information of the scientific activities in a certain field [47, 46]. Bibliometric analysis can be used to provide a comprehensive overview and the development evaluation for the given field, especially for the interdisciplinary and frontier fields. Bibliometrics has been used widely in many scientific fields, e.g., in big data [44, 45] , in deep learning [42, 43], in education [41] and in ecology [39, 40], etc. To the best of the author's knowledge, bibliometrics has never been utilized to study copula modeling. One of the contributions of this paper is to cover this gap.

2.1 Data Collection

The data used in this paper was retrieved from the Web of Science, which gives access to multiple databases that reference cross-disciplinary research for in-depth exploration of specialized sub-fields. Furthermore, in this study, the field tags TS(=Topic), which contains titles, abstracts, and keywords, were used to create the query. A research sample of 6519 bibliographic records published from 2000 to 2019 were collected from the Web of Science using the search strategy "TOPIC: (copula) " on May 19th, 2019 (the data of 2019 are incomplete). The data set contains all available factors including the title, author(s), keywords, abstract, research areas, citation information, etc. Note that the United Kingdom (UK) in this study refers to England, Scotland, Wales and Northern Ireland. China refers only to Mainland China.

2.2 Collaboration and Dominance Analysis

Collaboration in research can take many forms of activity ranging from offering general advice and opinion to active and sustained participation and contribution of physical and intellectual resources [38]. The collaboration degree is a measure of the scientific research's collaborative relation at the levels of authors, institutions, and countries/regions. The degree of collaboration varies from one discipline to another. In general, it is high in the intensely collaborative scientific and technical fields, but low in the humanities in which the lonely scholar produces much of the scholarly literature. In this study, the calculation formula at these three levels are given

respectively as follows [37]:

$$C_{AU} = \frac{\sum_{i=1}^{N} \alpha_i}{N},\ C_{IN} = \frac{\sum_{i=1}^{N} \beta_i}{N},\ C_{CR} = \frac{\sum_{i=1}^{N} \gamma_i}{N}, \tag{2}$$

where: $\alpha_i,\ \beta_i,\ \gamma_i$ are the number of authors, institutions and countries/regions for each research item; N is the total number of research items; and $C_{AU},\ C_{IN}$, and C_{CR} are the collaboration degree at the levels of authors, institutions, and countries/regions, respectively.

Dominance is described as "a relational, behavioral, and interaction state that reflects the actual achievement of influence or control over another via communicative actions" [36]. In collaborative research, Dominance is seen as a property of an interpersonal relationship rather than of an individual, and emphasis is on the social skills and communication practices that contribute to dominance. The dominance is usually reflected in the order of authors appeared in a publication in bibliometrics. In this study, the Dominance Factor is a ratio indicating the fraction of multi-authored articles in which a scholar appears as first author.

2.3 Citation Analysis

Citation analysis is a way of measuring the relative importance or impact of f an article, an author, or an institution by counting the number of times that article, author, or institution has been cited by other works. There are two major approaches in citation analysis: bibliography coupling and co-citation analysis. At the document level, two documents are said to be bibliographically coupled if both of them share at least one reference in common [35]. Their bibliographic coupling frequency is then the number of references they have in common . Two documents are said to be co-cited if they both are cited in anther document [33, 34, **?**]. The co-citation frequency is defined as the frequency that they are cited together. Besides document bibliographic couplings (or co-citations), one may also study author bibliographic couplings (or co-citations) [32, 31, 28, 29, 30].

In this paper, we will mainly use bibliography coupling networks to study the relationships among documents in our analysis. Social networks area collection of points and edges, where

points represent the various social documents and edges represent a variety of social relationships between them [27]. A formal description of social networks usually includes a social network graph and a matrix of social relations.

2.4 Keywords Co-occurrence Analysis

Keywords are an important part of a scientific document that reflects its content. Using co-occurrence analysis of keywords, we can understand patterns and trends in a specific discipline by measuring the association strengths of terms representative of relevant publications produced in this area [26]. Two keywords co-occurring within the same paper are an indication of a link between the topics to which they refer. More such occurrences indicate stronger relation between these referred topics. keywords co-occurrence analysis helps to determine the relation between and visualizes the intellectual structure of the various research themes among disciplines represented by the literature collection. In this paper, we apply high frequency keyword co-occurrence cluster analysis methods to explore research trends and hot spots in the copula modeling field.

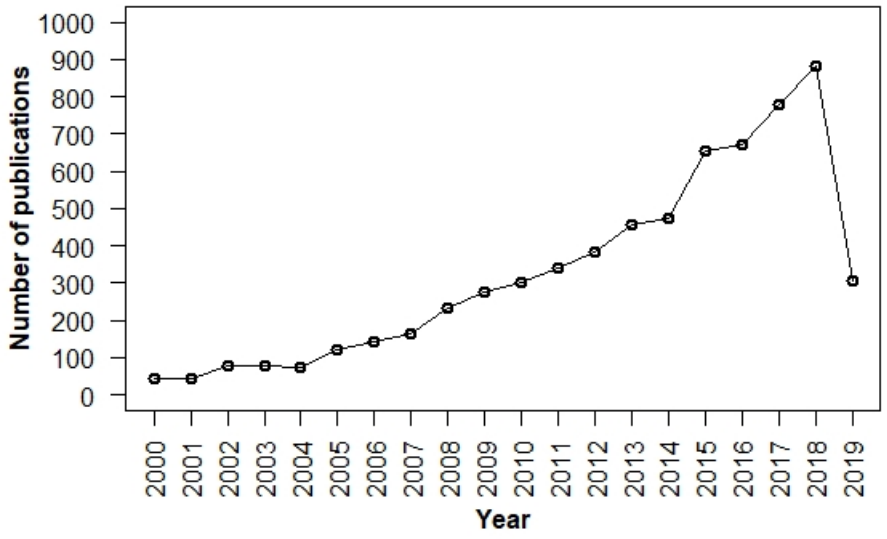

Figure 1: Distribution of publishing articles in Copula modeling by year

3 Results and discussion

3.1 Basic characteristics

The yearly publication distribution are demonstrated in Figure 1, which shows a steady increasing trend in copula study from 2000 to 2019. The annual number of publication has grown about twenty times from 44 in 2000 to 882 in 2018 (the number of the publications in 2019 is incomplete). In particular, the number of publications experienced the most rapid increase of nearly 200 publications in a single year from 2014 to 2015.

