

Warin, Thierry; Stojkov, Aleksandar

## Article

# Machine learning in finance: A metadata-based systematic review of the literature

Journal of Risk and Financial Management

## Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

*Suggested Citation:* Warin, Thierry; Stojkov, Aleksandar (2021) : Machine learning in finance: A metadata-based systematic review of the literature, Journal of Risk and Financial Management, ISSN 1911-8074, MDPI, Basel, Vol. 14, Iss. 7, pp. 1-31, <https://doi.org/10.3390/jrfm14070302>

This Version is available at:

<https://hdl.handle.net/10419/258406>

## Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

## Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

## Article

# Machine Learning in Finance: A Metadata-Based Systematic Review of the Literature

Thierry Warin <sup>1</sup>  and Aleksandar Stojkov <sup>2,\*</sup> <sup>1</sup> HEC Montreal, Montréal, QC H3T 2A7, Canada; thierry.warin@hec.ca<sup>2</sup> Iustinianus Primus Law Faculty, Ss. Cyril and Methodius University in Skopje, Skopje 1000, North Macedonia

\* Correspondence: a.stojkov@pf.ukim.edu.mk

**Abstract:** Machine learning in finance has been on the rise in the past decade. The applications of machine learning have become a promising methodological advancement. The paper's central goal is to use a metadata-based systematic literature review to map the current state of neural networks and machine learning in the finance field. After collecting a large dataset comprised of 5053 documents, we conducted a computational systematic review of the academic finance literature intersected with neural network methodologies, with a limited focus on the documents' metadata. The output is a meta-analysis of the two-decade evolution and the current state of academic inquiries into financial concepts. Researchers will benefit from a mapping resulting from computational-based methods such as graph theory and natural language processing.

**Keywords:** efficient market hypothesis; machine learning; network analysis; sentiment analysis



**Citation:** Warin, Thierry, and Aleksandar Stojkov. 2021. Machine Learning in Finance: A Metadata-Based Systematic Review of the Literature. *Journal of Risk and Financial Management* 14: 302. <https://doi.org/10.3390/jrfm14070302>

Academic Editor: Reza Tajaddini

Received: 25 April 2021

Accepted: 21 June 2021

Published: 2 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The theory and practice of finance have undergone a remarkable evolution in the past five decades. The emergence and acceptance of the Efficient Market Hypothesis (EMH), its subsequent mixed empirical record, the rise of pragmatically driven 'Chartism', and the present co-evolution of quantitative and behavioral finance represent some exciting significant developments in the financial domain.

The vibrancy of finance can also be observed by two methodological revolutions bringing sophisticated technical analysis of financial phenomena. Machine Learning Algorithms (MLAs) application in explaining and forecasting financial market trends has been a significant methodological advancement in the past three decades. Another critical research direction has been the rise of sentiment analysis of unstructured data relating to relevant news for financial markets.

In this article, we propose to take a comprehensive look at machine learning in finance. For that, we will use *neural network* as a keyword in our data collection. Using neural network as a keyword does not limit us to just neural networks approaches, because the source data will also contain other terms such as machine learning, deep learning, etc. The rationale behind using neural network as a core keyword is that the most influential papers introducing machine learning in finance used neural networks as a methodology of choice (i.e., Gencay and Stengos 1998).

Conventional systematic literature reviews (SLR) are a process that enables the collection of relevant evidence on a given topic that meets predefined eligibility criteria and provides an answer to the research questions formulated. A meta-analysis necessitates descriptive and/or inferential statistical methods to synthesize data from multiple studies on a particular subject. The techniques facilitate the generation of knowledge from a variety of studies, both qualitative and quantitative. The conventional method consists of four fundamental steps: search (define the search string and database types), appraisal (pre-defined literature inclusion and exclusion criteria, and quality assessment criteria), synthesis (extract and categorize the data), and analysis (narrate the results and finally reach a conclusion) (SALSA)

(Mengist et al. 2020). SLR is defined as a “systematic, explicit, and reproducible method for identifying, evaluating, and synthesizing the existing body of completed and recorded work” (del Amo et al. 2018). According to Grant and Booth (2009), the SALSA framework is a methodology for determining the search protocols that the SLR should follow. This ensures methodological precision, standardization, comprehensiveness, and reproducibility. The majority of scientific work employed this methodological approach to mitigate the risk of publication bias and increase the work’s acceptability (del Amo et al. 2018; Grant and Booth 2009; Malinauskaite et al. 2019; Perevochtchikova et al. 2019). Thus, most review articles followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocol and the Search, Appraisal, Synthesis, and Analysis (SALSA) framework (Grant and Booth 2009).

From SALSA, this article adds a pre-processing step to reduce potential human biases and highlights new results based on text-based analyses of the data collected.

Indeed, our main contribution is a computational systematic literature review of machine learning (and neural networks in particular) in finance between 1990 and 2021. We believe it is crucial to map the evolution of these new technologies and methodologies in our field. When scholars in the computer science field essentially develop the Artificial Intelligence (AI) sub-domain and machine learning techniques, including deep learning and reinforcement learning, it is interesting to look at the bridges between these developments and the ones in finance.

A second contribution is methodological. We indeed perform a metadata-based systematic review of the relevant literature. In the methodology section, we will provide a precise definition of the approach. We believe it is an essential methodological complement to conventional qualitative reviews and econometric-based meta-analyses. A metadata analysis means we will collect more articles than in a traditional systematic literature review and use algorithms to filter and sort the initial dataset. The methodological approach will be twofold: (1) we will use Natural Language Processing (NLP) techniques to extract text-as-data information, and (2) we will use graph theory to visualize potential collaboration networks. These two methodological approaches combined will provide us a different analysis than a formal systematic review. It is not to be seen as a substitute, but instead as a complement to the more conventional approach.

As an aside, and although we will not spend time on this aspect, a third contribution could be an epistemological one in nature and leverages our first contribution on the mapping of machine learning in finance to reflect on the implications of its significance on the old debate between theorists and chartists in finance. Markowitz (1952); Sharpe (1963, 1964), EMH emerged as a dominant paradigm providing a formal explanation of financial markets’ behavior. Empirical approaches emerged under the umbrella of “Chartism” (e.g., Berardi 2011). Chartists-or empirically minded technical analysts-have used extrapolative rules to discover statistical regularities in the time series for prices (e.g., Hsieh 1989; Frankel and Froot 1990; Taylor and Allen 1992; Menkhoff 1997, 2010; Lo 2004; Neely et al. 2009; Kaucic 2010; Gradojevic and Gencay 2013; Neely et al. 2014; Gerritsen et al. 2020). Additionally, a burgeoning literature on agent-based financial market models emerged, allowing various interactions between chartists and fundamentalists (e.g., Day and Huang 1990). Thanks to ML techniques, induction generates causal relationships based on information at the moment of estimation (Popper 1962; Warin 2005). These causal relationships are at the root of the predictive power of ML models. In the ML context, causality and prediction seem to get theorists and technical analyses closer.

The structure of the paper is as follows. In the next section, we provide a metadata-based systematic review of the academic literature on finance, published between January 1990 and May 2021. The third section elaborates the conceptual structures behind the relevant literature by exploring the keywords, keywords co-occurrences, and the topics’ evolution based on a topic modeling technique. In the next section, we examine the intellectual structures behind the evolution of analytical thinking on finance by focusing on what vehicles and which organizations are the main engines in this topic dynamics. The

fifth section critically examines the social structures of our sample, encompassing different measures to capture the social connections of authors, co-citations, and collaborations across institutions. The concluding remarks summarize the potential of machine learning, neural networks, and in general, the augmented technical analysis in analyzing financial markets.

## 2. Materials and Methods

A standard introduction to financial theory would often distinguish several valuation models that might be useful for analyzing securities and managing portfolios (see [Lee and Lee 2010](#)). Since the 1970s, the evolution of financial theory has been greatly influenced and informed by the emergence and acceptance of the EMH and the Modern Portfolio Theory (MPT) ([Prasch and Warin 2016](#)). Given the vast literature on financial analytics models, we confine our critical review only to the main strands of the relevant academic literature.

To illustrate the development of neural networks in finance, we conduct a scientometric study of the academic literature on finance, published between January 1990 and May 2021.

### 2.1. Methodology

The methodology used here is a systematic literature review with a different approach to more conventional reviews. In usual literature reviews, the author selects the relevant literature based on her domain or methodological expertise. Then, the analysis is based on the content found in the sample that has been created in the initial stage. The primary characteristics of SLR and its associated procedure, meta-analysis, are the following: (1) a clearly stated research question that the study will address; (2) explicit and reproducible objectives; (3) search strings that include all related studies that meet the eligibility criteria; and (4) an assessment of the quality/validity of the selected studies.

