Scientometrics
https://doi.org/10.1007/s11192-021-03884-4

Growth and dynamics of Econophysics: a bibliometric and network analysis

Kiran Sharma[1] · **Parul Khurana**[2]

Received: 4 November 2020 / Accepted: 21 January 2021

Abstract

Digitization of publications, advancement in communication technology, and the availability of bibliographic data have made it easier for the researchers to study the growth and dynamics of any discipline. We present a study on "Econophysics" metadata extracted from Web of Science managed by the Clarivate Analytics from 2000 to 2019. The study highlights the growth and dynamics of the discipline by measuring the number of publications, citations on publications, other disciplines contribution, institutions participation, country-wise spread, etc. We investigate the impact of self-citations on citations with every five-years interval. Also, we find the contribution of other disciplines by analyzing the cited references. Results emerged from micro, meso and macro-level analyses of collaborations show that the distributions among authors collaboration and affiliations of authors follow a power law. Thus, very few authors keep producing most of the papers and are from few institutions. We find that China is leading in the production of a number of authors and a number of papers; however, shares more of national collaboration rather than international, whereas the USA shares more international collaboration. Finally, we demonstrate the evolution of the author's collaborations and affiliations networks from 2000 to 2019. Overall the analysis reveals the "small-world" property of the network with average path length 5. As a consequence of our analysis, this study can serve as an in-depth knowledge to understand the growth and dynamics of Econophysics network both qualitatively and quantitatively.

Keywords Econophysics · Bibliometric analysis · Citations analysis · Scientific collaboration · Co-authorship network

✉ Kiran Sharma
kiransharma1187@gmail.com

Parul Khurana
parul11183@gmail.com

[1] Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, USA

[2] School of Computer Applications, Lovely Professional University, Phagwara, Punjab 144401, India

Published online: 02 March 2021

Introduction

Scientific collaborations have seen considerable growth in recent times and have emerged as an important factor for productive and qualitative research. The citation analysis of the scientific publications have become a tool to analyze the individual's performance, journal's impact as well as the discipline's growth (Zeng et al. 2017; Radicchi et al. 2017). Bibliometrics analysis not only makes a decision on the researcher's growth, in fact, it also measures the growth of a discipline. Many new interdisciplinary and multidisciplinary fields have arisen over time which in turn have increased and strengthened the interdisciplinary collaborations (Amaral et al. 1999; Stanley et al. 1999; Chakraborti et al. 2016). One such interdisciplinary field is "Econophysics" which was coined by H. Eugene Stanley in 1995 (Stanley and Mantegna 2000; Chakrabarti et al. 2006). Initially, physicists and economists contributed together to start this field and started applying theories and methods of physics to address problems in economics and stock markets (Carbone et al. 2007; Roehner 2010; Chakraborti et al. 2011; Pereira et al. 2017; Abergel et al. 2019). Later on, with the acceptance of the idea, scholars from other disciplines started contributing. Before the term Econophysics was coined, many people from different branches of science had worked and applied their knowledge in the field of economics leading to the evolution of Econophysics (Dash 2014).

Citations play a significant role in understanding the link between scientific works, and to understand the future research tendencies (Filser et al. 2017; Yang 2018; Tahamtan and Bornmann 2019; Chain et al. 2019). Nowadays, most of the research publications are created by teams of researchers instead of single individuals (Guimera et al. 2005). To investigate the patterns and trends of scientific collaboration, researchers have been working on publications data for a long time. There are different methods available in the literature to study collaborations and among them investigating the co-authorship network is the popular one (Sun and Rahwan 2017). A co-authorship network is a social network built on scientific collaborations, and thus it is amenable to social network analysis (Barabási 2016; Singh et al. 2020). With the development of complex network theory, researchers have been using network science to re-investigate the structural properties of co-authorship networks (Price 1965; Newman 2001, 2003; Newman et al. 2006; Zheleva et al. 2009).

Over time many such networks have been studied in different domains of social aspects like the author's collaborations (Newman 2001; Andrikopoulos et al. 2016), author's affiliations collaborations (Zheleva et al. 2009), and countries collaborations networks (Ortega and Aguillo 2013). Finding communities inside network (Good et al. 2010) and calculating centralities have been a major focus of social network analysis (Freeman 1977; Valente et al. 2008). It identifies critical pointers in the network and often used to equate popularity and leadership. The above-mentioned social networks are either directed or undirected where nodes act as authors and edges represent the collaboration among authors. The author's collaboration analysis is a micro-level study; however, such interactions among authors also give rise to institutional collaborations at meso-level and cross-country collaborations at the macro-level. Investigating the co-authorships network can help to identify entrants, leading researchers, and new collaborations. Co-authorship, institutional, and cross-country collaboration networks jointly reveal scientific collaboration and its growth (Chakrabarti and Chakraborti 2010; Ghosh 2013; Sinatra et al. 2016). This way we captured the changes in network structure at the microscopic, mesoscopic, and macroscopic levels and identified the key leaders at all levels.

The scholars have studied the econophysicists collaboration network earlier (Fan et al. 2004; Li et al. 2007); however, to the best of our knowledge, no one has performed systematic empirical research highlighting the patterns in data, key disciplines by cited references, and the patterns of collaborations at micro, meso, and macro-levels. At the micro-level, an author's collaboration, at the meso-level author's affiliation, and at macro-level countries' collaboration networks have been analyzed that demonstrate the in-depth knowledge of the growth of the discipline. This is the first time we are showing the detailed analysis of Econophysics through bibliometric and network analyses which cover the gap of the previous studies accomplished on it. It demonstrates the current state of Econophysics and provides researchers and practitioners with up-to-date knowledge. Thus, the objective of this study is to appraise the scientific evolution of Econophysics through various factors involved in information productivity and diffusion of knowledge.

To demonstrate the progress, growth and dynamics of Econophysics, the study is organized as follows: Section 2 provides the data description. Section 3 highlights the results which are further divided into three subsections: subsection 3.1 discusses the results on dynamics of citation patterns in the data and the key disciplines of the cited references. Section 3.2 presents a detailed discussion on the collaboration networks at micro- (3.2.1 and 3.2.2), meso- (3.2.3) and macro- (3.2.4) levels. Subsection 3.3 shows the growth of co-authorship and institutional networks over years. Section 4 concludes this study and discusses the limitations and future directions.

Data description

We collected the data from Web of Science managed (WoS) by Clarivate Analytics. The data mining API (https://apps.webofknowledge.com/) of WoS is used to fetch the records. We searched for the papers that match the keyword *Econophysics* in the "keyword" field of WoS published during 2000–2019. During 1995–1999 significant publication count is not available in WoS, so we could not perform the analysis since 1995. A total of 1458 records are retrieved including all document types. We further filtered the data based on the *Document Type* and included papers which are: *Articles, Reviews, Proceedings, Editorial Material, and Book Chapter* as these categories are having a sufficient number of papers. Hence, we finally filtered 1437 records. All records contain the full description of the paper like author name, affiliation, citations, publication journals, references, etc.

To retrieve the disciplines of the cited references of each paper, first, we extracted the title of each reference and then searched for that title in the WoS database. Not all cited references are listed in the WoS and this allowed us to match 74% of the references. This way we get the list of relative disciplines of all cited references in 1437 papers. To get the list of author's collaborations, we identified the author's unique ID provided by WoS (DAIS number) as there could be two authors with the same name. Similarly, corresponding to the author's ID, we identified the institutions. The corresponding author's location information is extracted from the reprinting address in the paper. Many scholars have studied the economy and the stock market behavior by using the methods of statistics, mathematics, computer science, etc. However, the focus of our study is to select papers where physics concepts have been used to study the economy and stock market behavior.