The top 5 languages used by these documents are English, Spanish, French, German, and Russian. English is the dominant language with 6398 records, accounting for 98.14% of the total publications. It is followed by Spanish (25; 0.38%), French (21; 0.32%), German (18; 0.28%), and Russian (13, 0.20%). Analysis on the types of publications is demonstrated in table 1. We retrieved a sum of 11 different document types in this research. From table 1 we observe that articles have the highest frequency in terms of the most published document type which is 6108 representing 93.7 % of the total publications. This was followed by Proceedings paper, having a total count of 242 which represents 3.71% of the total publications. Book Review came in the 3rd place with 55 counts representing 0.84% of the total publications.

Table 1 Documents type distribution of Copula related Literature

	Documnet Type	Count	Percent
1	ARTICLE	6108	93.70
2	PROCEEDINGS PAPER	242	3.71
3	REVIEW	55	0.84
4	EDITORIAL MATERIAL	53	0.81
5	MEETING ABSTRACT	25	0.38
6	BOOK REVIEW	16	0.25
7	CORRECTION	8	0.12
8	BOOK CHAPTER	5	0.08
9	LETTER	4	0.06
10	REVIEW, BOOK CHAPTER	2	0.03
11	REPRINT	1	0.02

3.2 Countries and regions

We gathered an overall 6469 documents out of the 6519 from January 2000 to March 19th 2019 and dropped 50 documents that had no author information and could not be traced to any country or region. Therefore, the obtainable documents we used, based on copula research, for the purpose of country and regional analysis provided affiliations for at least one author. Table 2 shows the top 20 countries or regions and their statistical information with respect to their total publications.

From table 2 we observe that the USA is the most productive country with 1112 documents amounting to 17.19% of the total retrieved documents. The second most productive country is China(870,13.35%), followed by Germany(496, 7.61%). The USA again tops in terms of the h-index. Having an h-index of 58 reflects the level of academic influence the USA has on the Copula related research field. China on the other hand has an h-index of 34 and is ranked 4th although she is the 2nd in terms of total publications on copula. This indicates that China has a relatively smaller influence on the copula research related field compared to Canada(2nd) although it takes

an important role in the field.

Table 2 Top 20 most productive countries and regions in research on Copula research

	Country/Region	TP	(%)	H Index (rank)
1	USA	1112	17.19	58 (1)
2	CHINA	870	13.45	34 (4)
3	GERMANY	496	7.67	38 (3)
4	CANADA	400	6.18	41 (2)
5	UNITED KINGDOM	330	5.10	31 (7)
6	ITALY	291	4.50	33 (6)
7	FRANCE	271	4.19	31 (7)
8	SPAIN	245	3.79	27 (10)
9	AUSTRALIA	212	3.28	28 (9)
10	BELGIUM	181	2.80	27 (10)
11	IRAN	161	2.49	13 (17)
12	SWITZERLAND	145	2.24	34 (4)
13	TAIWAN	134	2.07	19 (14)
14	BRAZIL	129	1.99	12 (18)
15	AUSTRIA	111	1.72	24 (12)
16	INDIA	108	1.67	17 (15)
17	NETHERLANDS	94	1.45	20 (13)
18	KOREA	92	1.42	11 (19)
19	POLAND	89	1.38	10 (20)
20	JAPAN	80	1.24	15 (16)

The annual publication output for the top 5 most productive countries and regions from the year 2000 to 2019 is displayed in Figure **??**. We find that there is an overall rising trend in the number of publications across these years. The USA occupied a leading position from the year 2000 until the year 2016. During 2016, USA and China produced similar quantities of publications. China has seen a very sharp increase in the number of publications in the years 2017 and 2018, making her the country with the most publications in those years. During these years(2000-2019), the focus of research was mainly on Copula itself and the applications of copula to different distributions and their extensions. Figure **??** also shows that the remaining three countries have been very close in terms of the quantity of copula related publications produced yearly.

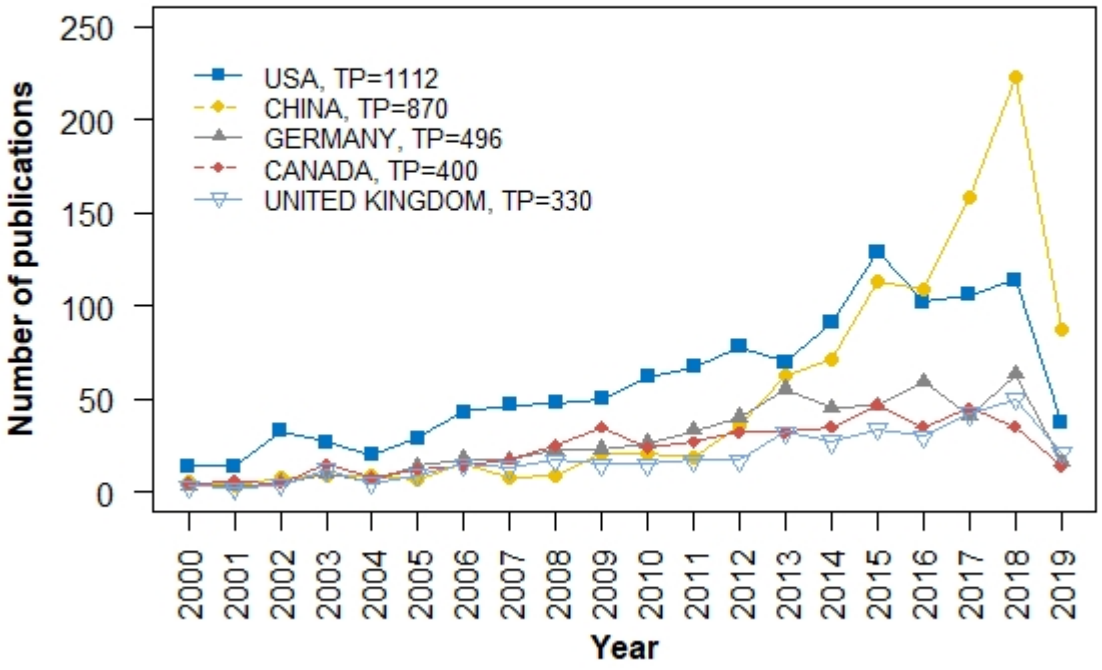

Figure 2: The Yearly output performance of the top 7 countries (TP=total publication).

3.3 Subject Categories

In total, 9519 publications were from 117 ISI subject categories. Out of the 117 categories, 53 (45.3%) subject categories contain less than 10 articles. 2680 (28.2%) items were published in the category of mathematics, and 1303 (13.7%) items were published in the category of business. Table 3 shows the top 20 subject categories with the most publications including the total times being cited, the average citations per publication, and the h-indices. Mathematics has the highest h-index of 60, which shows that it not only held the first place in the quantity of publications, but also had the greatest academic influence in copula research. In addition, Table 5 shows that copula research can be viewed as an interdisciplinary field, with Mathematics, Business, Engineering, and so on being the top subject categories with the most productive areas in copula research.