To have a comprehensive look, conventional systematic literature might not be the best choice. Considering the pace of the new developments in the artificial intelligence field, we propose here to map the extent of the usage of these new technologies and methodologies in finance. Systematic literature is a mapping exercise of a knowledge area, and it is also really focused, with between 50 to 200 papers being analyzed. Here, we also want to map the machine learning knowledge area while collecting a significant number of documents. The large dataset size will allow us to build an analysis based on the documents' metadata, such as authors' affiliations, universities, etc. This research protocol built around a metadata-based systematic literature review could be considered the first phase in a systematic literature review.

In contrast to more conventional methods, we have two phases: First, similar to a traditional systematic review, the selection of the relevant articles is performed via a search engine, except the expert does not select the relevant articles from the results presented to her. Here, the expert chooses the keywords and creates a comprehensive dataset of all the documents matching the keywords in the title, abstract, keyword, and keyword + section. The first phase, being automated, allows the utilization of quantitative criteria to filter down the dataset. Then, in the second phase, a dataset reduction to 50–200 documents is made by an expert.

To summarize, one of the critical contributions of a metadata-based systematic literature review is to reduce—though not wholly—potential human biases. Another significant contribution of this new methodology based on these two phases is that it allows us to consider the documents' metadata in a text format. By adding a computational treatment based on Natural Language Processing (NLP) techniques to transform the text into data, we can then provide analyses that would not be possible otherwise, leveraging analytical approaches such as graph theory. It is particularly relevant to discover research patterns, research history, the actual research vehicles, or to be able to associate discoveries with institutions, to name a few examples. These sophisticated techniques allow us to perform a literature mapping thanks to this computational approach.



Another critical point is the large size of the dataset, which has a lot of favorable statistical properties. We will also use algorithms to help us analyze quantities of papers that we would not be able to do otherwise due to the sheer quantity of information analyzed by a human.

Finally, another important dimension is using each document's reference section to perform metrics that allow researchers to understand the knowledge transmission patterns.

Beyond the computational treatment and to leverage the results obtained from these computations, we use the following theoretical framework. [Aria et al. \(2017\)](#) propose to look at three different structures: the conceptual, intellectual, and social structures. The conceptual structures are about leveraging the metadata to help us understand which concepts and topics are used in the academic conversation and how they have evolved through time. The intellectual structure will help us understand who produced these concepts, which journals played a pivot role in this nascent literature, and which articles were among the most referenced that fueled this literature. Lastly, the social structure will allow us to look at authors' collaborations and the knowledge support from universities and countries through their collaborations.

The data collection will be conducted using a "human-in-the-loop" (HIL) approach. It consists of proceeding to a purely automated data collection with an ex-post validation based on the field expertise.

First, we use an automated process in two phases as described earlier. The search was performed on the publisher-independent citation database "Web of Science" (WoS), *Clarivate Analytics*, by using combinations of keywords (and simultaneously removing the duplicates): "neural network\*" AND "finance\*".

These keywords allow us to build our sample. This sample does not aim at being representative of the domain. Instead, it intends to analyze the dynamics of the conversation about neural networks in finance. By building a sample about a modeling technique, we risk overfitting the true representativity of neural networks in finance if someone is interested in generalizing; this is not our intent.

We then use human-based field expertise to review the references anyway while adding some potential missing references based on the domain expertise (see Appendix A for a list of the added references). HIL allows us to have a combined qualitative assessment with pure automatic data collection. This second step is marginal in terms of added articles, but it is crucial for quality control.

Our approach differs at these two levels: in the sample creation, we try to be as comprehensive as possible on a particular topic, here "neural network\*" AND "finance\*". The stars mean that we collect any occurrence with a declination of the word's root. We use neural networks as a proxy for machine learning techniques as authors who use neural networks also reference machine learning in their keywords (among 10,160 used keywords and 3606 keywords Plus, see Table 1). So, the sample includes papers on machine learning as well. The sample is likely not comprehensive, as in any systematic literature review, but it is larger than conventional methods. The sample is collected by finding matches in the text title, the abstract, the keywords, and the keywords + in Web of Science. It helps us create a 5053 rich sample, a larger sample than regular, systematic reviews. We can deal with a larger sample thanks to the second differentiation point of our methodology: leveraging the sample metadata through computational techniques. The dataset can be found on the following webpage, including a search engine: <https://warin.ca/posts/article-machine-learning-finance/> (accessed date: 29 June 2021).

In this second level of differentiation, we create and use the metadata from the title, the abstract, the keywords, and the keywords +. The creation of metadata is conducted via Natural Language Processing (NLP) techniques. We prepare the dataset by selecting tokens, n-grams, etc. ([Aria and Cuccurullo 2017](#)).

These metadata are helpful to provide quantitative analysis to the sample. Using these machine learning tools allows us to have a research synthesis that can be leveraged with

other techniques such as social network analysis. We can also look at the dynamics of the research contributions, the collaborations, the idea generation, and propagation.

**Table 1.** Preliminary information about data, overall period, and per year.

Description	Overall Time Period (1990–2021)	2017	2018	2019	2020	2021
Sources (Journals, Books, etc.)	2533	265	329	374	333	107
Documents	5053	355	436	578	592	157
Average years from publication	7.74	4	3	2	1	0
Average citations per documents	14.66	10.9	8.278	5.005	2.255	0.465
Average citations per year per document	1.699	2.18	2.069	1.668	1.128	0.465
References	105,684	10,844	13,281	18,239	22,817	7313

Let us first look at the descriptive statistics before studying the dynamics of the research in this sample. We present the main descriptive statistics and empirical findings from the systematic literature review in the next step.

## 2.2. Descriptive Statistics

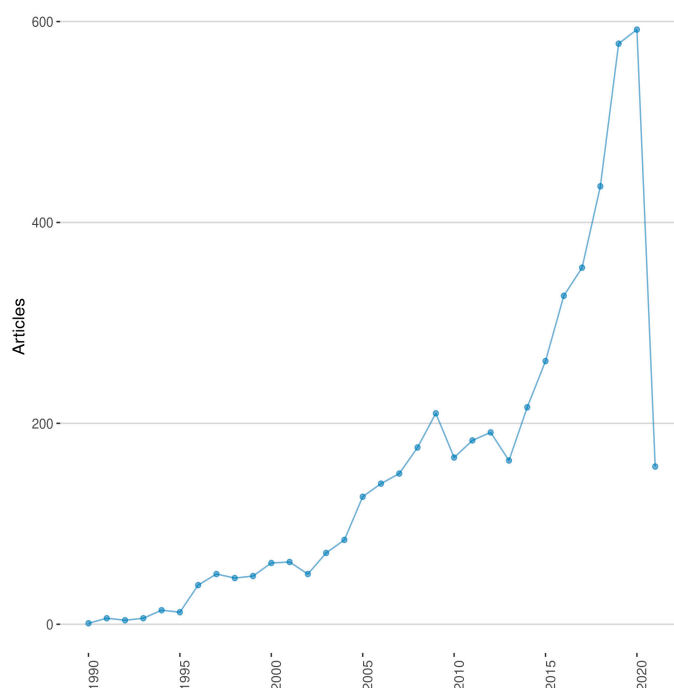
The relevant ‘universe’ of the literature consists of references identified in the HIL-Web of Science citation database (see Table 1) totaling 5053 documents, most of which are published in refereed journals (see Table 2). The literature review covers the period between 1 January 1990, and 10 May 2021 (see Figure 1).

**Table 2.** Document type, overall period, and per year.

Description	Overall Time Period (1990–2021)	2017	2018	2019	2020	2021
Article	2719	196	222	339	484	143
Article; easy access	67	0	0	0	0	0
Article; proceedings paper	143	1	4	2	0	1
Article; retracted publication	1	0	1	0	0	0
Bibliography	1	0	0	0	0	0
Biographical item	1	0	0	0	0	0
Book review	6	0	0	0	0	0
Correction	3	0	0	1	0	1
Editorial material	9	0	2	1	0	1
Letter	3	0	0	0	0	0
Meeting abstract	3	0	0	0	1	0
Proceedings paper	1974	150	194	216	79	0
Review	120	8	13	19	28	11
Review; early access	3	0	0	0	0	0

The overall number of documents in our sample is 5053 (see Table 1). This number is the cumulative result of each year, and we can observe a significant rise in the number of documents per year. The average citations per document are 14.66 but have evolved through time to numbers ranging between 1 and 2. As a reference point, the total citations per paper in economics and business for the highly cited papers were 3.04 for the 2011–2015 period and 3.91 for the 2017–2021 period. In Social Sciences in general, the total citations per paper for the highly cited papers were 2.89 for the 2011–2015 period and 3.30 for the 2017–2021 period. These results show the normalization of machine learning in finance-related documents.