Results

Dynamics of citation patterns

We have presented the characterization of number of publications, citations, self-citations, etc. in Fig. 1. The number of papers published from 2000 to 2019 are reported in Fig. 1a. In Fig. 1b we have shown the cumulative growth of only those papers that have received more than 200 citations (highly cited) and the inset shows the growth

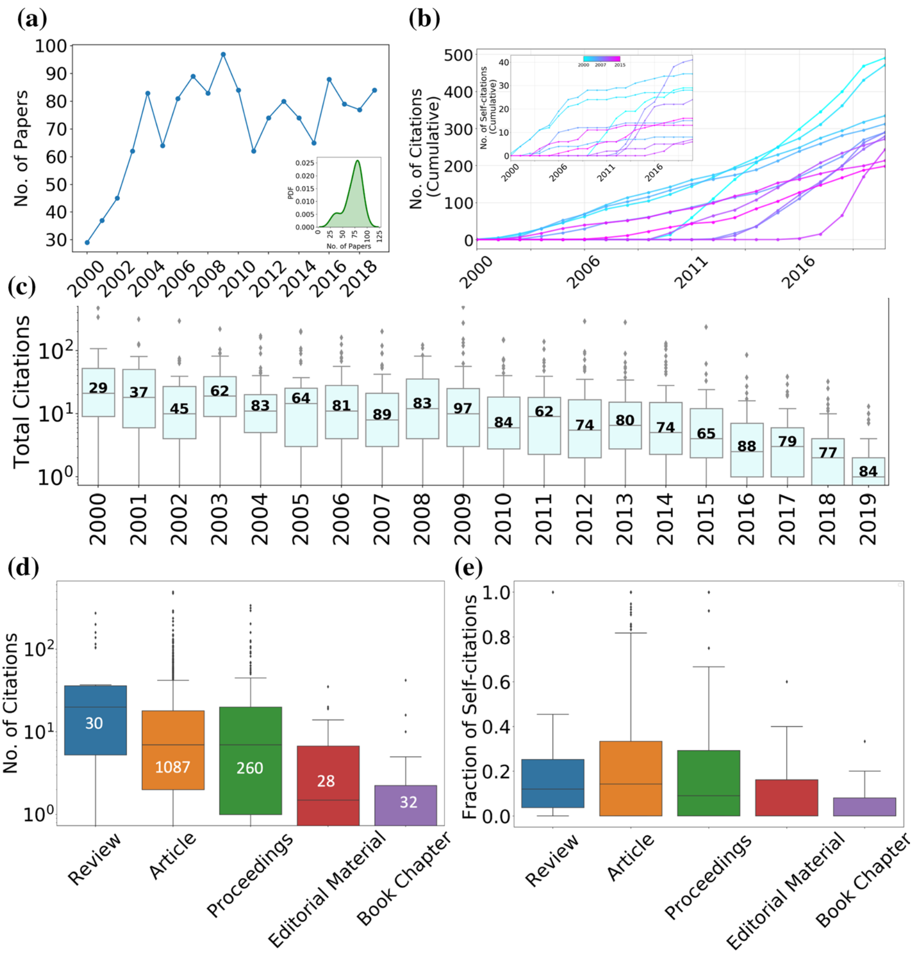


Fig. 1 Characterizing publications, citations, and self-citations. **a** Total number of papers published during 2000–2019. (Inset) Probability density of the number of papers. **b** The cumulative growth of only those papers that have received more than 200 citations. (Inset) Cumulative plot of the number of self-citations received. The color code corresponds to the publication year of each paper. **c** Citations received by papers published over years. The number inside the box shows the publication count corresponding to years. **d** The median number of citations received by different documents published during 2000–2019. The numeric value inside the box is the total number of published papers in that document category. **e** The fraction of self-citations received by each document category

of self-citations received by the same set of papers published from 2000 to 2019. Figure 1c represents the median number of citations received by all papers published over years. The numeric value inside the box plot represents the count for the total number of papers published in the respective year. The median number of citations received by papers published as *Articles*, *Reviews*, *Book Chapter*, etc. is shown in Fig. 1d and corresponding median self-citations is shown in Fig. 1e. The numeric value inside the plot is the number of papers published. The bars are arranged according to the median number of citations rather than the number of publications. For example, papers published as *Articles* and *Proceedings* have received equal median number of citations; however, the number of publications as *Articles* are higher than *Proceedings*. On the other hand, *Review* papers are less published as compared to other document types but have received the highest median number of citations. The median self-citations received by *Reviews* and *Articles* are almost same.

Figure 2 represents the dynamics of citations and self-citations over the years. Figure 2a shows the average age of a paper when it has received the first citation which is not a self-citation during 2000–2019. Similarly, the average age of a paper when it has received first self-citation is shown in Fig. 2b. On an average, the paper receives first citation and self-citation within the first two years after its publication. Figure 2c shows the overall citations and self-citations received by papers from 2000 to 2019. Higher the number of citations, the higher the self-citations. During the first five years of a publication, the count of self-citations has increased with the increase of citations as shown in Fig. 2d (Fowler and Aksnes 2007). This shows that during the initial year's authors tend to cite their papers quite often to maintain the visibility of the papers. This association decreases with the increase of the time interval (Fig. 2e, f).

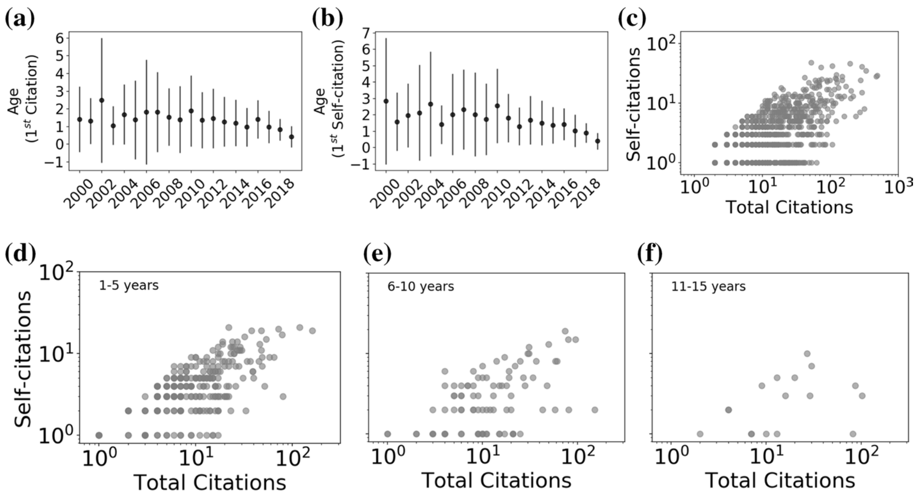


Fig. 2 Dynamics of citations and self-citations over the years. **a** Average age of the paper when it has received first citations which is not a self-citation during 2000–2019. **b** The average age of the paper when it has received the first self-citation. **c** A number of citations and self-citations received by papers from 2000 to 2019. **d** Relationship between the citations and self-citations received by each paper in the first five years after publication. **e**, **f** Five-five years' time interval behavioral change in citations and corresponding self-citations

Referencing disciplines in the papers

To understand which disciplines have contributed more to the growth of Econophysics, we analyzed the references cited by each paper. We retrieved the disciplines of all the cited references and analyzed the contribution of disciplines. Figure 3a highlights the disciplines according to the number of cited references (in %). It is evident that major references were quoted from *Physics* followed by *Economics* which clearly represents the true nature of Econophysics. The proportion of physics references also revealed the major contribution of physicists' in the field. Figure 3b highlights the journals based on the median number of citations received by the papers. The bars are arranged according to the median of citations rather than the number of citations. *Physica A* has published more papers (739) than *Physics Review E* (34); however, *Physics Review E* has received higher the median number of citations (20) than *Physica A* (10). The first few journals are also physics-based journals where papers have gained higher citations.

Collaboration network

Here we presented the scientific collaborations at micro, meso, and macro-level.