Table 3 Top 20 ISI subject categories with the most publications

	Subject Category	TP	TC	AC	H-Index(R)
1	MATHEMATICS	2680	30369	11.33	69 (1)
2	BUSINESS \& ECONOMICS	1303	16418	12.60	55 (2)
3	ENGINEERING	996	13165	13.22	48 (5)
4	COMPUTER SCIENCE	693	7977	11.51	41 (6)
5	WATER RESOURCES	578	11752	20.33	51 (3)
6	MATHEMATICAL METHODS IN SOCIAL SCIENCES	475	7239	15.24	39(7)
7	ENVIRONMENTAL SCIENCES \& ECOLOGY	411	8996	21.89	51(3)
8	LINGUISTICS	330	2523	7.65	27(11)
9	GEOLOGY	244	4409	18.07	33(8)
10	OPERATIONS RESEARCH \& MANAGEMENT SCIENCE	236	2503	10.61	25(14)
11	MATHEMATICAL \& COMPUTATIONAL BIOLOGY	210	2153	10.25	26(13)
12	METEOROLOGY \& ATMOSPHERIC SCIENCES	158	2375	15.03	25(14)
13	ENTOMOLOGY	141	2177	15.44	25(14)
14	ZOOLOGY	119	2189	18.39	27(11)
15	SCIENCE \& TECHNOLOGY - OTHER TOPICS	114	779	6.83	17(18)
16	ENERGY \& FUELS	107	1198	11.20	20(17)
17	LIFE SCIENCES \& BIOMEDICINE - OTHER TOPICS	105	3947	37.59	29(9)
18	PHYSICS	101	1188	11.76	17(18)
19	MARINE \& FRESHWATER BIOLOGY	93	2604	28.00	29(9)
20	ENVIRONMENTAL \& OCCUPATIONAL HEALTH	82	567	6.91	12(20)

3.4 Well-known institutes, and journals

3.4.1 Institutes

A total of 3031 different research institutes were collected from these 9519 documents in this study. Of these 3031 research institutes, 2433 contributed 1-5 publications each, 259 institutes 6-10 papers each, 183 institutes 11-20 publications each, 122 institutes 21-50 publications each, 25 institutes 51-100 publications each, and 5 institutes 108-188 papers each. Therefore, the distribution by research institutes are highly right-skewed. The top 20 most productive research institutes based on the number of total publications are listed in Table (4. Among the top 20 institutes, three are from the USA and four are from China, which again shows their dominant positions in the copula research field. TEXAS AANDM UNIV(USA) ranked at the top in terms of both total publication (188) and the h-index (25), which indicated its leading position on the copula research field. UNIV ALMERIA(Spain) was the second in terms of the total publications

Table 4 Top 20 most productive institutes in copula research

	Institution	TP	TC	AC.	H-Index(R)
1	TEXAS AANDM UNIV(USA)	188	2152	11.45	25 (1)
2	UNIV ALMERIA(Spain)	129	975	7.56	18 (10)
3	UNIV GHENT(Belgium)	121	1154	9.54	21 (4)
4	TECH UNIV MUNICH(Germany)	120	2102	17.52	20 (6)
5	SLOVAK UNIV TECHNOL BRATISLAVA(Slovakia)	108	1095	10.14	20 (6)
6	WUHAN UNIV(China)	98	1013	10.34	19 (9)
7	UNIV LAVAL(Canada)	94	3076	32.72	25 (1)
8	SUN YAT SEN UNIV(China)	81	597	7.37	15 (11)
9	BEIJING NORMAL UNIV(China)	74	227	3.07	10 (17)
10	JOHANNES KEPLER UNIV LINZ(Austria)	67	1203	17.96	22 (3)
11	UNIV WATERLOO(Canada)	65	530	8.15	12 (16)
12	CATHOLIC UNIV LOUVAIN(Belgium)	62	651	10.50	14 (12)
13	UNIV WISCONSIN(USA)	62	643	10.37	14 (12)
14	UNIV MINNESOTA(USA)	58	555	9.57	13 (14)
15	HOHAI UNIV(China)	57	194	3.40	8 (19)
16	UNIV SALENTO(Italy)	57	1307	22.93	20 (6)
17	UNIV ZURICH(Switzerland)	56	1622	28.96	21 (4)
18	KATHOLIEKE UNIV LEUVEN(Belgium)	55	432	7.85	10 (17)
19	MCGILL UNIV(Canada)	53	584	11.02	13 (14)
20	UNIV SAO PAULO(Brazil)	52	244	4.69	8 (19)

(129). However, its h-index (18) just ranked the 10th, which means that its academic influence in this field is relatively low.

We provide the time-trend analysis of the top 5 most productive institutes in Figure **??**. It shows one institutes each in the USA, Spain, Belgium, Germany, and Slovakia. The evolution trend can be divided into two stages: from 2000 to 2009, Ghent University, University of Almeria, and Slovak University of Technology in Bratislava dominated the study of copula, even though the general study is volatile; from 2009-2019, nearly all these institutes showed an increasing but fluctuating trend, especially for Texas A&M University which experienced the most rapid growth, from the lowest rank of these 5 institutes in 2009 to the the highest rank among them in 2018.