The number of articles dominates the sample for the overall period (see Table 2) with 2719 occurrences, followed by 1974 proceedings papers. So, short contributions (articles and proceedings papers) represent the actual output in this sample. Authors indeed tend to produce the knowledge body about machine learning in finance through short contributions (e.g., [Gu et al. 2020](#)).



**Figure 1.** Article count through time.

Our database of references covers 308 keywords and 946 author appearances (see Table 3). Most of the publications are multi-authored documents, indicating the increasingly collaborative nature of research in the finance domain.

**Table 3.** Document content and authors, overall period, and per year.

Description	Overall Time Period	2017	2018	2019	2020	2021
Keyword Plus (ID)	3607	604	693	849	950	234
Author's Keywords (DE)	10164	1251	1429	1804	2044	688
Authors	9648	939	1210	1655	1651	492
Author Appearances	14628	1056	1350	1972	1985	519
Authors of single-authored documents	520	44	40	37	47	8
Authors of multi-authored documents	9128	895	1170	1618	1604	484

The descriptive statistical analysis also reveals that, on average, there are 2.32 authors per publication and 2.72 co-authors per publication (see Table 4). Most of the documents are collectively written. Only 661 documents have a single author.

**Table 4.** Authors' collaboration, overall period, and per year. Note: The Collaboration Index (CI) is calculated as total authors of multi-authored articles/total multi-authored articles.

Description	Overall Time Period	2017	2018	2019	2020	2021
Single-authored documents	661	46	42	37	49	9
Documents per Author	0.524	0.378	0.360	0.349	0.359	0.319
Authors per Document	1.91	2.65	2.78	2.86	2.79	3.13
Co-Authors per Documents	2.89	2.97	3.10	3.41	3.35	3.31
Collaboration Index	2.08	2.90	2.97	2.99	2.95	3.27

To conclude this descriptive statistics section, we observed a similar trend in the academic production about machine learning in finance based on short documents and co-authorship. Let us now analyze the three different structures: conceptual, intellectual, and social.

### 3. Conceptual Structures of Our Sample

The application of AI in the domain of finance is not a recent phenomenon in the academic literature (e.g., [Hutchinson et al. 1994](#); [Lo et al. 2000](#); [Gavrishchaka and Banerjee 2006](#); [De Spiegeleer et al. 2018](#); [Huang et al. 2020](#)). However, the last decade witnessed empirical studies using Machine Learning Algorithms (MLAs) to examine credit risk analysis and forecasting stock returns. As [Dixon et al. \(2020, p. vii\)](#) highlight, “ML in finance sits at the intersection of several emergent disciplines, including pattern recognition, financial econometrics, statistical computing, probabilistic programming, and dynamic programming”. One of the main competitive advantages of ML is that computers have an outstanding ability to process large amounts of financial information.

From a methodological perspective, the empirical studies rely not only on conventional MLAs such as support vector machine (SVM) and k-nearest neighbors (kNN) but also on Deep Learning (DL) (e.g., [Krauss et al. 2017](#); [Fischer and Krauss 2018](#); [Huang et al. 2020](#)), an advanced technique based on artificial neural network algorithms (e.g., [Chung-Ming and White 1994](#); [Donaldson and Kamstra 1997](#); [Hans and Griensven 1998](#); [Gencay and Stengos 1998](#); [Blake and Kapetanios 2000](#); [Garcia and Gencay 2000](#); [Fernandez-Rodriguez et al. 2000](#); [Bekiros and Georgoutsos 2008](#); [Kristjanpoller and Minutolo 2018](#); [Atsalakis et al. 2019](#)). Some DL models were also used to predict stock prices (e.g., [Kraus and Feuerriegel 2017](#); [Minh et al. 2017](#); [Jiang et al. 2018](#); [Matsubara et al. 2018](#)). For instance, [Schumaker and Chen \(2010\)](#) make a stock market forecasting based on financial news articles using a text classification approach. [Glasserman et al. \(2020\)](#) study using the supervised Latent Dirichlet Allocation (sLDA) framework to select news articles topics to explain stock returns.

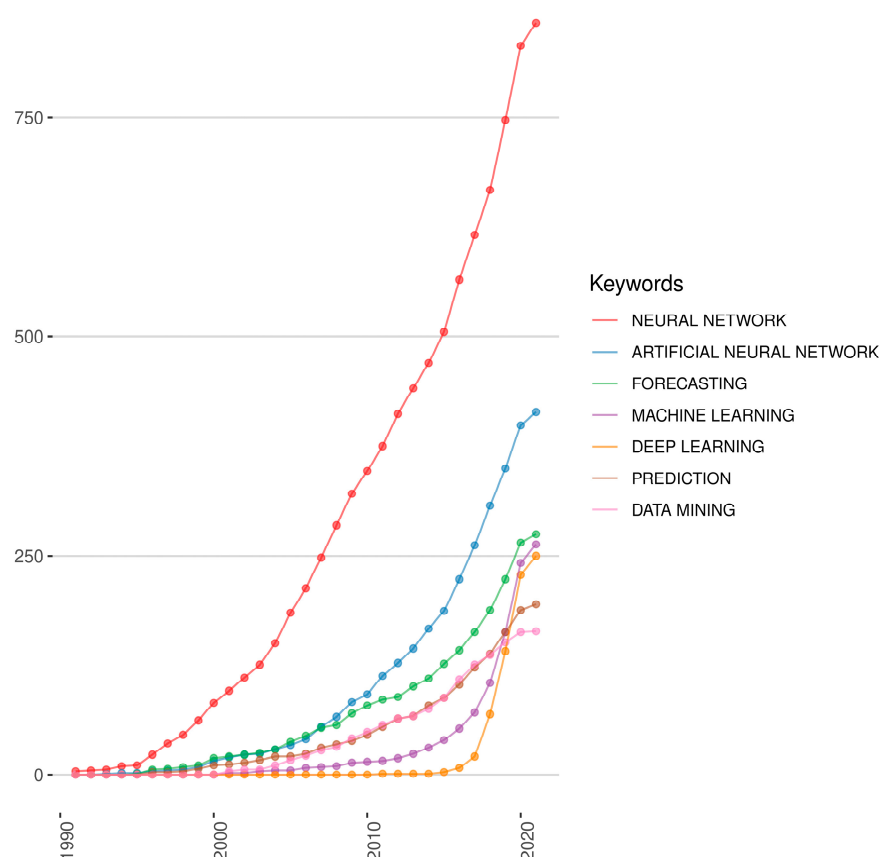
The network analysis has been used more in the context of financial stability analysis and financial linkages. Another strand of the literature examines the impact of views and opinions of investors-also known as investor sentiment-on stock price movements. The sentiment analysis aims to capture news by traditional and/or social media and assess the investors’ views and market mood (e.g., [Mitra and Mitra 2011](#); [Mitra and Yu 2016](#)). The assessment of market sentiment-often captured by market indices-can be strengthened by sentiment analysis of the market mood or investors’ emotions. A popular approach is to extract relevant news articles, preprocess the text, and assign a sentiment score to each article. The sentiment score is then commonly calculated as the difference between the number of positive and negative words in the article divided by the total number of words. The studies use a reputable lexicon of financial terms-such as [Loughran and McDonald \(2011\)](#) lexicon-to determine positive and negative words.

In the following sub-sections, we will consider the conceptual structures of our sample by looking at the keywords, the keywords co-occurrences, and the evolution of the topics based on a topic modeling technique.

#### 3.1. Keywords Analyses

We consider here the entire words that we find in the keyword section of every document. Remember that the sample was created using “neural network\*” AND “finance\*” (see [Figure 2](#)). It is thus expected that authors would again put neural networks as keywords in the keyword section. They will also associate other keywords such as prediction, forecasting, or machine learning, including deep learning. This is evidence that our sample goes beyond just neural networks but also covers other related topics.

It is interesting to see that deep learning is a very recent addition to the fintech field, as approximated by our sample. It is also interesting to notice that it is recently that the reasons why we would use the new techniques in finance have appeared, for instance, the role of these new methodologies in prediction. Machine learning techniques are indeed a paradigm shift when it comes to their predictive power.



**Figure 2.** Keywords count through time.

Table 5 represents the top keywords in the overall sample and the top keywords per year. It is interesting to see keywords ranking through time and how the literature has evolved in machine learning ownership and maturity, with deep learning papers moving up the ladder.

To go beyond a single-dimensional perspective of the keywords, let us look now at the co-occurrences matrix.