Micro-level analysis: author's collaboration network

In the co-authorship network (Fan et al. 2004), we have constructed an undirected weighted network consists of 1834 nodes and 4590 edges (3137 unique edges) as shown in Fig. 4a, where nodes correspond to authors and edges represent the collaboration (when two or more authors write a paper together). Single-authored articles are excluded from the data set since they do not contribute to the co-authorship network. The first five largest connected components of the network are colored differently. The giant component (colored in purple) contains the 30% of the total nodes of the network. The giant component is further elaborated in Fig. 5. The second-largest component (colored in green) contains 2% nodes, and so on (see network statistics table in Fig. 4). It is often perceived that certain authors are actively engaged in collaboration than others. Figure 4b, c shows the complementary cumulative density function (CCDF) of the degree of the nodes and edges strength which

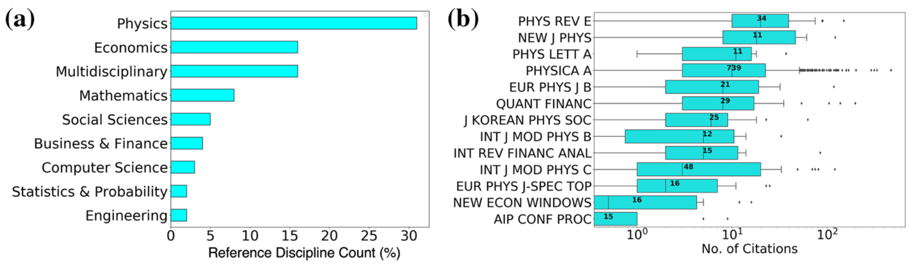
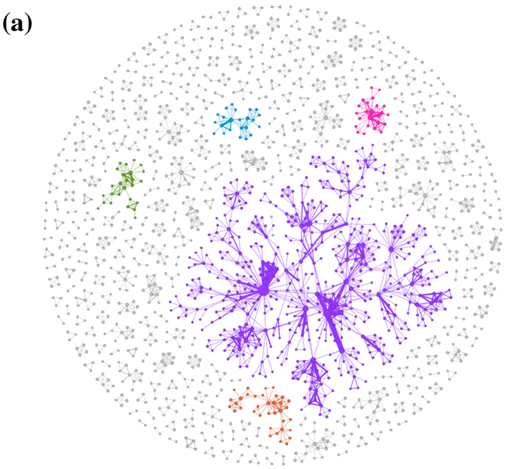


Fig. 3 Key disciplines by cited references and publication journals. **a** Bar plot shows the number of times (in %) a reference has been cited from a discipline. 31% of the references are cited from physics discipline and 16% are from economics, which clearly depicts the true nature of Econophysics. **b** The median number of citations received by different journals that published Econophysics papers. The numeric value inside the box is the total number of published papers. The bars are arranged according to the median of citations rather than the number of citations. The majority of the papers are published in physics journals



Network Statistics	
Total Nodes	1832
Total Edges	4589
Unique Edges	3137
Avg. Authors per Paper	2.58
Avg. Degree	3.42
Avg. Weighted Degree	5.01
Avg. Clustering Coefficient	0.87
No. of Triangles	2454
Size of Giant Component (%)	547 (30%)
2nd Largest Component (%)	39 (2%)
3rd Largest Component (%)	26 (1.4%)
4th Largest Component (%)	26 (1.4%)
5th Largest Component (%)	23 (1.2%)

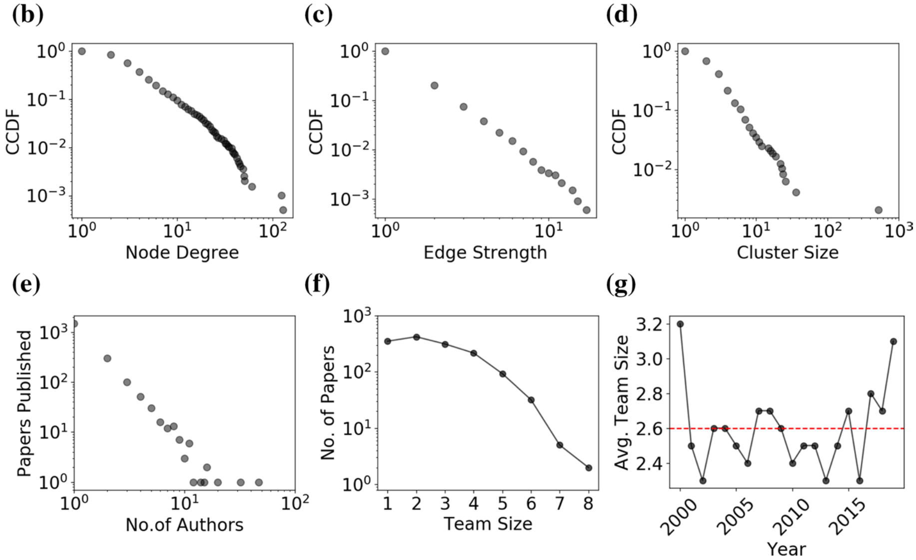
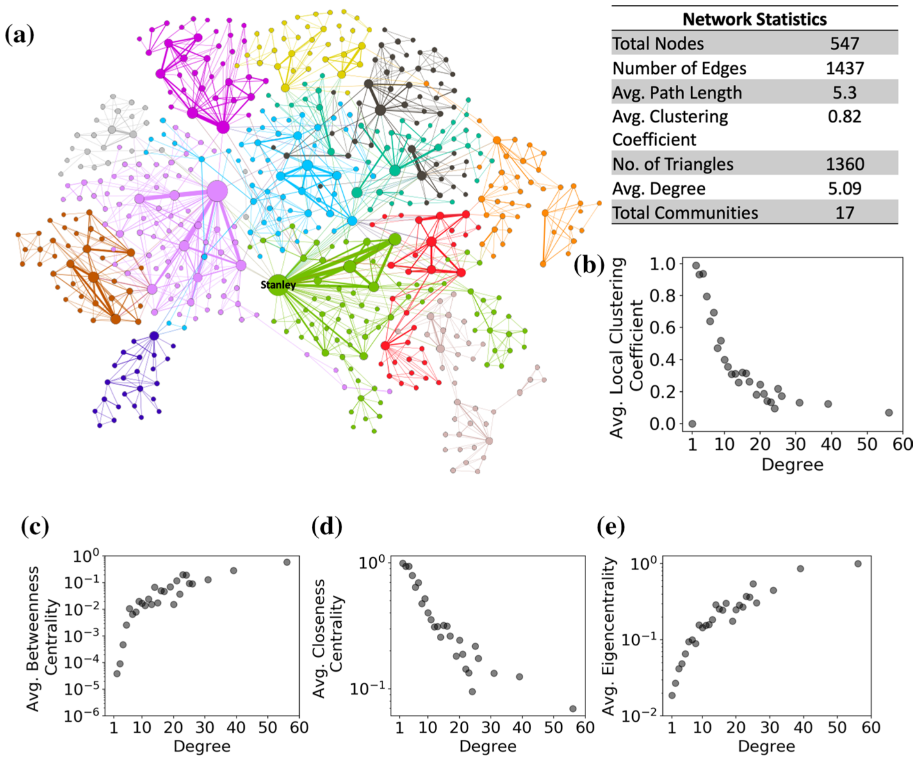


Fig. 4 Co-authorship network. **a** An undirected weighted co-authorship network having 1834 nodes and 4590 edges (3137 unique edges). The nodes represent authors and edges represent the collaboration among authors. We have filtered the self-loops in the network representation. The size of the node corresponds to the weighted degree of the node and the width of the edge represents the strength of the collaboration. Different colors represent the first five largest connected components. The giant component (colored in purple) contains 547 nodes which are 30% of the total nodes of the network. **b**-**d** The statistical properties of the network as complementary cumulative density functions (CCDF's): weighted degree, edge weight, and cluster size, respectively. **e** Number of papers published by authors represents the contribution of authors in the field. A few authors have published a large number of papers. **f** Papers published by teams of varying sizes. **g** Time evolution of the typical number of team members. The red line represents the average team size. The table shows the network statistics. The network is constructed in *Gephi 0.9.2*

represents the author's collaborations and the strength of the collaboration. The power-law behavior of CCDF shows that there are few authors who share a large number of collaborations. The CCDF's of the cluster size or connected components are shown in Fig. 4d. The power-law behavior of the cluster size distribution clearly shows that only one component



Network Statistics	
Total Nodes	547
Number of Edges	1437
Avg. Path Length	5.3
Avg. Clustering Coefficient	0.82
No. of Triangles	1360
Avg. Degree	5.09
Total Communities	17