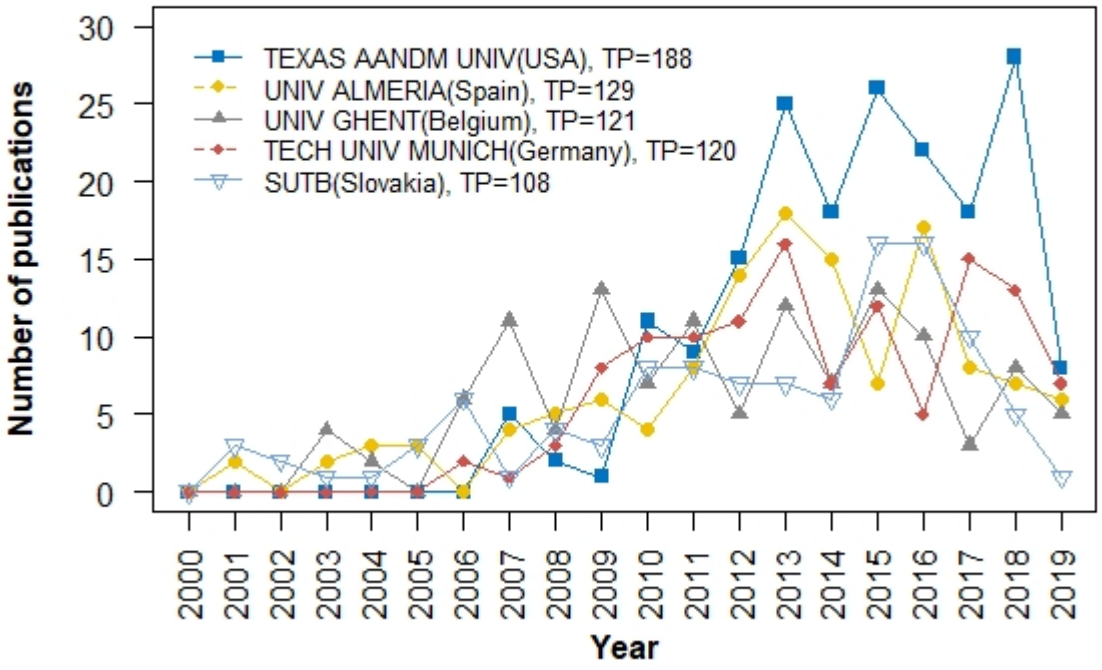

Figure 3: The Yearly output performance of the top 5 research institutes (TP=total publication, SUTB=SLOVAK UNIV TECHNOL BRATISLAVA).

3.4.2 Journals

The 6519 documents analyzed in this paper was affiliated with 1455 different journals. Table **??** shows the first 20 journals, their respective total number of publications with respect to copula research papers, the number of times cited, the average number of citation per document, the H-index and their rank by h- index. We observe that the peer-reviewed open access journal on water science and technology including the ecology and management of water resources called "Water", topped the list having the highest number of published materials(207), times being cited(4082) and the relative influence in the area of copula research(h-index of 33). "Water" was followed by the journal of multivariate analysis having a total count of 191, total citation of 3454, and an h-index of 30. Insurance mathematics and economics(176) came out to be the

third with the highest number of publication but the fourth most influential journal in the area of copula. Likewise, in terms of the total number of publications, Journals like the Communication in Statistics theory and methods(124), Fuzzy Sets and Systems(117), Computational Statistics and Data Analysis(99), Journal of Hydrology(98), Statistics and Probability letters(94), Water Resource Research(69) and Quantitative Finance(68) constitutes the remaining part of the top ten.

3.5 Collaboration and dominance analysis

3.5.1 Collaboration among authors, institutes, and countries

Excluding 50 anonymous documents due to lack of author address information, the remaining 6469 documents were used for collaboration and dominance analysis. The time-trends of the collaboration degree at the author, institute, and country levels are demonstrated in Figure **??**. The three levels of collaboration degree increase over time in general, which indicates that the collaboration among them got strengthened over time. Among these three levels, the author's level shows the fastest growing pattern, while the country's level was experiencing the slowest growth. The relative higher collaboration degree at the author's level suggests that the authors tended to collaborate more with those within the same institutes and within the same country. For instance, in 2012, the author collaboration degree was 2.51, the institute collaboration degree was 1.77, and the country collaboration degree was 1.37. It indicated that each document has 2.51 authors on average, while only has 1.77 institutes and 1.37 countries on average in 2012.

Table 5 The top 20 most productive journals in copula research. TP=total publications,TC=total number of times being cited, AC=average number of times being cited per publication.

	Journal	TP	TC	AC	H-Index(Rank)
1	WATER	207	4082	19.72	33 (1)
2	JOURNAL OF MULTIVARIATE ANALYSIS	191	3454	18.08	30 (2)
3	INSURANCE MATHEMATICS \ \& ECONOMICS	176	3093	17.57	25(4)
4	COMMUNICATIONS IN STATISTICS-THEORY AND METHODS	124	539	4.35	13 (15)
5	FUZZY SETS AND SYSTEMS	117	1015	8.68	17 (9)
6	COMPUTATIONAL STATISTICS \ \& DATA ANALYSIS	99	1518	15.33	20 (8)
7	JOURNAL OF HYDROLOGY	98	2050	20.92	24 (5)
8	STATISTICS \ \& PROBABILITY LETTERS	94	917	9.76	16 (10)
9	WATER RESOURCES RESEARCH	69	2349	34.04	27 (3)
10	QUANTITATIVE FINANCE	68	901	13.25	13 (15)
11	JOURNAL OF HYDROLOGIC ENGINEERING	63	2248	35.68	21 (7)
12	STOCHASTIC ENVIRONMENTAL RESEARCH AND RISK ASSESSMENT	60	810	13.50	16 (10)
13	KYBERNETIKA	55	690	12.55	14 (13)
14	STATISTICS IN MEDICINE	51	363	7.12	10 (18)
15	COMMUNICATIONS IN STATISTICS-SIMULATION AND COMPUTATION	49	199	4.06	8 (19)
16	JOURNAL OF BANKING \ \& FINANCE	49	1617	33.00	24 (5)
17	JOURNAL OF STATISTICAL PLANNING AND INFERENCE	49	702	14.33	13 (15)
18	ENERGY ECONOMICS	49	738	15.06	14 (13)
19	DEPENDENCE MODELING	48	48	1.00	3 (20)
20	INFORMATION SCIENCES	43	540	12.56	15 (12)