### 3.2. Keywords Co-Occurrences Network Analyses

Now, we are interested in looking at the keywords co-occurrences. When a keyword is used, it is possible to build a count matrix and compute its relationships with other keywords. From there, we can compute some relevant network indicators (centrality, density, etc.). Several figures will plot the relevance degree (centrality, or notions of ‘importance’) against the development degree (density). Degree centrality counts the number of links held by each node and points at themes that can easily connect with the broader network. The density of a network is the frequency of realized edges relative to potential edges.

In Figure 3, we represent the graphs based on the network indicators. The first figure is the network of keywords for the entire sample, while each other graph represents a network for 2021, 2020, 2019, 2018, and 2017, respectively.

When we consider the co-occurrences networks, particularly the years 2021 and 2017, we observe that most of the conversations are organized around two groups, representing both computer techniques and mathematical approaches. Only recently, applications in finance are starting to appear, such as the prediction of bankruptcies.

In Table 6, we compute the mathematical features of the networks. We observe that the size of the networks has been on the rise in the past years, showing an increase in the spread of the concepts. It is accompanied by a decrease in density through time with a

slight increase in the average path length, confirming potentially that the literature opens up to applications.

**Table 5.** Top keywords, overall period, and per year.

Author Keywords (DE)	Articles	Keywords-Plus (ID)	Articles
<b>Overall Time Period</b>			
Neural Network	867	Neural Networks	800
Artificial Neural Network	423	Prediction	482
Forecasting	277	Model	402
Machine Learning	274	Neural Network	340
Deep Learning	257	Classification	305
<b>2021</b>			
Neural Network	26	Neural Networks	13
Artificial Neural Network	22	Model	12
Forecasting	21	Prediction	10
Machine Learning	15	Market	8
Deep Learning	10	Classification	7
<b>2020</b>			
Deep Learning	87	Neural Networks	81
Neural Network	85	Prediction	66
Machine Learning	79	Model	63
Artificial Neural Network	49	Neural Network	50
Forecasting	42	Models	40
<b>2019</b>			
Neural Network	80	Neural Networks	96
Deep Learning	72	Prediction	51
Machine Learning	58	Model	49
Artificial Neural Network	43	Neural Network	38
Forecasting	35	Classification	36
<b>2018</b>			
Neural Network	51	Neural Networks	83
Deep Learning	48	Prediction	44
Artificial Neural Network	45	Model	42
Machine Learning	35	Classification	26
Forecasting	25	Neural Network	25
<b>2017</b>			
Neural Network	51	Neural Networks	68
Artificial Neural Network	39	Prediction	38
Forecasting	21	Model	34
Prediction	20	Neural Network	31
Machine Learning	18	Classification	30

### 3.3. Topic Modeling-Based Analyses

In the following analysis, we will add a new dimension based on structural topic modeling. The goal here is to complement the information we obtained from the keywords co-occurrences. A structural topic modeling first means that we will leverage words including the keywords section and beyond: the title section, the abstract, and the keyword + section.

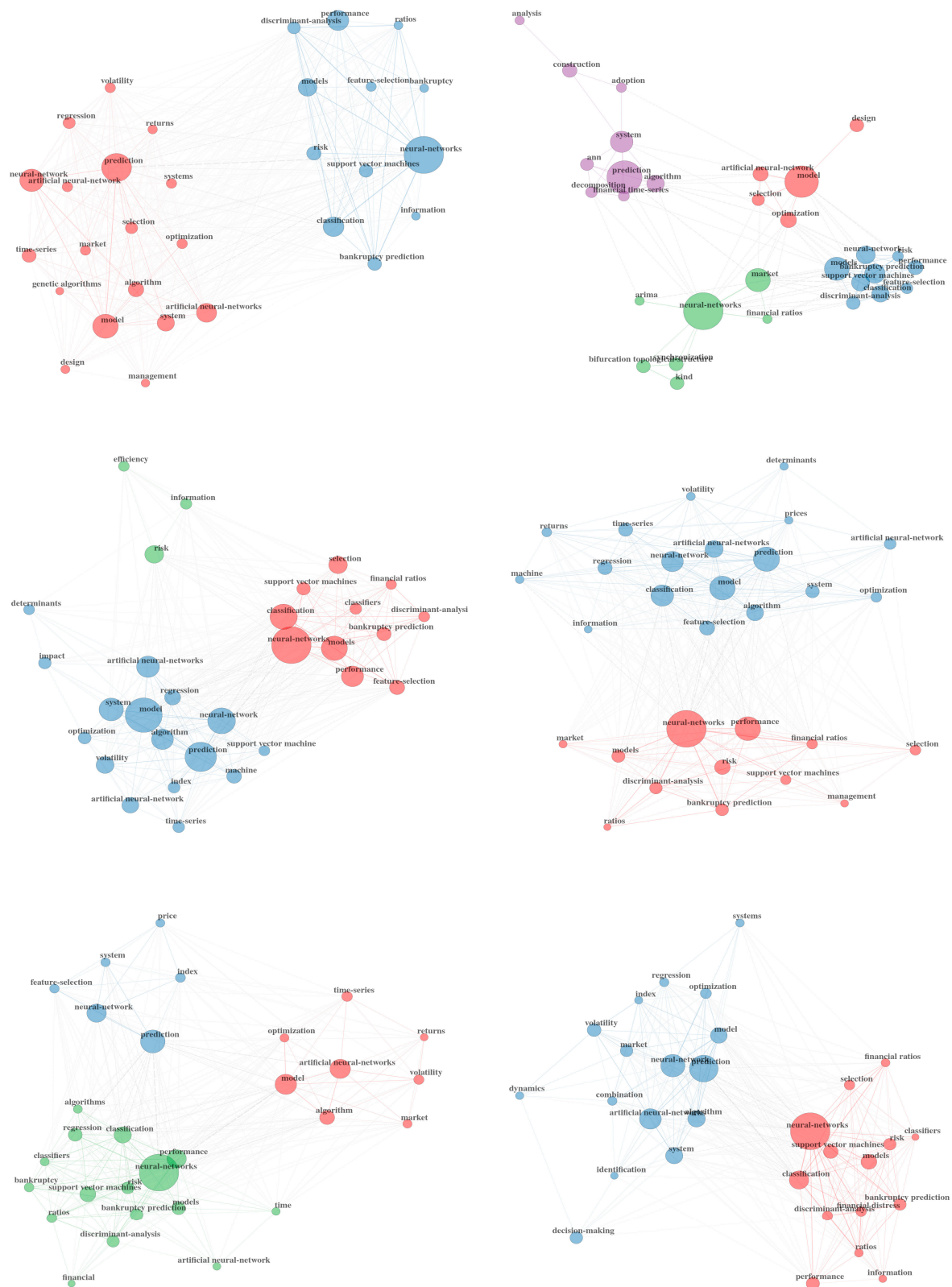
We tokenize all the words, and we compute the latent variables to identify potential topics.

In the following figures, we represent this analysis. The top-left figure covers the whole period, while the other figures represent each year, 2021, 2020, 2019, 2018, and 2017, respectively.

**Table 6.** Graph indicators, overall period, and per year.

Statistics	Overall Time Period	2021	2020	2019	2018	2017
Size	3607.000	234.000	950.000	849.000	693.000	604.000
Density	0.005	0.036	0.014	0.016	0.018	0.021
Transitivity	0.128	0.538	0.238	0.232	0.266	0.269
Diameter	6.000	6.000	6.000	6.000	6.000	6.000
Degree Centralization	0.298	0.188	0.229	0.303	0.317	0.333
Average path length	2.752	3.067	2.792	2.716	2.732	2.682





**Figure 3.** Network of authors' keywords, overall period, and per year.

We found the topics mapped in four dimensions: basic themes, emerging or declining, niche themes, and motor themes.

Interestingly, data mining and neural networks were part of the fundamental themes in 2017 (see Figure 4). Since we consider mostly finished documents in our sample, it means the work from the researchers has started a bit earlier, likely one or two years before.