Fig. 5 A giant component of the co-authorship network. **a** A zoomed-in view of the giant component. A modularity detection algorithm has been used to detect the communities among the network. The node with the largest connectivity is labeled by the author's name. **b** Degree versus the average local clustering coefficients of the nodes. On an average nodes of higher degrees exhibit lower local clustering. **c** Degree versus average betweenness centrality of the nodes. Nodes with higher betweenness centrality represent the potential key authors and nodes with higher degree represent hubs in the network. It highlights that the nodes with the highest degree act as the bridge to compute the shortest-path among all nodes in the network. **d** Degree versus average closeness centrality of the nodes. On an average, nodes of higher degrees share a low closeness. **e** Degree versus average eigencentrality of the nodes. The eigencentrality measures the prestige of the node in the network. On an average, nodes of higher degrees have higher prestige. The table shows the network statistics. The network is constructed in *Gephi 0.9.2*

contains a large number of nodes. In the network, 55% of nodes have clustering coefficient 1, and 28% have 0. The highest clustering coefficient represents how well the nodes are connected to their neighbors. The highest average clustering coefficient (0.87) shows that almost everyone is connected to others in the network. Figure 4e shows the relation between the number of authors and the number of papers published by them. A few authors have published more than 10 papers, whereas a large number of authors have published less than 10 papers. Also, how big is the team size of authors is studied in Fig.4f. The majority of the papers are either published as a single author or two authors. There are few papers that have been written by 7 to 8 authors which is also the largest team size. Figure 4g shows the evolution of team size in scientific collaborations (Guimera et al. 2005). Over the years the team size fluctuates from an average of 2 to an average of 3.

Several collaborative teams have worked on different topics and spread the growth in different parts of the globe. To name a few, we start with H.E. Stanley from Boston

University, USA coined the word “Econophysics” and started the field. He has published a large number of papers focused on stylized facts, random matrix theory, detrended fluctuation analysis to understand the behavior and dynamics of financial markets (Kantelhardt et al. 2002; Gabaix et al. 2003). Similarly, in China, Wei-Xing Zhou, a professor in Finance at East China University of Science and Technology has published a large number of papers highlighting the concept of multifractals, detrended fluctuation analysis to analyze the cross-correlations among stocks. He has also shed light on the bubble diagnosis and prediction (Jiang et al. 2010; Song et al. 2011). In Japan, H. Takayasu from Sony Computer Science Laboratories has worked on correlation networks among currencies, banking transactions, etc (Mizuno et al. 2006). In Italy, R.N. Mantegna from Palermo University has explained the structure and scaling behavior of the market, correlation and complexity in finance, and how to filter information in complex systems (Mantegna and Stanley 1995; Stanley and Mantegna 2000; Tumminello et al. 2005). In the United Kingdom, Tiziana Di Matteo and Tomaso Aste have worked on long-term memories of emerging markets and blockchain technologies (Di Matteo et al. 2005; Aste et al. 2017). In Switzerland, D. Sornette explained the critical events in complex systems and why the stock market crash (Sornette 2017). In Belgium, M. Ausloos has worked on foreign exchange currency and currency exchange rates (Vandewalle and Ausloos 1998). In India, A. Chakraborti has worked on financial market characterizations and kinetic wealth exchange models (Chakrabarti et al. 2013; Pharasi et al. 2018; Chakraborti et al. 2020). Obviously, this is not an exhaustive list.

Giant component

Figure 5a shows the zoomed-in view of the giant component of the co-authorship network extracted from Fig. 4a. A modularity maximization algorithm is used to find out the communities inside the giant component (Chen et al. 2014). Different colors represent different communities in the giant network. There is a total of 17 communities in the network and the node with the highest number of connections is labeled with the author’s name. Communities that are smaller in size are mainly having national collaborations while the communities that are larger in size have international collaborations. The largest community (colored in green) has 62 authors with H.E. Stanley and L.A.N. Amaral as one of the authors with a large number of collaborations. The second-largest community has 55 authors (colored in black) with T. Aste, M. Ausloos, T. DiMatteo as one of the authors in the community. The community colored in brown has 40 authors and presents a strong collaboration between China and Taiwan. The community colored in light pink has 39 authors and represents the national collaborations. All authors in this community are from China and WX Zhou shares large collaborations. The community colored in dark pink with 26 authors also represents national collaboration in Japan. The community colored in light blue has 34 authors and are mainly from South Korea. This community represents the strong national and weak international collaboration. Every community has at least one author that shares large number of collaborations (see Table 1).

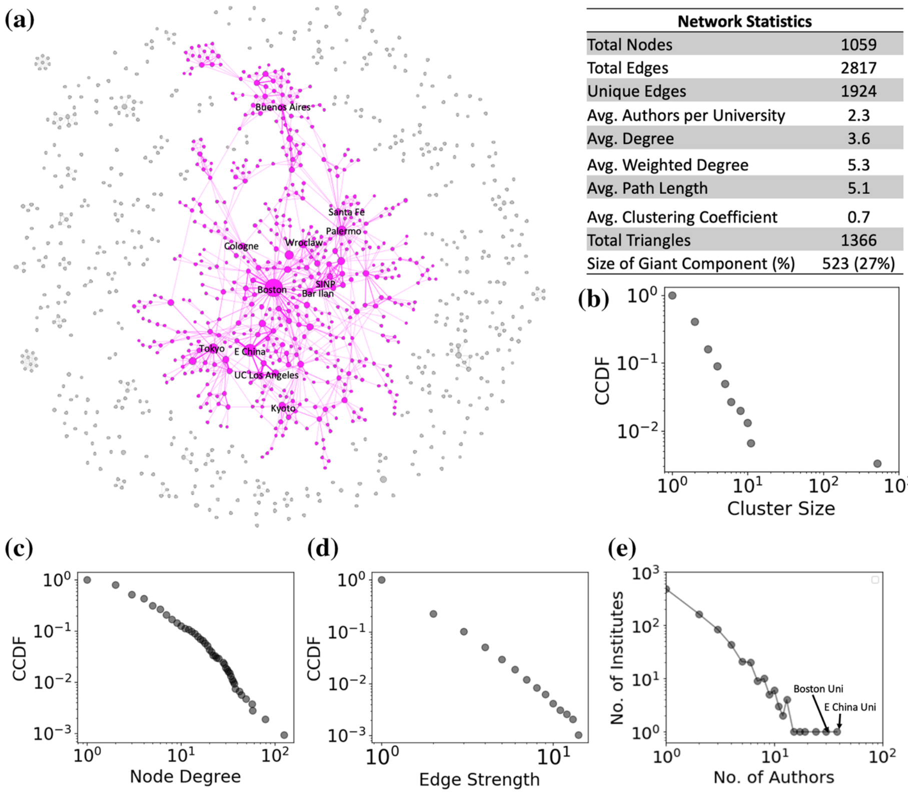
In addition, the average path length of the network is 5.3 which reveals the “small-world” property of the network (Watts and Strogatz 1998). In the community of econophysicists’, everyone is connected to others in ≈ 5 steps. The relationship between nodes degree and local clustering coefficient is shown in Fig. 5b. On an average, nodes of higher degree exhibit lower local clustering. A list of top 50 authors based on the degree (collaboration) is shown in Table 1. Figure 5c–e represents the relationship among nodes degree and the centrality measures: betweenness, closeness, and eigencentrality. On an average, nodes of higher degree exhibit higher betweenness and eigencentrality but lower closeness centrality. Nodes with higher betweenness centrality represent the potential key authors in the network, whereas nodes with higher degree represent hubs in the network. It highlights that the nodes with the highest degree act as the bridge to compute the shortest-path among all nodes in the network. Also, the largest value of eigencentrality represents the prestige/ influence of the node in the network.

Meso-level analysis: authors' affiliations network

After the institutionalization of Econophysics in 1995, many reputed institutes have initialized research on it and some institutes have started courses on it (Dash 2015; Ortega and Aguillo 2013). To investigate the contribution of different institutions, an undirected weighted authors' affiliations (institutions) network is constructed (see Figure 6 a). The network consists of 1059 institutions/universities and shows 2817 possible collaborations between institutions across the globe. Self-loops are removed while plotting the network; however, included in the analysis. The giant component (colored in dark pink) contains 27% of the network nodes shown in Fig. 6b as CCDF of cluster size. Fig. 6(c-d)) show the CCDF's of nodes degree and edges strength, respectively. Figure 6 e shows the number of authors corresponding to the number of institutions working on Econophysics. A large number of authors belong to a few institutions. The top two institutions in terms of the number of authors and collaborations are *East China University of Science and Technology* (ECUST) and *Boston University*. ECUST produces a large number of authors, whereas Boston University shares a large number of collaborations (see Table 2 for institutions details).