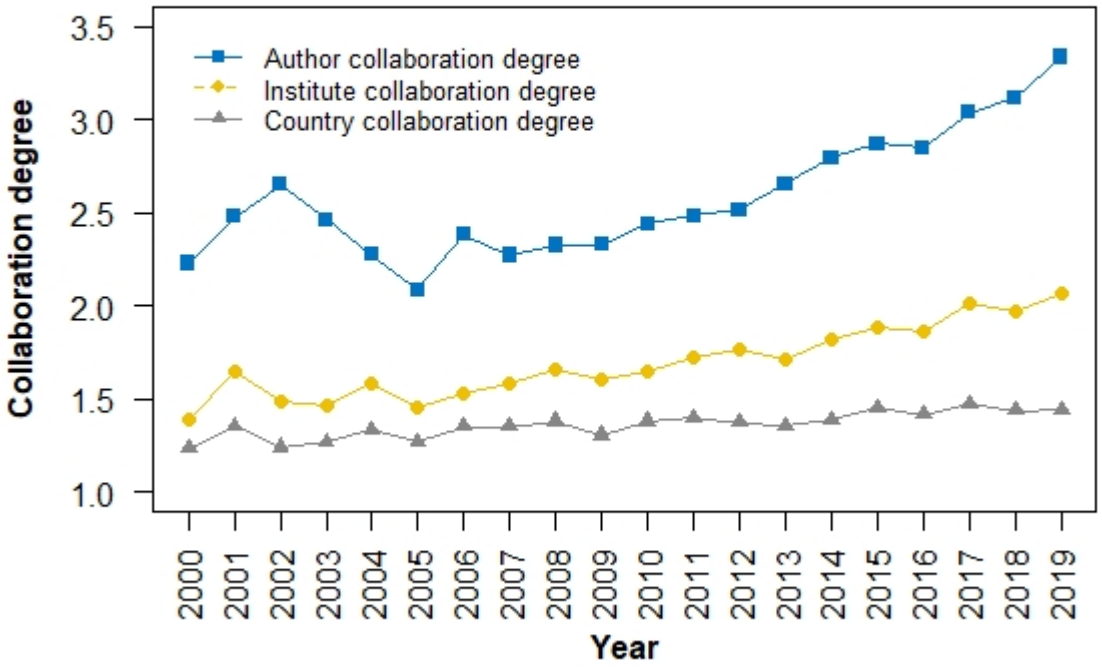

Figure 4: Collaboration degrees at author, institute, and country levels in copula research from 2000 to 2019.

Let's now delve further into the author collaboration analysis by focusing on the top 20 most productive authors, since it experienced a much faster growth. We collected 9060 different authors from 6469 documents in this study. Of these 9060 different authors, 6086 (67.17%) contributed only 1 publications each, 2379 (26.26%) authors 2-5 publications each, and 595 (6.57%) authors 6 or more publications each. Table 6 shows the statistical information of the top 20 authors with respect to the number of total publications, the total number of times being cited, the average number of times being cited per publication, and the h-indices together with the corresponding ranks. DURANTE F ranked first in terms of the total publications (95). SINGH VP took the leading position in terms of h-index (25). GENEST C ranked first in both total citations (2855) and average citations (52.87). These shows their importance and authority of these authors in the field of copula research.

Table 6 The top 20 most productive authors in copula research. TP=total publications,TC=total number of times being cited, AC=average number of times being cited per publication.

	Author	TP	TC	AC	H-Index(R)
1	DURANTE F	95	1517	15.97	22 (4)
2	MESIAR R	80	1268	15.85	23 (3)
3	SINGH VP	77	2097	27.23	25 (1)
4	WANG Y	66	623	9.44	14 (8)
5	DE BAETS B	60	1222	20.37	21 (5)
6	FERNANDEZ SANCHEZ J	58	319	5.50	11 (13)
7	UBEDA FLORES M	57	598	10.49	13 (10)
8	GENEST C	54	2855	52.87	24 (2)
9	LI H	52	519	9.98	11 (13)
10	LI X	51	255	5.00	10 (16)
11	CZADO C	50	1443	28.86	16 (6)
12	ZHANG Y	45	161	3.58	7 (20)
13	CHEN X	42	875	20.83	14 (8)
14	DE MEYER H	41	677	16.51	16 (6)
15	LI J	41	650	15.85	13 (10)
16	ZHANG X	39	270	6.92	10 (16)
17	ZHANG L	38	855	22.50	10 (16)
18	ZHANG J	36	245	6.81	8 (19)
19	JOE H	32	795	24.84	11 (13)
20	SEGERS J	31	496	16.00	13 (10)

3.5.2 Dominance analysis

Table 7 shows the results of the dominance analysis for the top 10 most dominant authors in copula research. The dominance factor is the fraction of multi-authored (M-A) articles in which a scholar appears as the first author(F-A). Durante, F is the most dominant author in this sense. He first-authored 74 among his 88 multi-authored publications, yielding a dominance factor of 0.84. In addition, Durante, F is also the most productive researcher from Table 6. Rank 1 in terms of both total publication and dominance factor tells the contribution of Durante, F in this field. The second most dominant researcher is Genest, C who first-authored 39 among his 53 multi-authored publications.

Table 7 The top 10 most dominant authors in copula research. DF=Dominance Factor,TP=total publications, S-A=Single-Authored, M-A=Multi-Authored, F-A=First-Authored.

	Author	DF	TP	S-A	M-A	F-A	TP Rank	DF Rank
1	DURANTE F	0.84	95	7	88	74.00	1	1
2	GENEST C	0.74	54	1	53	39.00	4	2
3	LI C	0.50	30	0	30	15.00	10	3
4	ZHANG L	0.39	38	5	33	13.00	8	4
5	ZHANG J	0.37	36	1	35	13.00	9	5
6	DE BAETS B	0.37	60	0	60	22.00	2	6
7	ZHANG Y	0.36	45	6	39	14.00	6	7
8	FERNANDEZ SANCHEZ J	0.33	58	0	58	19.00	3	8
9	LI H	0.32	52	15	37	12.00	5	9
10	LI J	0.31	41	9	32	10.00	7	10

3.6 Citation analysis

Among these 6519 publications, there are 1498 publications which have never been cited before. However, 1140 of them were published in 2016 or later. There are 1301 publications being cited once or twice, among which 756 were published in 2016 or later. 1017 publications are cited

between 3 and 5 times. 904 publications are cited between 6 and 10 times. 1175 publications are cited between 11 and 30 times. The remaining 624 publications are cited over 30 times.

Table 8 presents the top 10 most cited publications [17, 18, 58, 19, 20, 21, 22, 23, 24, 25]. Among these 10 publications, 4 were published in economics-related journals, 3 were published in statistics-related journals, and 3 were published in hydrology-related journals. Furthermore, a paper by PATTON AJ (2006) [24] ranked first in total citations. The researcher studied the asymmetric dependence between exchange rates using an extension of the copula theory allowing for conditioning variables. They found that the mark-dollar and yen-dollar exchange rates are more correlated when they are depreciating against the dollar than when they are appreciating.

Bibliographic coupling is applied to construct a article coupling network to establish a similarity relationship among the 10 articles in this field from 2000 to 2019. In Figure **??**, the size of the circle represents the coupling degree among these articles and the thickness of the edges between pairs of articles represent the strength of coupling of the corresponding pairs of articles. We found that the size of the vertex representing Patton, AJ (2012) was the largest, indicating that the coupling degree of it among these selected articles. We also found that the Patton, AJ (2012) and Fan,Y (2014) are bibliographically coupled the most, since the edge between them is the thickest one among all the edges in the network.

Table 8 The top 10 most cited publications in copula research. TC=total number of times being cited, PY=publication year.