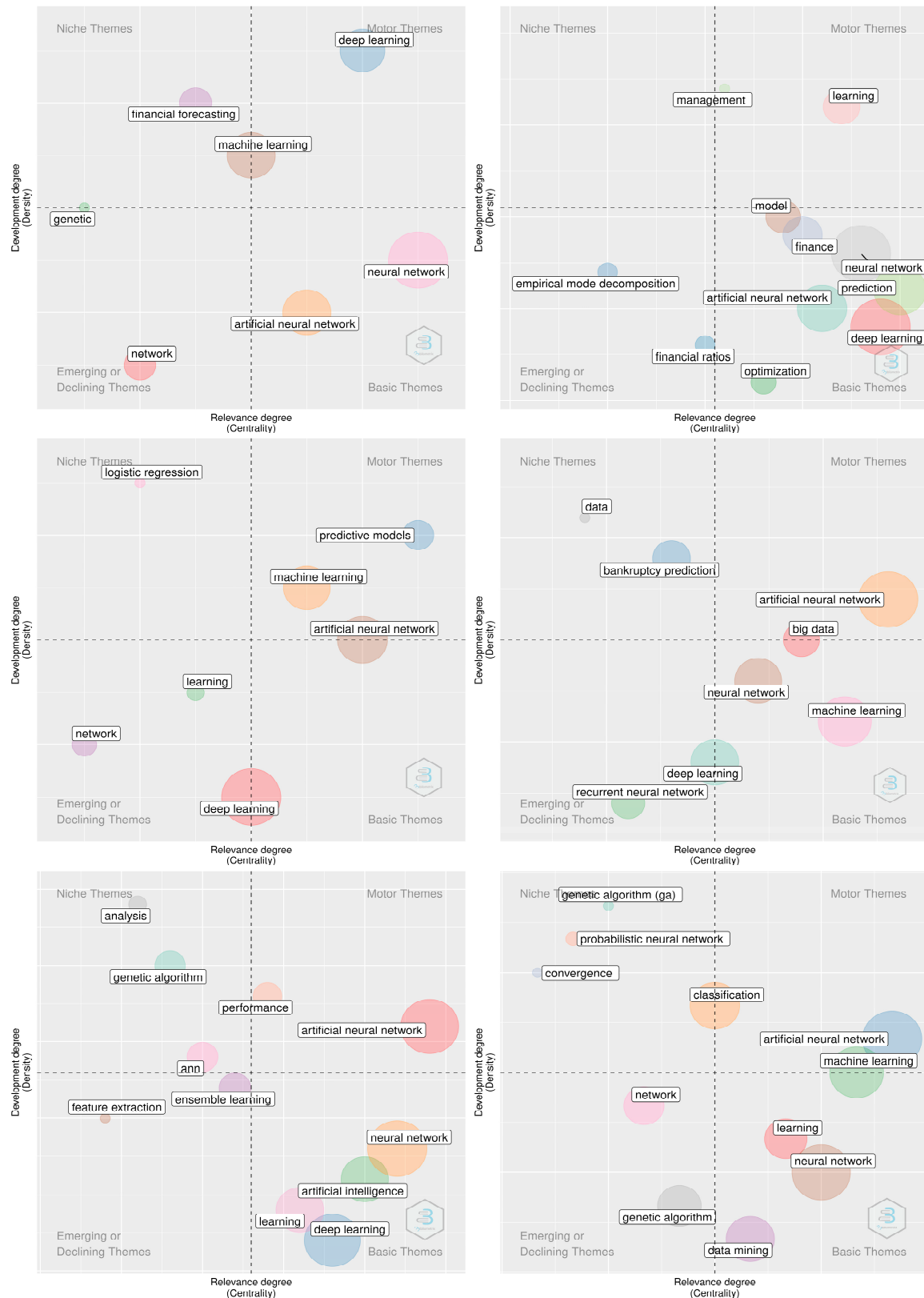


Figure 4. Topic modeling, overall period, and per year.

In 2017, a generic algorithm was an emerging theme as well as network theory. We see here a burgeoning reflection about what will become the contribution from data science in finance. Comparing 2017 and 2020, and 2021, it is interesting to see that the motor themes are about the predictive capacity of machine learning-based models. We can also observe the emerging sub-field of deep learning in finance. We can easily extrapolate and imagine that deep learning in finance will have a prominent future in the field.

We want to insist on the inductive nature of machine learning: it is inductive by nature but does not come with the former empirical baggage of being potentially biased and lacking theoretical grounds (the falsification potential, etc.). Inductive in the context of ML implies finding causal patterns in empirical data.

#### 4. Intellectual Structures of Our Sample

An interesting analysis stems from the investigation of which authors and organizations are driving the dynamics of this topic.

##### 4.1. Authors

In the intellectual structure, authors are interesting to consider. We can see that the top authors have published more than 30 papers on this topic in our sample (Figure 5).

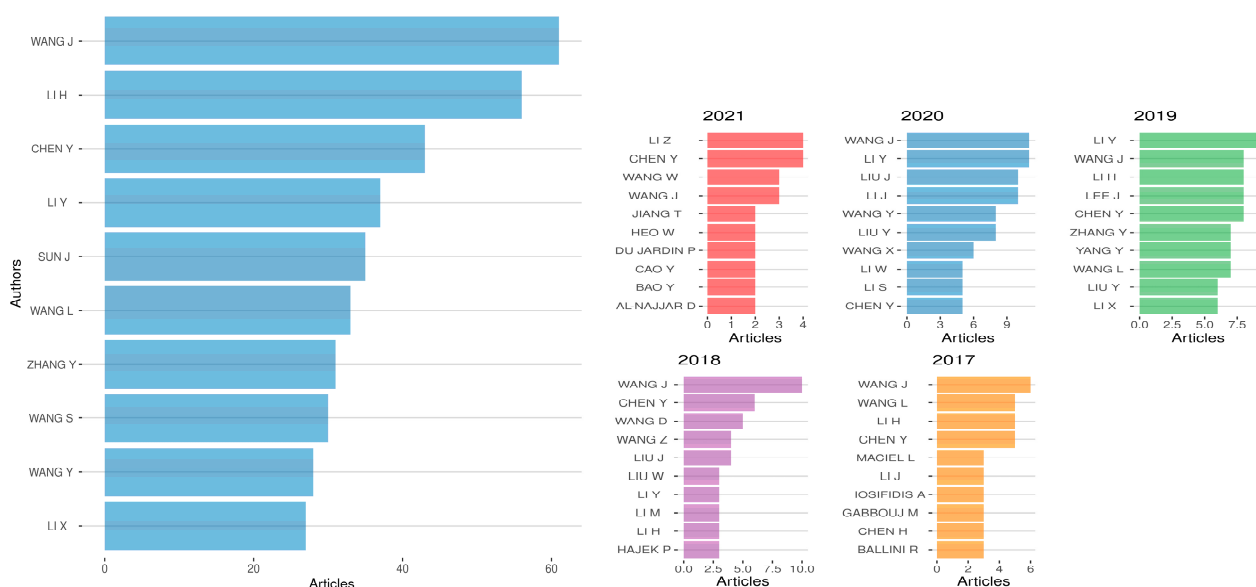


Figure 5. Top authors in terms of production, overall period, and per year.

We can go a little deeper and look at the average productivity of all the authors (see Figure 6). It has not evolved much through time, and on average, every author produces two articles a year on this topic.

We can also look at the authors' dominance ranking through time (see Figure 7). The authors' dominance is computed by looking at how many times an author is a first author in a multi-authored paper. It can be a weak indicator as the alphabetical order is respected most of the time, irrespective of the marginal contributions, as assumed by this indicator.

Interestingly, it is interesting to see that authors unfavored by the alphabetical order, such as Zhang or Wang, are still making the top 10 of this ranking.