Macro-level analysis: countries' collaborations network

To visualize the expansion of the econophysicists' across the globe we have studied the geolocations of authors. Figure 7a represents the number of authors in different countries' (in %) working on Econophysics. The violet-colored bars represent the corresponding authors who lead the projects and cyan colored bars represent the co-authors of the papers. Here, we displayed results only for few countries' as per the number of corresponding authors. China is leading in terms of the number of corresponding as well as co-author's participation. Figure 7b highlights the number of papers published by the number of authors in the respective countries. The results are presented in 71 countries. The trend reveals the signature of scaling behavior in terms of the author's publications across the globe. Further, an undirected weighted network of countries' with 71 nodes and 1716 edges is constructed in Fig. 7c. There are self-loops present in the network which correspond to either a single author paper or collaboration among the same country. The size of the node represents the number of authors in the respective country and the edge width represents the number of times a collaboration occurred. Results highlight a strong collaboration between the USA and France; however, the number of authors is higher in the USA rather than in France which shows there might be a small but active community of researchers in the field. We also find that the within-country collaboration is more active as compared to cross-country. Hence, China, the USA, Italy, Japan, Germany, France, etc. have a large number of authors, a large number of publications, and strong connectivity/ collaboration among them (self-loops not shown in-network). We can say that these



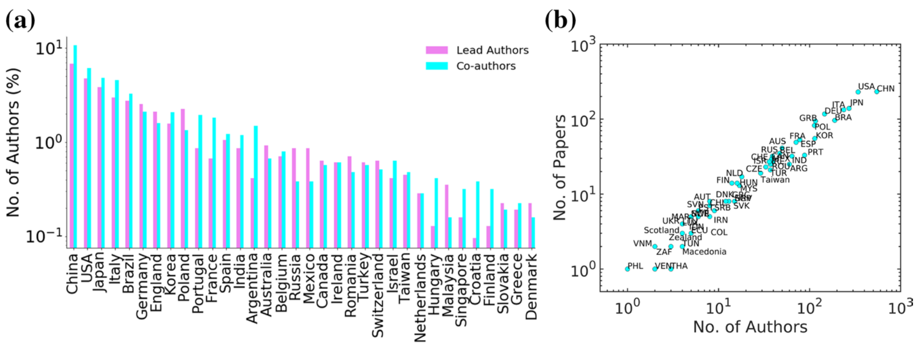
Network Statistics	
Total Nodes	1059
Total Edges	2817
Unique Edges	1924
Avg. Authors per University	2.3
Avg. Degree	3.6
Avg. Weighted Degree	5.3
Avg. Path Length	5.1
Avg. Clustering Coefficient	0.7
Total Triangles	1366
Size of Giant Component (%)	523 (27%)

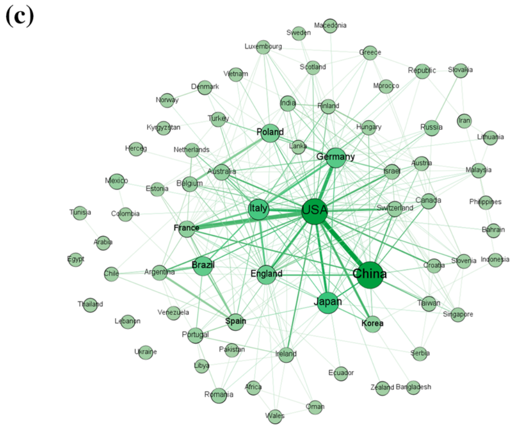
Fig. 6 Author's affiliations network. **a** An undirected weighted network of institutions having 1059 nodes and 2817 edges (1924 unique edges) where nodes represent the institutions and edges represent the collaboration among the institutes across the globe. The giant component (colored in dark pink) comprises 27% of the nodes. The institutes with strong collaborations are labeled with the names. There is an isolated institution in the network that corresponds to within institution collaboration; however, we have filtered the self-loops in the network representation. The size of the node represents the weighted degree and the width of the edge represents the collaboration strength. **b** CCDF of cluster size. **c** CCDF of nodes degree. **d** CCDF of edges strength. **e** A number of authors corresponding to a number of institutions. A large number of authors correspond to a few institutions. The table shows the network statistics. The network is constructed in *Gephi 0.9.2*

leading countries' are driving the discipline; however, other countries are also contributing to the growth of the discipline and getting connecting to the leading countries. Figure 7d shows the evolution of the cumulative growth of international and national collaborations. Results highlights that national collaboration is higher than international collaborations. China shares more nationals, whereas the USA shares more international collaborations. There is a dip in the international collaborations trend during 2007–2008, this was the time when the stock market crashed due to the bankruptcy of Lehman Brothers.

Network growth

People form a team through collaborations due to the need to incorporate individuals with different skills, ideas, and resources. The main products of the collaborations at micro,





Network Statistics	
Total Nodes	71
Number of Edges	1716
Unique Edges	310
Avg. Path Length	2.26
Avg. Clustering Coefficient	0.57
No. of Triangles	413

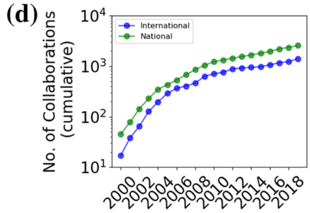


Fig. 7 Countries collaborations network. **a** Number of papers published as a corresponding author (colored violet) and as one of the authors (colored cyan) listed for few countries. The countries are arranged in descending order based on the number of corresponding authors. **b** Scattered plot for 71 countries representing the number of papers published by authors. **c** An undirected weighted network of countries corresponding to the author's location contains 71 nodes and 1716 edges (310 unique edges) where nodes represent countries and edges represent the scientific collaboration. There are a few isolated countries too. For simplicity, we filtered the self-loops from the network representation which correspond to within the country collaboration. The size of the node represents the weighted degree and the color gradient of the nodes varies according to the degree. The edge width represents the number of connections/collaborations among the nodes. The cross-country network shows the countries that have strong ties among them. **d** The evolution of the cumulative growth of international and national collaborations. The table shows the network statistics. The network is constructed in *Gephi 0.9.2*

meso, and macro-levels are the complementary skill sets or the expertise used in the development and growth of the discipline (Guimera et al. 2005). Also, such collaborations at different levels provide fresh insights into new problems and challenges (Chakrabarti and Chakraborti 2010). Econophysics had started by using tools of physics to solve financial and economics problems. Now the field has emerged and has been expanding by gaining insights from other disciplines/domains like machine learning, artificial intelligence, deep learning, game theory, etc. Also, other research communities have started accepting and appreciating it widely.

The evolution of fundamental statistical properties of the scientific collaboration networks in terms of the average degree of the nodes ($< k >$), average clustering coefficient ($< cc >$), and size of the giant component (GC(%)) during 2000–2019 is shown in Fig. 8.