First Author	Title	TC	SO	PY
PATTON AJ	MODELLING ASYMMETRIC EXCHANGE RATE DEPENDENCE	725	INTERNATIONAL ECONOMIC REVIEW	2006
GENEST C	EVERYTHING YOU ALWAYS WANTED TO KNOW ABOUT COPULA MODELING BUT WERE AFRAID TO ASK	631	JOURNAL OF HYDROLOGIC ENGINEERING	2007
AAS K	PAIR-COPULA CONSTRUCTIONS OF MULTIPLE DEPENDENCE	526	INSURANCE MATHEMATICS \& ECONOMICS	2009
GENEST C	GOODNESS-OF-FIT TESTS FOR COPULAS: A REVIEW AND A POWER STUDY	516	INSURANCE MATHEMATICS \& ECONOMICS	2009
DEMARTA S	THE T COPULA AND RELATED COPULAS	322	INTERNATIONAL STATISTICAL REVIEW	2005
VARIN C	AN OVERVIEW OF COMPOSITE LIKELIHOOD METHODS	313	STATISTICA SINICA	2011
FAVRE AC	MULTIVARIATE HYDROLOGICAL FREQUENCY ANALYSIS USING COPULAS	306	WATER RESOURCES RESEARCH	2004
MCNEIL AJ	MULTIVARIATE ARCHIMEDEAN COPULAS, D-MONOTONE FUNCTIONS AND L(1)-NORM SYMMETRIC DISTRIBUTIONS	288	ANNALS OF STATISTICS	2009
JONDEAU E	THE COPULA-GARCH MODEL OF CONDITIONAL DEPENDENCIES: AN INTERNATIONAL STOCK MARKET APPLICATION	275	JOURNAL OF INTERNATIONAL MONEY AND FINANCE	2006
MISHRA AK	DROUGHT MODELING - A REVIEW	271	JOURNAL OF HYDROLOGY	2011

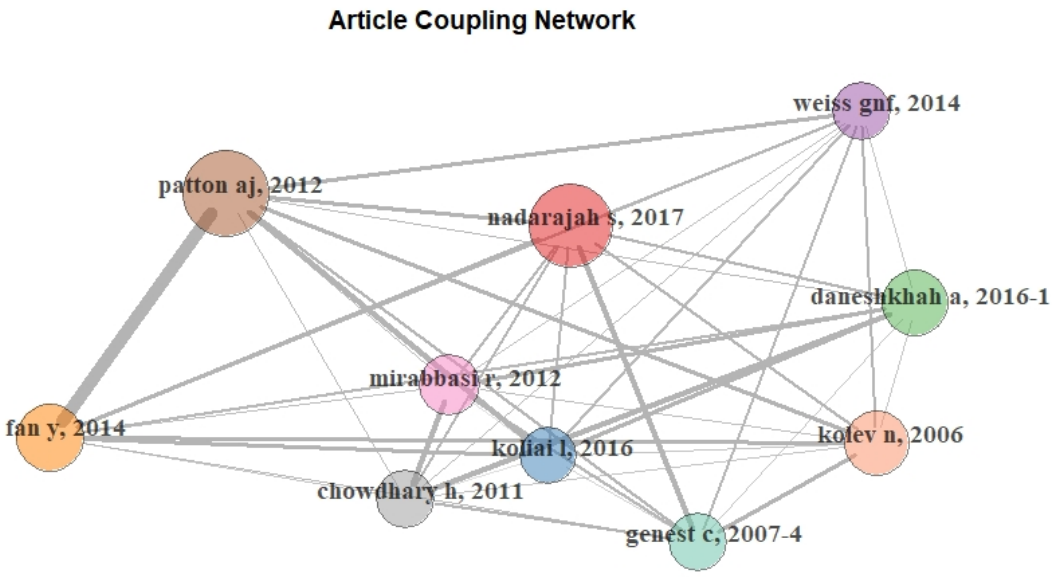

Figure 5: Bibliographic coupling association network of the most frequently cited studies

3.7 Academic hotspots

We use the frequency of keywords in copula publications to study the academic concerns. The 9519 documents had 14263 different keywords, among which 12530 were used once or twice, 1475 were used between 3 and 10 times, 239 were used between 11 and 50 times, and 19 were used 51 or more times. Since there are keywords with difference forms but the exam meanings (e.g. "copula" and "copulas"), we applied a preliminary treatment of the keywords and then constructed a co-occurrence network for the high-frequency keywords (frequency more than 50), see Figure **??**.

From the results on the keywords frequency and co-occurrence analysis, we list the following academic hotspots in copula research:

(1) Families of copulas and their applications: Several families of copulas, such as Gaussian copulas [16], Archimedian copulas [15], and vine copulas [14] have been proposed and studied in literature. A delicate description about families of copula and their applications can be

found in [11, 12, 13].

(2) Using copula to measure dependence: Copulas provides an ideal tool in studying the dependence degree and structure in multivariate modeling. A copula can connect the marginal distributions to restore the joint distribution using a copula function. The main purpose in copula dependence modeling is to choose an appropriate dependence function which captures the dependence nature among the variables of interests. In the literature, individual copulas and mixed copulas have been proposed and used in many applications of interest [12, 10, 9, 8].

(3) Copula modeling in business and economics. In business and economics, we show that copulas can be extensively used to solve many financial problems. For example, copulas have been used to study financial contagion among different markets [4, 3]; many studies uses copula to conduct risk management by computing portfolio's Profit & Loss series and corresponding risk measures curves [2]; researchers utilized the adoption of copula functions to price multivariate contingent claims by embedding the marginal distributions extracted from vertical spreads in the options markets in a multivariate pricing kernel [1].

(4) Copula modeling in hydrology. In hydrology, complex events like floods and storms always appear to be multivariate events that are represented by a joint distribution of several random variables [6, 7]; the design of hydropower dam requires the study of the risks associated with multiple peak discharges at the future dam site [5].