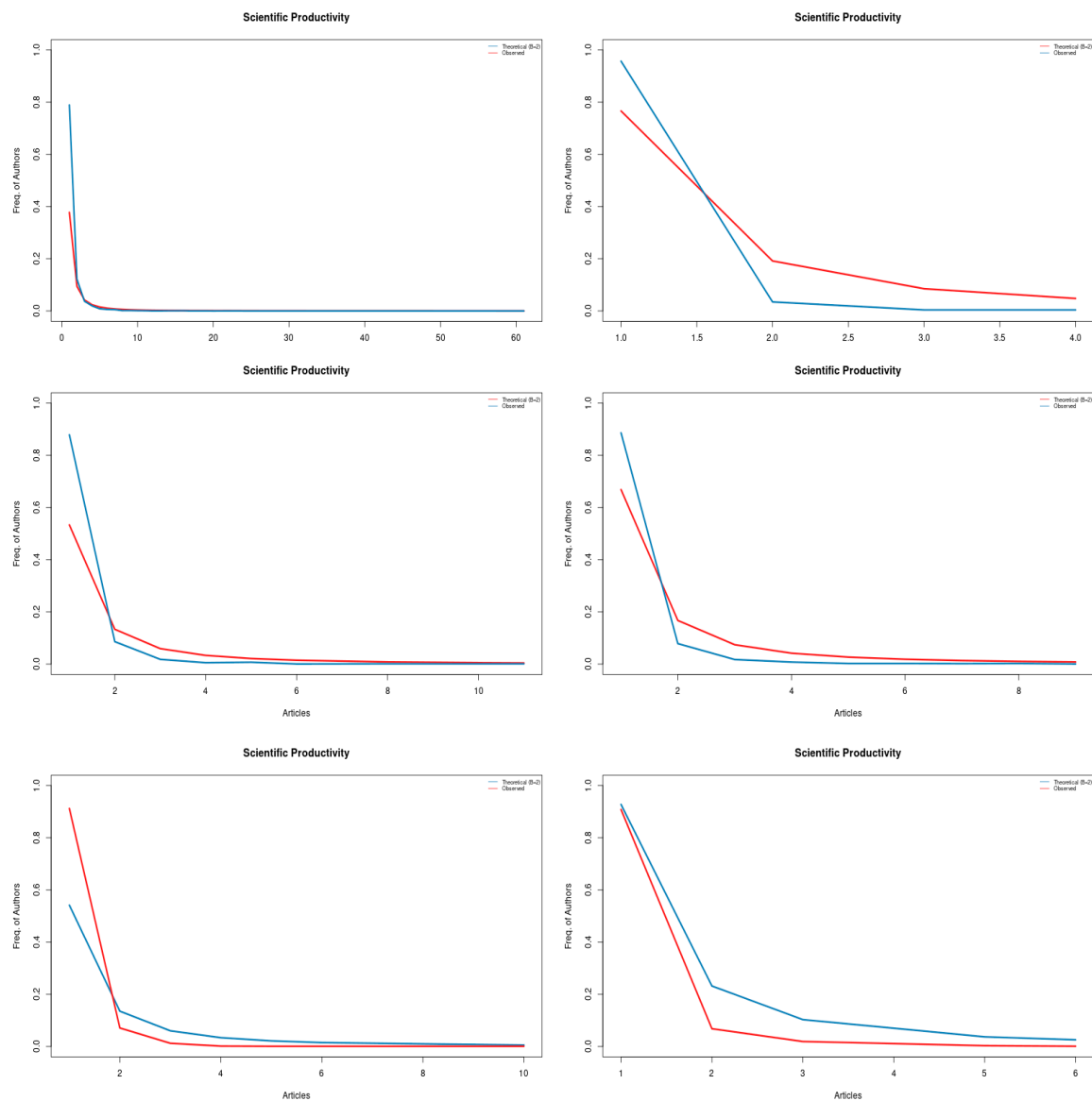


Figure 6. Scientific productivity, overall period, and per year.

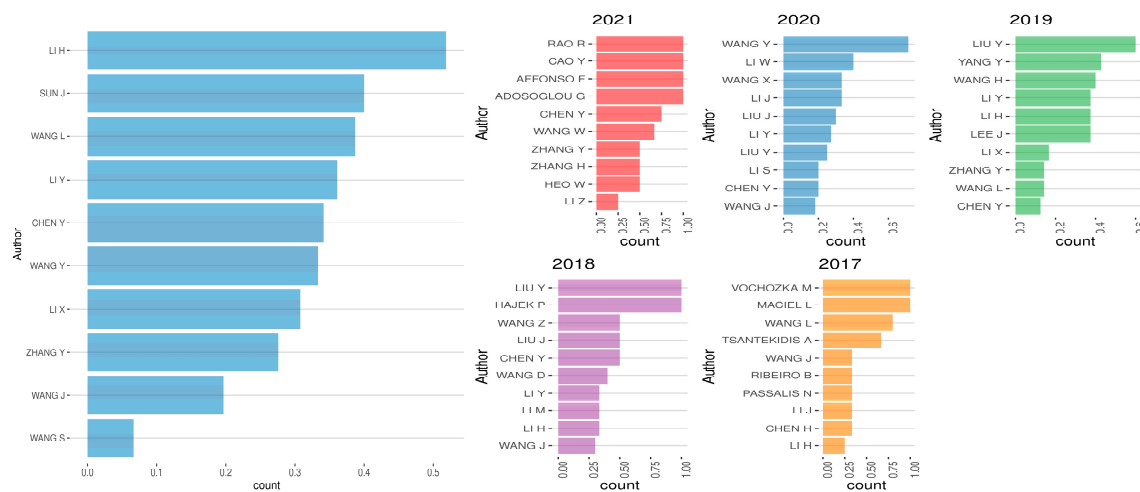


Figure 7. Author dominance ranking, overall period, and per year.

#### 4.2. Articles

Table 7 illustrates the citations of the articles in our sample.

**Table 7.** Most cited manuscripts, overall period, and per year.

Article	Total Citations	Total Citations per Year	NTC
<b>Overall Time Period</b>			
Schaap Mg., 2001, J Hydrol	1361	64.8	20.06
Jordan Mi, 2015, Science	1189	169.9	78.27
Kim Kj, 2003, Neurocompeting	748	39.4	18.34
Pan Wt, 2012, Knowledge-Based Syst	725	72.5	33.93
Tay Feh, 2001, Omega-Int H Manage Sci	596	28.4	8.79
<b>2017</b>			
Wei, Y, 2017, Ieee Trans Pattern Anal Mach Intell	199	39.8	18.25
Bao W, 2017, Plos One	198	39.6	18.16
Deng Y, 2017, Ieee Trans Neural Netw Learn Syst	142	28.4	13.03
Barboza F, 2017, Expert Syst Appl	135	27.0	12.38
Krauss C, 2017, Eur J Oper Res	115	23.0	10.55
<b>2018</b>			
Fischer T, 2018, Eur J Oper Res	258	64.5	31.17
Termeh Svr, 2018, Sci Total Environ	144	36.0	17.40
Han J, 2018, Proc Natl Acad Sci USA	129	32.2	15.58
Kim Hy, 2018, Expert Syst Appl	108	27.0	12.38
Cai Y, 2018, Remote Sens Environ	102	25.5	12.32
<b>2019</b>			
Altan A, 2019, Chaos Solitons Fractals	90	30.0	17.98
Cao J, 2019, Physica A	60	20.0	11.99
Long W, 2019, Knowledge-Based Syst	55	18.3	10.99
Strubell E, 2019, 57th Annual Meeting of the Association for Computational Linguistics (ACI 2019)	48	16.0	9.59
Plawiak P, 2019, Appl Soft Comput	43	14.3	8.59
<b>2020</b>			
Pang X, 2020, J Supercomput	44	22.0	19.51
Akhtar Ms, 2020, Ieee Comput Intell Mag	41	20.5	18.18
Ahmed R, 2020, Renew Sust Energ Rev	38	19.0	16.85
Sezer Ob, 2020, Appl Soft Comput	32	16.0	14.19
Gu S, 2020, Rev Financ Stud	29	14.5	12.86
<b>2021</b>			
Marcelino P, 2021, Int J Pavement Eng	12	12	25.81
Talwar M, 2021, J Retail Consum Serv	8	8	17.21
Carta S, 2021, Expert Syst Appl	6	6	12.90
Brodny J, 2021, J Clean Prod	5	5	10.75
Hu Z, 2021, Appl Syst Innov	4	4	8.60

We can go a little further and look now at the articles that authors in our sample include in their references. As such, those references are the foundations of this nascent literature in machine learning in finance. Let us look at the top authors in the references of each paper (see Figure 8).

We can also look at the most cited references in terms of journals beyond their authors. The most cited authors and the most cited references will match, but it is interesting to see the nuances (see Figure 9).

It is interesting to note that the literature has not moved too much from the top papers from 2017 to 2021.