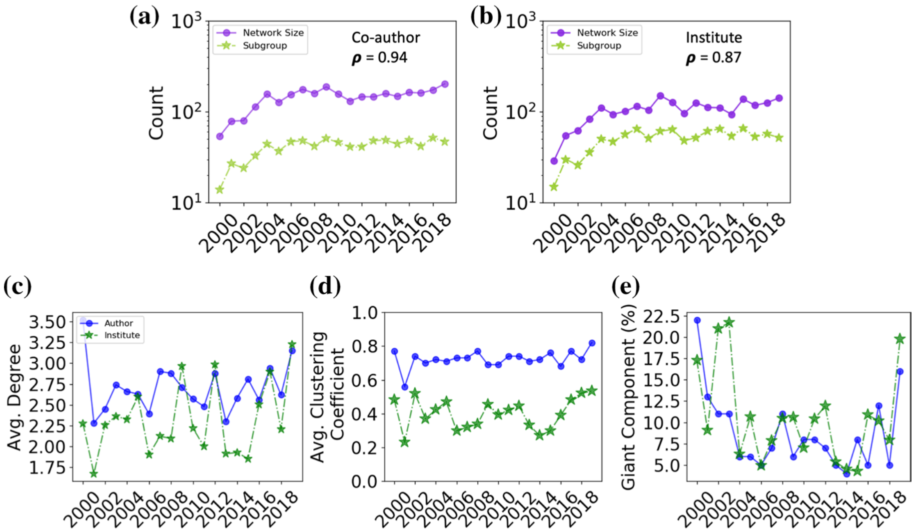


Fig. 8 Network growth over the years. **a, b** Size of the network and number of connected components of co-authorship and institutional networks over years, respectively. The time series of network size evolution is highly correlated with the time series of the evolution of the number of connected components for both the networks. **c**-**e** Growth of co-authorship network and affiliation network over years: **c** average degree, **d** average clustering coefficient, and **e** size of the giant component (in %)

The evolution of the co-authorship network is shown in Fig. 8a where the time series of network growth and a number of connected components shares a high amount of correlation (0.94). Similarly, the evolution of the author's affiliation network is shown in Fig. 8b where the time series of network growth and a number of connected components also shares a high amount of correlation (0.87). The evolution of the network's average degree, average clustering coefficient, and size of the giant component (in%) are shown for both the co-authorship and affiliations networks in Fig. 8c–e, respectively. Dynamics in the size of the giant component reveals the growth of the new groups. On an average, the degree of the co-authorship network varies between 2 and 3, and the clustering coefficient varies between 0.6 and 0.8. Similarly, on an average, the degree of the affiliations network varies between 1 and 3 and the clustering coefficient varies between 0.2 and 0.6 over years. The average path length of the network lies between 2 and 3 which reveals the "small-world" behavior of the network at every time step. A higher average clustering coefficient shows that nodes are grouped into communities.

Discussion and conclusion

Several important developments in Econophysics research have already taken place in the last two decades. Econophysics has answered various questions of the financial market and helped us to understand the market dynamics. The work done by H.E. Stanley, J.P. Bouchaud, J.D. Farmer, etc. on the empirical characterization of financial time series has provided a strong mathematical foundation to the field (Farmer and Sidorowich 1987; Bouchaud and Potters 2003; Farmer et al. 2005). The time series analysis and modeling

Table 1 List of 50 authors based on the degree (collaboration). The table shows the Author name. country, affiliation, number of collaboration (k), and clustering coefficient (cc)

S. no.	Country	Author	Affiliation	k	cc
1	USA	Stanley, HE	Boston University	56	0.07
2	China	Zhou, WX	East China University of Science and Technology	39	0.13
3	Japan	Takayasu, H	Sony Computer Science Laboratories	31	0.13
4	Italy	Mantegna, RN	University of Palermo	26	0.17
5	China	Ren, FE	China University of Science and Technology	25	0.22
6	Belgium	Ausloos, M	University of Liege	24	0.10
7	England	Di Matteo, T	Kings College London	23	0.13
8	Switzerland	Sornette, D	Swiss Finance Institute	22	0.10
9	Japan	Takayasu, M	Tokyo Institute of Technology	22	0.19
10	Japan	Kaizoji, T	Int Christian University	21	0.20
11	USA	Yakovenko, VM	University of Maryland	21	0.18
12	China	Qiu, T	Nanchang Hangkong University	20	0.24
13	Ireland	Richmond, P	Univ Dublin Trinity College	19	0.17
14	Italy	Gallegati, M	Univ Politecn Marche	19	0.22
15	South Korea	Lee, JW	Inha University	19	0.15
16	Canada	Li, SP	University of Toronto	17	0.20
17	China	Jiang, ZQ	East China University of Science and Technology	17	0.33
18	China	Huang, JP	Fudan University	17	0.23
19	China	Xiong, X	Tianjin University	17	0.40
20	Japan	Takahashi, T	University of Tokyo	17	0.15
21	China	Zhong, LX	Dianzi University	16	0.33
22	China	Chen, W	Shenzhen Stock Exchange	16	0.42
23	England	Scalas, E	University of Sussex	16	0.15
24	Italy	Lillo, F	Scuola Normale Super Pisa	16	0.37
25	Japan	Fujiwara, Y	University of Hyogo	16	0.30
26	China	Gu, GF	East China University of Science and Technology	15	0.45
27	South Korea	Jung, WS	Pohang University of Science and Technology	15	0.34
28	USA	Johnson, NF	University of Miami	15	0.16
29	England	Schinckus, C	University Leicester Finance	14	0.26
30	China	Jiang, XF	Collaborat Innovat Ctr Adv Microstruct	13	0.32
31	China	Zheng, B	Collaborat Innovat Ctr Adv Microstruct	13	0.26
32	China	Zhang, W	Tianjin University	13	0.59
33	Croatia	Podobnik, B	University of Rijeka	13	0.24
34	England	Preis, T	University College London	13	0.33
35	India	Chakrabarti, BK	Saha Institute of Nuclear Physics	13	0.24
36	Ireland	McCauley, JL	NUI Galway	13	0.26
37	Japan	Mizuno, T	University of Tsukuba	13	0.33
38	South Korea	Kim, SY	Korea Advance Institute of Science & Technology	13	0.24
39	Australia	Aste, T	Australian National University	12	0.21
40	China	Wang, GJ	Hunan University	12	0.26
41	China	Xie, C	Hunan University	12	0.30
42	South Korea	Yang, JS	Korea University	12	0.36
43	South Korea	Moon, HT	Korea Advance Institute of Science and Technology	12	0.41
44	Germany	Mimkes, J	University of Gesamthsch Paderborn	11	0.36

Table 1 (continued)

S. no.	Country	Author	Affiliation	k	cc
45	South Korea	Kim, S	Korea Advance Institute of Science and Technology	11	0.29
46	South Korea	Oh, G	Pohang University of Science and Technology	11	0.31
47	USA	Amaral, LAN	Northwestern University	11	0.46
48	China	Yang, G	Fudan University	10	0.49
49	China	Zhang, YJ	Tianjin University	10	0.51
50	India	Chakraborti, A	Jawaharlal Nehru University	10	0.29

of the financial market by M. Ausloos, A. Carbone, E. Scalas (Carbone et al. 2004; Scalas 2006) has enriched the filed both mathematically and statistically. Further, the field has emerged with the work of A. Chakraborti, D. Sornette, F. Lillo, R.N. Mantegna, T. Aste, etc. on financial market crash predictions. D. Garlaschelli, G. Caldarelli, G. Iori, etc. have highlighted the concepts of money market, world trade, and equities in financial markets in terms of networks (Bonanno et al. 2004; Garlaschelli and Loffredo 2005; Iori et al. 2008). Also, the groups of R.N. Mantegna, M. Marsili, J. Kertesz, K. Kaski, Sinha and others have provided the various network models and characterization of market correlations among different stocks/sectors. The groups of J.P. Bouchaud, T. Lux, D. Stauffer, M. Gallegati, D. Sornette, T. Kaizoji and others have contributed to the development of behavioral models, and analyses of market bubbles and crashes. The mathematical and agent-based modeling approaches have enhanced the depth of the field. The models on wealth distribution by A. Chakraborti, B. K. Chakrabarti, M. Patriarca, J.P. Bouchaud, V.M. Yakovenko, S. Solomon, P. Richmond, etc. (Dragulescu and Yakovenko 2000; Solomon and Richmond 2001; Chakrabarti et al. 2013; Boghosian 2019), and game-theoretic models by Challet et al. (2013); Chakraborti et al. (2015) have provided new insights and directions to the field. The importance and proliferation of the interdisciplinary research of Econophysics is highlighted in this special issue of Science & Culture, which presents a collection of twenty-nine papers written by more than forty renowned experts in physics, mathematics or economics, from all over the world (Chakrabarti and Chakraborti 2010). Nowadays formal and introductory courses in Econophysics are being offered by many distinguished universities like the Leiden University, the ETH Zurich, the Casimir Research School, etc. along with faculty positions in Econophysics.