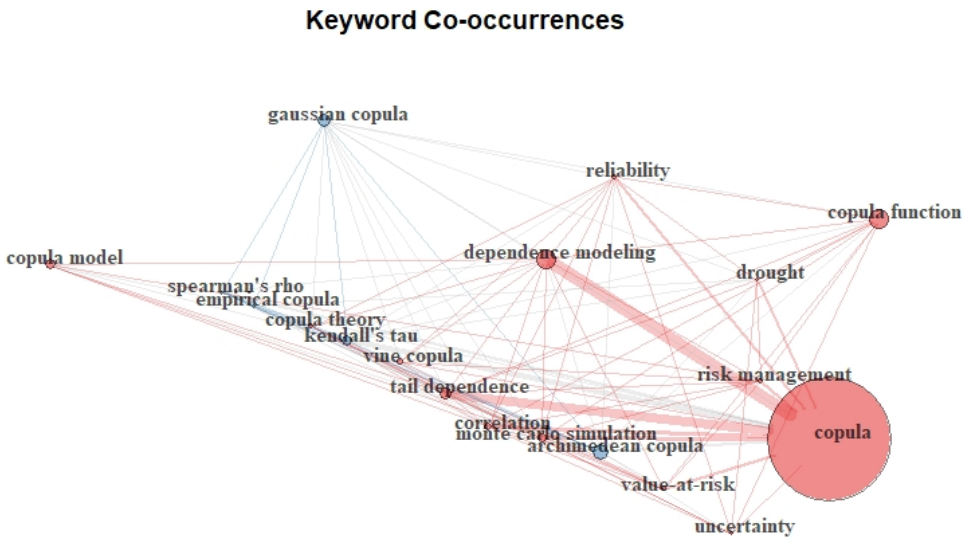

Figure 6: The co-occurrence network of the most frequently used keywords

4 Conclusions

As a bibliometric analysis of copula modeling literature, this paper thoroughly analyzes the research development status from 2000 to 2019 through frequency analysis, collaboration analysis, citation analysis, social network analysis. This study presents a comprehensive description for the first time in copula modeling. It presents a number of interesting observations and involve a few related hot topics during this period of time. It is shown that scientific productions in copula modeling experienced substantial growth in terms of annual number of publication and citations of documents from 2000 to 2019. The major output in copula modeling come from the researchers in USA and China, indicated by the article distribution among countries. English is the dominant language, as high as 98.14%. The overall collaboration degree at the author level shows a upward trend over this period of time, which, in some sense, is due to the fact that copula are used in many interdisciplinary subjects involving mathematics, business & economics,

engineering, hydrology, etc. We also presented the top 20 most productive ISI subject categories, in which "Mathematics" was the most productive ISI subject category followed by "Business & Economics", "Engineering", "Computer Science", and "Water Resources". The top three journals that publishing the most articles are WATER, JOURNAL OF MULTIVARIATE ANALYSIS, and INSURANCE MATHEMATICS, publishing 207, 191, and 176 articles during this period of time, respectively. Bibliographic coupling was used to establish a similarity relationship among the top 10 most cited articles in this field from 2000 to 2019. We found that the size of the vertex representing Patton, AJ (2012) was the largest, indicating that the coupling degree of it among these selected articles. We also found that the Patton, AJ (2012) and Fan,Y (2014) are bibliographically coupled the most, since the edge between them is the thickest one among all the edges in the network. Based on the frequency and co-occurrence analysis of keywords used in the collected research sample, we identified the hot research areas in copula modeling: families of copulas and their applications, using copula to measure dependence, copula modeling in business and economics, and copula modeling in hydrology.

References

[1] Rosenberg, J. V. (2003). Non-parametric pricing of multivariate contingent claims. The Journal of Derivatives, 10(3), 9-26.

[2] Semenov, M., & Smagulov, D. (2019). Copula Models Comparison for Portfolio Risk Assessment. In Global Economics and Management: Transition to Economy 4.0 (pp. 91-102). Springer, Cham.

[3] Goel, A., & Mehra, A. (2019). Analyzing Contagion Effect in Markets During Financial Crisis Using Stochastic Autoregressive Canonical Vine Model. Computational Economics, 53(3), 921-950.

[4] BenSaïda, A. (2018). The contagion effect in European sovereign debt markets: A regime-switching vine copula approach. International Review of Financial Analysis, 58, 153-165.

[5] Yan, B., Guo, S., Guo, J., Chen, L., Liu, P., & Chen, H. (2010). Regional design flood composition based on Copula function. Shuili Fadian Xuebao(Journal of Hydroelectric Engineering), 29(6), 60-65.

[6] Vernieuwe, H., Vandenberghe, S., De Baets, B., & Verhoest, N. (2015). A continuous rainfall model based on vine copulas. Hydrology and Earth System Sciences, 19(6), 2685-2699.

[7] Zhang, L. S. V. P., & Singh, V. P. (2006). Bivariate flood frequency analysis using the copula method. Journal of hydrologic engineering, 11(2), 150-164.

[8] Ji, Q., Liu, B. Y., & Fan, Y. (2019). Risk dependence of CoVaR and structural change between oil prices and exchange rates: A time-varying copula model. Energy Economics, 77, 80-92.

[9] Papaefthymiou, G., & Kurowicka, D. (2008). Using copulas for modeling stochastic dependence in power system uncertainty analysis. IEEE Transactions on Power Systems, 24(1), 40-49.

[10] Hu, L. (2006). Dependence patterns across financial markets: a mixed copula approach. Applied financial economics, 16(10), 717-729.

[11] Demarta, S., & McNeil, A. J. (2005). The t copula and related copulas. International statistical review, 73(1), 111-129.

[12] Embrechts, P., Lindskog, F., & McNeil, A. (2001). Modelling dependence with copulas. Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich.

[13] Nelsen, R. B. (2007). An introduction to copulas. Springer Science & Business Media.

[14] Joe, H., & Kurowicka, D. (Eds.). (2011). Dependence modeling: vine copula handbook. World Scientific.

[15] Genest, C., & Rivest, L. P. (1993). Statistical inference procedures for bivariate Archimedean copulas. Journal of the American statistical Association, 88(423), 1034-1043.

[16] Li, D. X. (2000). On default correlation: A copula function approach. The Journal of Fixed Income, 9(4), 43-54.

[17] McNeil, A. J., & Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and L1-norm symmetric distributions. The Annals of Statistics, 37(5B), 3059-3097.

[18] Parker, G. A., & Pizzari, T. (2010). Sperm competition and ejaculate economics. Biological Reviews, 85(4), 897-934.

[19] Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. Statistica Sinica, 5-42.

[20] Demarta, S., & McNeil, A. J. (2005). The t copula and related copulas. International statistical review, 73(1), 111-129.

[21] Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and economics, 44(2), 199-213.

[22] Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2009). Pair-copula constructions of multiple dependence. Insurance: Mathematics and economics, 44(2), 182-198.

[23] Genest, C., & Favre, A. C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. Journal of hydrologic engineering, 12(4), 347-368.

[24] Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. International economic review, 47(2), 527-556.

[25] Jennions, M. D., & Petrie, M. (2000). Why do females mate multiply? A review of the genetic benefits. Biological Reviews, 75(1), 21-64.

[26] Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. Information processing & management, 37(6), 817-842.

[27] Marin, A., & Wellman, B. (2011). Social network analysis: An introduction. The SAGE handbook of social network analysis, 11.

[28] Zheng, L. Scientometrics (2019) 119: 1695. https://doi.org/10.1007/s11192-019-03098-9

[29] Ma, R. (2012). Author bibliographic coupling analysis: A test based on a Chinese academic database. Journal of Informetrics, 6(4), 532-542.