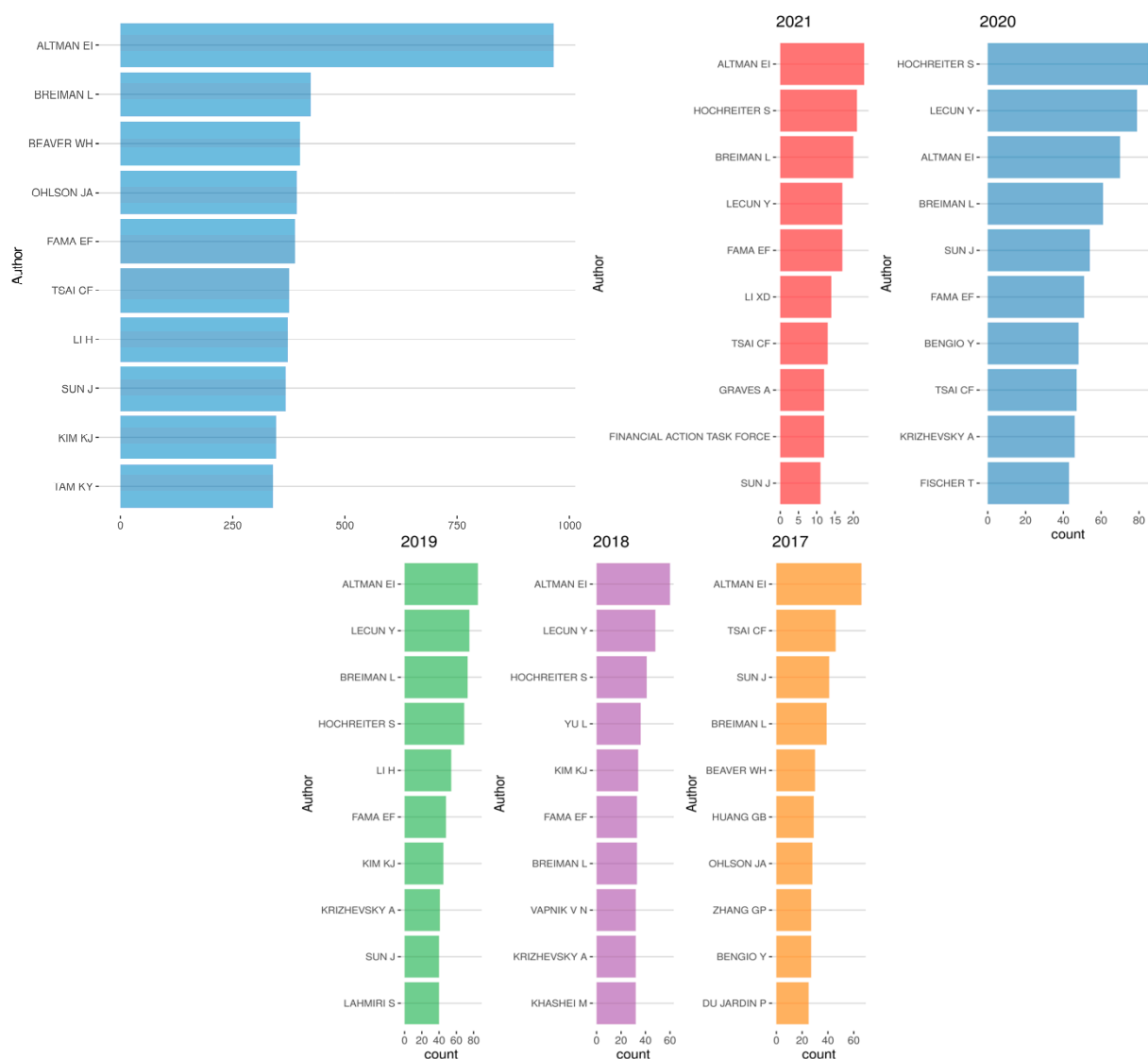


Figure 8. Analysis of cited references, overall period, and per year.



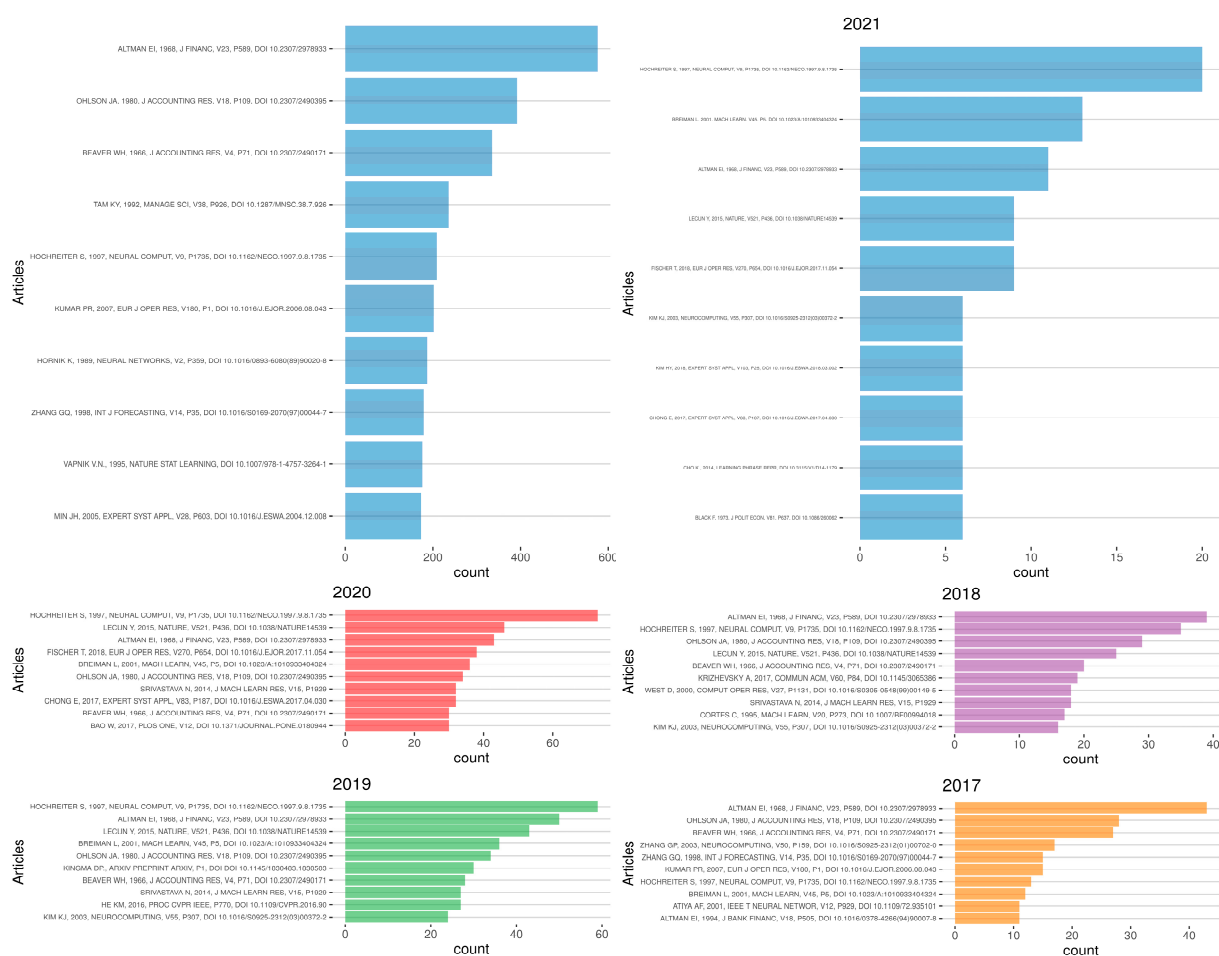


Figure 9. Most cited manuscripts, overall period, and per year.

## 5. Social Structures of Our Sample

In this section, we will spend time on different measures to capture the social connections: the co-citations of authors, the co-citations of articles, the co-citations of journals, and the collaborations across institutions.