To conclude, we have presented the detailed analysis of "Econophysics" in terms of the evolution and structure of collaborations networks from 2000 to 2019. We have performed a systematic empirical research highlighting the patterns in data, key disciplines by cited references, and the patterns of collaborations at micro, meso, and macro-levels. The key findings of the study are: (i) The impact of self-citations on citations reveals that in first few years the publications have received more self-citations and this trend goes down with time. Also, on an average a paper has received first self-citations in first two years after the publication. (ii) The disciplines extracted from cited references from all published papers highlights the higher contribution of *physics* and second highest of *economics*. The higher contribution of physicists' towards the growth of Econophysics reveals the true nature of the discipline. (iii) The co-authorship network at micro-level identifies the key authors and their contributions as an individual or in group. Also, the number of papers contributed by teams of varying sizes and the evolution of the team size over time is presented. We identified communities inside the giant component of the network and presented the relationships among nodes degrees and centrality measures (betweenness, closeness and

Table 2 List of 50 institutes based on the degree (collaboration). The table shows the institute name, number of collaborations (k) and number of authors (#a)

S. no.	Institutes	k	#a
1	Boston University, USA	56	30
2	East China University of Science and Technology, China	36	38
3	University of Palermo, Italy	34	13
4	University Buenos Aires, Argentina	26	5
5	University of Tokyo, Japan	25	17
6	Int Christian University, Japan	23	2
7	University of Leicester, England	23	8
8	Santa Fe Institute, USA	22	9
9	UCL, England	21	8
10	Kyoto University, Japan	19	7
11	Tokyo Institute of Technology, Japan	19	13
12	Aalto University, Finland	18	7
13	Ist Nazl Fis Nucl, Italy	18	2
14	Korea Advance Institute of Science and Technology, South Korea	18	11
15	University of Wroclaw, Poland	18	7
16	University Maryland, USA	17	5
17	Bar Ilan University, Israel	16	4
18	University Cologne, Germany	16	6
19	Kanazawa Gakuin University, Japan	15	2
20	National University of Singapore	15	5
21	University of California Los Angeles, USA	15	1
22	University Politecn Marche, Italy	15	2
23	Complexity Science Hub Vienna, Austria	14	3
24	Consejo Nacl Invest Cient & Tecn, Argentina	14	3
25	ETH, Switzerland	14	6
26	University of Evora, Portugal	14	3
27	CNRS, France	13	3
28	Saha Institute of Nuclear Physics, India	13	10
29	Sony Compter Science Labs, Japan	13	2
30	Swiss Federal Institute of Technology, Switzerland	13	6
31	Trinity College Dublin, Ireland	13	6
32	Federal University of Rio Grande do Sul, Brazil	13	6
33	University of Pavia, Italy	13	3
34	Artemis Capital Asset Management GmbH, Germany	12	1
35	Kings College London, England	12	10
36	Korea University, South Korea	12	3
37	Peking University, China	12	9
38	Tel Aviv University, Israel	12	6
39	University of Catolica Brasilia, Brazil	12	3
40	University of Electronic Science and Technology, China	12	7
41	University of Fed Alagoas, Brazil	12	8
42	University of Kiel, Germany	12	3
43	University of Piemonte Orientale, Italy	12	2
44	University of Politecn Madrid, Spain	12	6

Table 2 (continued)

S. no.	Institutes	*k*	#a
45	University of Porto, Portugal	12	6
46	Budapest University of Technology & Economics, Hungary	11	7
47	Pohang University of Science and Technology, South Korea	11	5
48	University of Adelaide, Australia	11	2
49	University of Liege, Belgium	11	1
50	Zhejiang University, China	11	27

eigencentrality). (iv) We also explored the authors' affiliations and country collaborations at meso and macro level. Results highlight that large number of authors are affiliated to a few numbers of institutions and China and USA has produced the higher authors as well as institutions. In terms of national and international collaborations, China share more national and USA shares more international collaborations. (v) Finally, the author's collaborations and affiliations networks are explored in terms of average degree, average clustering coefficient, average path length, size of giant component, etc. to study the networks evolution with a yearly resolution.

To conclude further, our study has provided an integrated view of citation dynamics and the growth of scientific collaborations networks of Econophysics metadata from 2000 to 2019. Our study justified the highest contribution of physicists' towards the field and to spread the visibility of the discipline, we suggest authors should publish more in interdisciplinary journals. However, the low number of publications reported under the Econophysics domain in Web of Science points out as a limitation of the study which further leads to the absence of the significant contribution of few authors. A possible future direction to extend the study is to integrate temporal data and quantify the evolution process of the co-authorship network and affiliations network (Börner et al. 2004). This could reveal how the importance of an author varies with time at different stages in his/her career.

Acknowledgements This work greatly benefited from discussions with and comments from A. Chakraborti and H. K. Pharasi. This work uses Web of Science data by Thomson Reuters provided by the Northwestern University.

Author Contributions KS conceived and designed the analysis, collected the data, performed analysis, prepared figures, and wrote the draft. PK collected the data, performed analysis, and reviewed the draft.

Compliance with ethical standards

Conflict of interest The author declare that they have no conflict of interest.

References

Abergel, F., Chakrabarti, B. K., Chakraborti, A., Deo, N., & Sharma, K. (2019). *New perspectives and challenges in econophysics and sociophysics*. Berlin: Springer.

Amaral, L. A., Cizeau, P., Gopikrishnan, P., Liu, Y., Meyer, M., Peng, C. K., & Stanley, H. (1999). Econophysics: Can statistical physics contribute to the science of economics? *Comput Phys Commun, 121,* 145–152.

Andrikopoulos, A., Samitas, A., & Kostaris, K. (2016). Four decades of the journal of econometrics: Coauthorship patterns and networks. *J Econ, 195*(1), 23–32.

Aste, T., Tasca, P., & Di Matteo, T. (2017). Blockchain technologies: The foreseeable impact on society and industry. *computer, 50*(9), 18–28.

Barabási, A. L., et al. (2016). *Network science*. Cambridge: Cambridge University Press.

Boghosian, B. (2019). Is inequality inevitable. *Sci Am, 321,* 70–77.

Bonanno, G., Caldarelli, G., Lillo, F., Micciche, S., Vandewalle, N., & Mantegna, R. N. (2004). Networks of equities in financial markets. *Eur Phys J B, 38*(2), 363–371.

Börner, K., Maru, J. T., & Goldstone, R. L. (2004). The simultaneous evolution of author and paper networks. *Proc Natl Acad Sci, 101*(suppl 1), 5266–5273.

Bouchaud, J. P., & Potters, M. (2003). *Theory of financial risk and derivative pricing: from statistical physics to risk management*. Cambridge: Cambridge University Press.

Carbone, A., Castelli, G., & Stanley, H. E. (2004). Time-dependent hurst exponent in financial time series. *Phys A Stat Mech Appl, 344*(1–2), 267–271.

Carbone, A., Kaniadakis, G., & Scarfone, A. (2007). Where do we stand on econophysics? *Phys A Stat Mech Appl, 382*, xi–xiv.

Chain, C. P., Santos, ACd., Castro, LGd., & Prado, JWd. (2019). Bibliometric analysis of the quantitative methods applied to the measurement of industrial clusters. *J Econ Surv, 33*(1), 60–84.

Chakrabarti, B. K., & Chakraborti, A. (2010). Fifteen years of econophysics research. *Sci Cult, 76*(9–10), 293–295.

Chakrabarti, B. K., Chakraborti, A., & Chatterjee, A. (2006). *Econophysics and sociophysics: Trends and perspectives*. New Jersey: Wiley.

Chakrabarti, B. K., Chakraborti, A., Chakravarty, S. R., & Chatterjee, A. (2013). *Econophysics of income and wealth distributions*. Cambridge: Cambridge University Press.

Chakraborti, A., Toke, I. M., Patriarca, M., & Abergel, F. (2011). Econophysics review: I. Empirical facts. *Quant Finance, 11*(7), 991–1012.

Chakraborti, A., Challet, D., Chatterjee, A., Marsili, M., Zhang, Y. C., & Chakrabarti, B. K. (2015). Statistical mechanics of competitive resource allocation using agent-based models. *Phys Rep, 552,* 1–25.