[30] Rousseau, R. (2010). Bibliographic coupling and co-citation as dual notions. The Janus faced scholar. A Festschrift in honour of Peter Ingwersen, 173-183.

[31] White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. Journal of the American society for information science, 49(4), 327-355.

[32] Zhao, D., & Strotmann, A. (2008). Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis. Journal of the American Society for Information Science and Technology, 59(13), 2070-2086.

[33] Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for information Science, 24(4), 265-269.

[34] Marshakova, I. V. (1973). System of document connections based on references. Nauchno-Tekhnicheskaya Informatsiya Seriya 2-Informatsionnye Protsessy I Sistemy, (6), 3-8.

[35] Kessler, M. M. (1963). Bibliographic coupling between scientific papers. American documentation, 14(1), 10-25.

[36] Burgoon, J. K., Johnson, M. L., & Koch, P. T. (1998). The nature and measurement of interpersonal dominance. Communications Monographs, 65(4), 308-335.

[37] Wei, Y. M., Mi, Z. F., & Zhang, H. (2013). Progress of integrated assessment models for climate policy. Syst Eng Theory Pract, 33(8), 1905-1915.

[38] Subramanyam, K. (1983). Bibliometric studies of research collaboration: A review. Journal of information Science, 6(1), 33-38.

[39] Neff, M. W., & Corley, E. A. (2009). 35 years and 160,000 articles: A bibliometric exploration of the evolution of ecology. Scientometrics, 80(3), 657-682.

[40] Borrett, S. R., Sheble, L., Moody, J., & Anway, E. C. (2018). Bibliometric review of ecological network analysis: 2010–2016. Ecological Modelling, 382, 63-82.

[41] Huang, C., Yang, C., Wang, S., Wu, W., Su, J., & Liang, C. (2019). Evolution of topics in education research: a systematic review using bibliometric analysis. Educational Review, 1-17.

[42] Gupta, B. M., & Dhawan, S. M. (2019). Deep Learning Research: Scientometric Assessment of Global Publications Output during 2004-17. Emerging Science Journal, 3(1), 23-32.

[43] Mao, M., Li, Z., Zhao, Z., & Zeng, L. (2018, June). Bibliometric Analysis of the Deep Learning Research Status with the Data from Web of Science. In International Conference on Data Mining and Big Data (pp. 585-595). Springer, Cham.

[44] Liao, H., Tang, M., Luo, L., Li, C., Chiclana, F., & Zeng, X. J. (2018). A bibliometric analysis and visualization of medical big data research. Sustainability, 10(1), 166.

[45] Khiste, G. P., Maske Dnyaneshwar, B., & Deshmukh, R. K. (2018). Big Data Output in J-gate during 2013 to 2017: A Bibliometrics Analysis. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 3(1), 1252-1257.

[46] Daim, T. U., Rueda, G., Martin, H., & Gerdsri, P. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. Technological Forecasting and Social Change, 73(8), 981-1012.

[47] Wasserman, P. D. (1989). Neural computing: theory and practice. Van Nostrand Reinhold Co..

[48] Yan, J. (2006). Multivariate modeling with copulas and engineering applications. Springer Handbook of Engineering Statistics, 973-990.

[49] Hao, Z., & Singh, V. P. (2016). Review of dependence modeling in hydrology and water resources. Progress in Physical Geography, 40(4), 549-578.

[50] Patton, A. J. (2012). A review of copula models for economic time series. Journal of Multivariate Analysis, 110, 4-18.

[51] Patton, A. J. (2009). Copula–based models for financial time series. In Handbook of financial time series (pp. 767-785). Springer, Berlin, Heidelberg.

[52] Zhang, Y., Kim, C. W., & Lin, J. (2018). Removing Environmental Influences in Health Monitoring for Steel Bridges Through Copula Approaches. International Journal of Steel Structures, 1-8.

[53] Kilgore, R. T., & Thompson, D. B. (2011). Estimating joint flow probabilities at stream confluences by using copulas. Transportation Research Record, 2262(1), 200-206.

[54] Zhou, X., Huang, G., Wang, X., Fan, Y., & Cheng, G. (2018). A coupled dynamical-copula downscaling approach for temperature projections over the Canadian Prairies. Climate dynamics, 51(7-8), 2413-2431.

[55] Schoelzel, C., & Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research–introduction to the copula approach. Nonlinear Processes in Geophysics, 15(5), 761-772.

[56] Kang, L., Jiang, S., Hu, X., & Li, C. (2019). Evaluation of Return Period and Risk in Bivariate Non-Stationary Flood Frequency Analysis. Water, 11(1), 79.

[57] Chebana, F., & Ouarda, T. B. M. J. (2011). Multivariate quantiles in hydrological frequency analysis. Environmetrics, 22(1), 63–78

[58] Favre, A. C., El Adlouni, S., Perreault, L., Thiémonge, N., & Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. Water resources research, 40(1).

[59] Cherubini, U., Luciano, E., & Vecchiato, W. (2004). Copula methods in finance. John Wiley & Sons.

[60] Genest, C., Gendron, M., & Bourdeau-Brien, M. (2013). The advent of copulas in finance. In Copulae and Multivariate Probability Distributions in Finance (pp. 13-22). Routledge.

[61] Jaworski, P., Durante, F., & Härdle, W. K. (2013). Copulae in mathematical and quantitative finance. Lecture Notes in Statistics-Proceedings, 213.

[62] Delatte, A. L., & Lopez, C. (2013). Commodity and equity markets: Some stylized facts from a copula approach. Journal of Banking & Finance, 37(12), 5346-5356.

[63] Karimalis, E. N., & Nomikos, N. K. (2018). Measuring systemic risk in the European banking sector: A Copula CoVaR approach. The European Journal of Finance, 24(11), 944-975.

[64] Oh, D. H., & Patton, A. J. (2018). Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. Journal of Business & Economic Statistics, 36(2), 181-195.

[65] Dastgir, S., Demir, E., Downing, G., Gozgor, G., & Lau, C. K. M. (2019). The causal relationship between Bitcoin attention and Bitcoin returns: Evidence from the Copula-based Granger causality test. Finance Research Letters, 28, 160-164.

[66] Embrechts, P.: Copulas: a personal view. J. Risk Ins. 76(3), 639–650 (2009)

[67] Sklar, A. (1959) "Fonctions de repartition a n dimentional et leurs marges", Publications de l' institut de Statistique de l' Universite de Paris, **8**, pp: 229-231.

[68] Joe, H. (1997). Multivariate models and multivariate dependence concepts. Chapman and Hall/CRC.

[69] Nelsen, R.B. (1999). An Introduction to Copulas. Lecture Notes in Statistics, vol. 139. Springer, New York, NY