Chakraborti, A., Raina, D., & Sharma, K. (2016). Can an interdisciplinary field contribute to one of the parent disciplines from which it emerged? *Eur Phys J Spec Top, 225*(17), 3127–3135.

Chakraborti, A., Sharma, K., Pharasi, H. K., Bakar, S., Das, S., & Seligman, T. H. (2020). Emerging spectra characterization of catastrophic instabilities in complex systems. *New J Phys, 22*(6), 063043.

Challet, D., Marsili, M., Zhang, Y. C., et al. (2013). Minority games: interacting agents in financial markets. OUP Catalogue, Oxford University Press, number 9780199686698.

Chen, M., Kuzmin, K., & Szymanski, B. K. (2014). Community detection via maximization of modularity and its variants. *IEEE Trans Comput Soc Syst, 1*(1), 46–65.

Dash, K. C. (2014). Evolution of econophysics. In F. Abergel, H. Aoyama, B. K. Chakrabarti, A. Chakraborti, & A. Ghosh (Eds.), *Econophysics of agent-based models* (pp. 235–285). Cham: Springer International Publishing.

Dash, K. C. (2015). *Judging the impact of 'econophysics' through response to questionnaire* (pp. 327–348). Cham: Springer International Publishing.

Di Matteo, T., Aste, T., & Dacorogna, M. M. (2005). Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *J Bank Finance, 29*(4), 827–851.

Dragulescu, A., & Yakovenko, V. M. (2000). Statistical mechanics of money. *Eur Phys J B Condens Matter Complex Syst, 17*(4), 723–729.

Fan, Y., Li, M., Chen, J., Gao, L., Di, Z., & Wu, J. (2004). Network of econophysicists: A weighted network to investigate the development of econophysics. *Int J Modern Phys B, 18*(17n19), 2505–2511.

Farmer, J. D., & Sidorowich, J. J. (1987). Predicting chaotic time series. *Phys Rev Lett, 59*(8), 845.

Farmer, J. D., Patelli, P., & Zovko, I. I. (2005). The predictive power of zero intelligence in financial markets. *Proc Natl Acad Sci, 102*(6), 2254–2259.

Filser, L. D., da Silva, F. F., & de Oliveira, O. J. (2017). State of research and future research tendencies in lean healthcare: A bibliometric analysis. *Scientometrics, 112*(2), 799–816.

Fowler, J., & Aksnes, D. (2007). Does self-citation pay? *Scientometrics, 72*(3), 427–437.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40,* 35–41.

Gabaix, X., Gopikrishnan, P., Plerou, V., & Stanley, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature, 423*(6937), 267–270.

Garlaschelli, D., & Loffredo, M. I. (2005). Structure and evolution of the world trade network. *Phys A Stat Mech Appl, 355*(1), 138–144.

Ghosh, A. (2013). Econophysics research in India in the last two decades. *IIM Kozhikode Soc Manag Rev, 2*(2), 135–146.

Good, B. H., De Montjoye, Y. A., & Clauset, A. (2010). Performance of modularity maximization in practical contexts. *Phys Rev E, 81*(4), 046106.

Guimera, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science, 308*(5722), 697–702.

Iori, G., De Masi, G., Precup, O. V., Gabbi, G., & Caldarelli, G. (2008). A network analysis of the Italian overnight money market. *J Econ Dyn Control, 32*(1), 259–278.

Jiang, Z. Q., Zhou, W. X., Sornette, D., Woodard, R., Bastiaensen, K., & Cauwels, P. (2010). Bubble diagnosis and prediction of the 2005–2007 and 2008–2009 Chinese stock market bubbles. *J Econ Behav Org, 74*(3), 149–162.

Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Phys A Stat Mech Appl, 316*(1–4), 87–114.

Li, M., Wu, J., Fan, Y., & Di, Z. (2007). *Econophysicists collaboration networks: Empirical studies and evolutionary model* (pp. 173–182). Milano: Springer Milan.

Mantegna, R. N., & Stanley, H. E. (1995). Scaling behaviour in the dynamics of an economic index. *Nature, 376*(6535), 46–49.

Mizuno, T., Takayasu, H., & Takayasu, M. (2006). Correlation networks among currencies. *Phys A Stat Mech Appl, 364,* 336–342.

Newman, M. E. (2001). Scientific collaboration networks. I. Network construction and fundamental results. *Phys RevE, 64*(1), 016131.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM Rev, 45*(2), 167–256.

Newman, M. E., Barabási, A. L. E., & Watts, D. J. (2006). *The structure and dynamics of networks*. Princeton: Princeton University Press.

Ortega, J. L., & Aguillo, I. F. (2013). Institutional and country collaboration in an online service of scientific profiles: Google scholar citations. *J Informetr, 7*(2), 394–403.

Pereira, EJd. A. L., da Silva, M. F., & Pereira, Hd. B. (2017). Econophysics: Past and present. *Phys A Stat Mech Appl, 473,* 251–261.

Pharasi, H. K., Sharma, K., Chatterjee, R., Chakraborti, A., Leyvraz, F., & Seligman, T. H. (2018). Identifying long-term precursors of financial market crashes using correlation patterns. *New J Phys, 20*(10), 103041.

Price, D. J. D. S. (1965). Networks of scientific papers. *Science, 149,* 510–515.

Radicchi, F., Weissman, A., & Bollen, J. (2017). Quantifying perceived impact of scientific publications. *J Informetr, 11*(3), 704–712.

Roehner, B. M. (2010). Fifteen years of econophysics: Worries, hopes and prospects. arXiv preprint arXiv:10043229.

Scalas, E. (2006). The application of continuous-time random walks in finance and economics. *Phys A Stat Mech Appl, 362*(2), 225–239.

Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A. L. (2016). Quantifying the evolution of individual scientific impact. *Science*. https://doi.org/10.1126/science.aaf5239.

Singh, C. K., Vasques Filho, D., Jolad, S., & O'Neale, D. R. (2020). Evolution of interdependent co-authorship and citation networks. *Scientometrics, 125*(1), 385–404.

Solomon, S., & Richmond, P. (2001). Power laws of wealth, market order volumes and market returns. *Phys A Stat Mech Appl, 299*(1–2), 188–197.

Song, D. M., Tumminello, M., Zhou, W. X., & Mantegna, R. N. (2011). Evolution of worldwide stock markets, correlation structure, and correlation-based graphs. *Phys Rev E, 84*(2), 026108.

Sornette, D. (2017). *Why stock markets crash: Critical events in complex financial systems*. Princeton: Princeton University Press.

Stanley, H. E., & Mantegna, R. N. (2000). *An introduction to econophysics*. Cambridge: Cambridge University Press.

Stanley, H. E., Amaral, L. A. N., Canning, D., Gopikrishnan, P., Lee, Y., & Liu, Y. (1999). Econophysics: Can physicists contribute to the science of economics? *Phys A Stat Mech Appl, 269*(1), 156–169.

Sun, L., & Rahwan, I. (2017). Coauthorship network in transportation research. *Trans Res Part A Policy Pract, 100,* 135–151.

Tahamtan, I., & Bornmann, L. (2019). What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. *Scientometrics, 121*(3), 1635–1684.

Tumminello, M., Aste, T., Di Matteo, T., & Mantegna, R. N. (2005). A tool for filtering information in complex systems. *Proc Natl Acad Sci, 102*(30), 10421–10426.

Valente, T. W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How correlated are network centrality measures? *Connections (Toronto, Ont), 28*(1), 16.

Vandewalle, N., & Ausloos, M. (1998). Multi-affine analysis of typical currency exchange rates. *Eur Phys J B Condens Matter Complex Syst, 4*(2), 257–261.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *Nature, 393*(6684), 440–442.
Yang, J. (2018). Information theoretic approaches in economics. *J Econ Surv, 32*(3), 940–960.
Zeng, A., Shen, Z., Zhou, J., Wu, J., Fan, Y., Wang, Y., & Stanley, H. E. (2017). The science of science: From the perspective of complex systems. *Phys Rep, 714,* 1–73.
Zheleva, E., Sharara, H., Getoor, L. (2009). Co-evolution of social and affiliation networks. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 1007–1016.