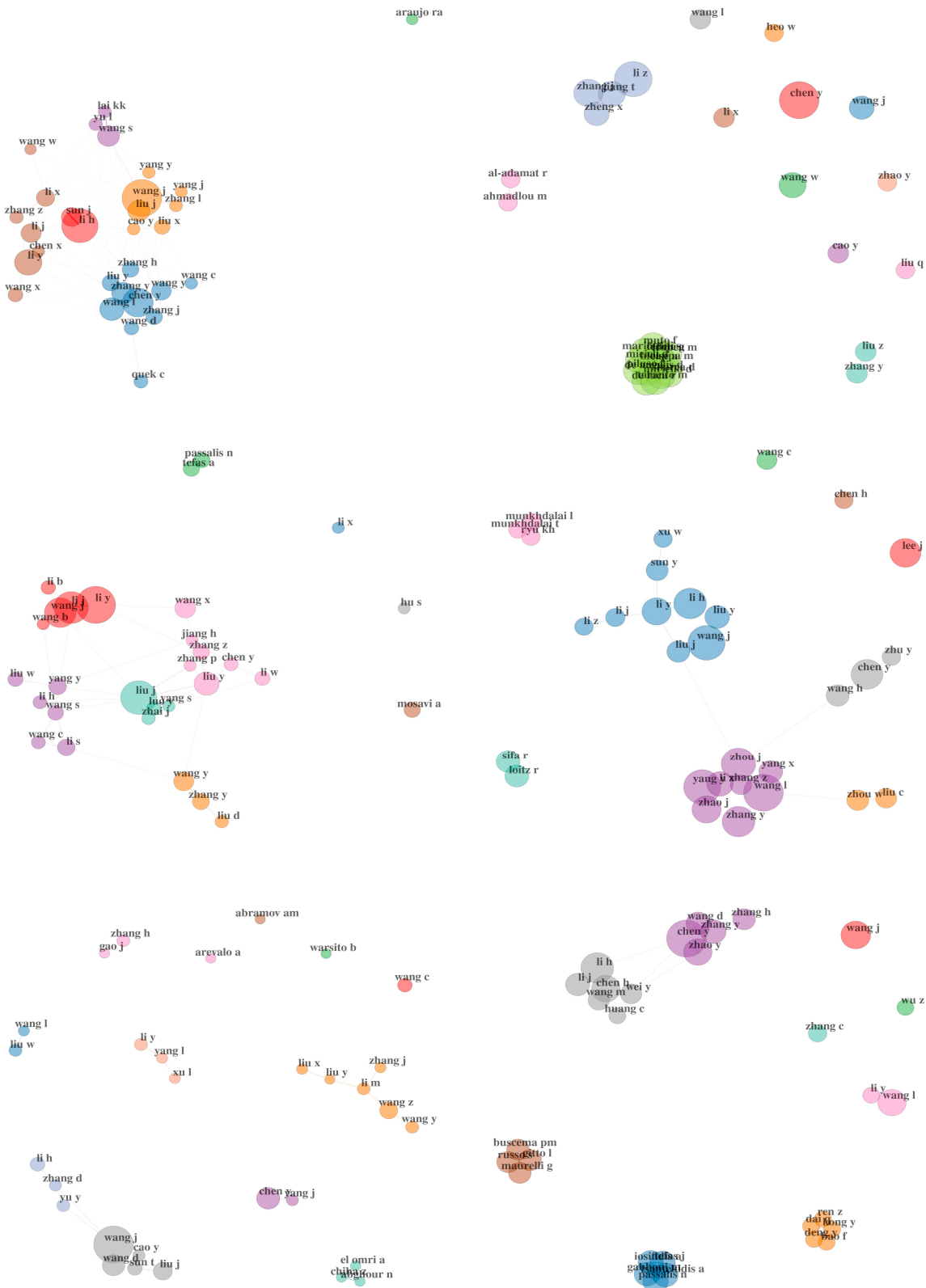
### 5.1. Co-Citations of Authors

Figure 10 highlights the evolution of authors' collaborations. We can observe that it is still a narrow network of collaborators. We are showing the nascent nature of the field. We represent here the network of the top authors.

As we can see in the previous figure, the top authors are still working nearby within their groups of collaborators. The next question is to know whether it is still the case for co-citations of articles.

### 5.2. Co-Citations of Articles

When a reference was addressed by two articles published in the same journal, this reference was included in the co-citation network of references (see Figure 11). Therefore, the co-citation network addressed the expected references to the concept of uncertainty in articles published by a journal.



**Figure 10.** Authors' collaboration networks, overall period, and per year.

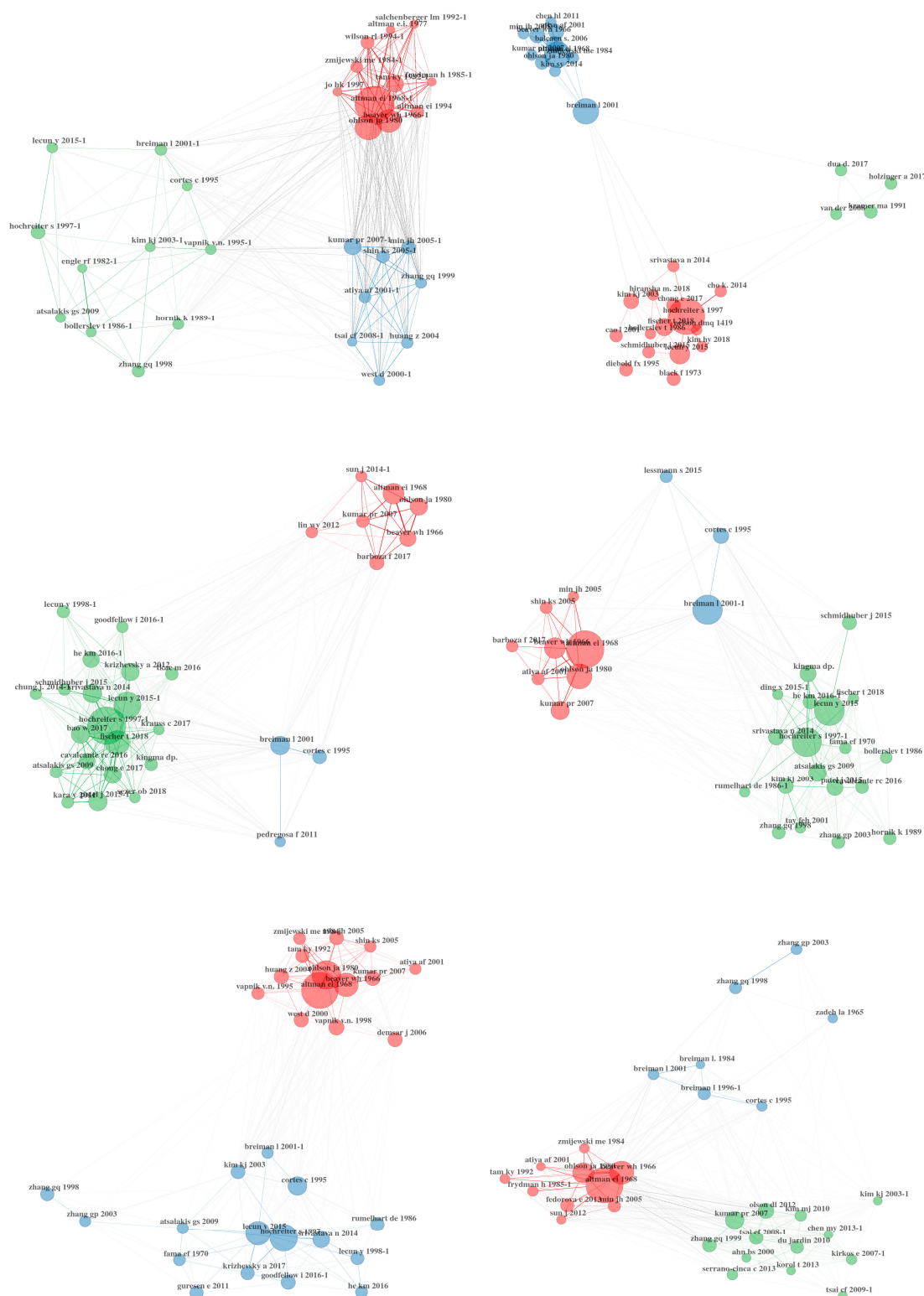


Figure 11. Co-citations of articles, overall period, and per year.

In our sample, most of the authors in finance are residents of the People's Republic of China, the United States, the United Kingdom, and India (see Table 8). While the dominant presence of authors from the advanced economies is undisputed, it is also noticeable that the law of large numbers ensures the participation of authors from several Emerging Market Economies (EMEs).

**Table 8.** Corresponding authors' countries, overall period, and per year.

Country	Articles	Frequency	SCP	MCP	MCP_Ratio
<b>Overall Time Period</b>					
China	1438	0.2885	1253	185	0.1287
United States	476	0.0955	389	87	0.1828
India	293	0.0588	268	25	0.0853
United Kingdom	256	0.0514	195	61	0.2383
Brazil	147	0.0295	138	9	0.0612
<b>2017</b>					
China	90	0.2535	74	16	0.1778
India	36	0.1014	33	3	0.0833
United States	28	0.0789	20	8	0.2857
Iran	18	0.0507	16	2	0.1111
Brazil	12	0.0338	11	1	0.0833
<b>2018</b>					
China	106	0.2437	89	17	0.1604
India	35	0.0805	32	3	0.0857
United States	34	0.0782	22	12	0.3529
Iran	18	0.0414	15	3	0.1667
Turkey	16	0.0368	15	1	0.0625
<b>2019</b>					
China	172	0.2976	136	36	0.2093
United States	55	0.0952	48	7	0.1273
India	36	0.0623	33	3	0.0833
Russia	23	0.0398	22	1	0.0435
Spain	19	0.0329	9	10	0.5263
<b>2020</b>					
China	177	0.2990	147	30	0.169
India	44	0.0743	35	9	0.205
United States	43	0.0726	34	9	0.209
United Kingdom	29	0.0490	20	9	0.310
Iran	21	0.0355	18	3	0.143
<b>2021</b>					
China	53	0.3397	42	11	0.208
India	13	0.0833	13	0	0.000
United States	9	0.0577	6	3	0.333
Italy	7	0.0449	7	0	0.000
Turkey	7	0.0449	6	1	0.143

Note: SCP = single country publications; MCP = multiple country publications; MCP\_Ratio = share of multiple country publications in the total number of publications.

Table 9 provides Supplementary Materials on the total citations per country. Asia and China, in particular, dominate the ranking.

Figure 12 shows an apparent increase in the contributions coming from Asia: China and India being at the forefront of academic production.

Starting from a bibliographic matrix, two groups of descriptive measures are computed: (1) the summary statistics of the network and (2) the leading indices of centrality and prestige of vertices.

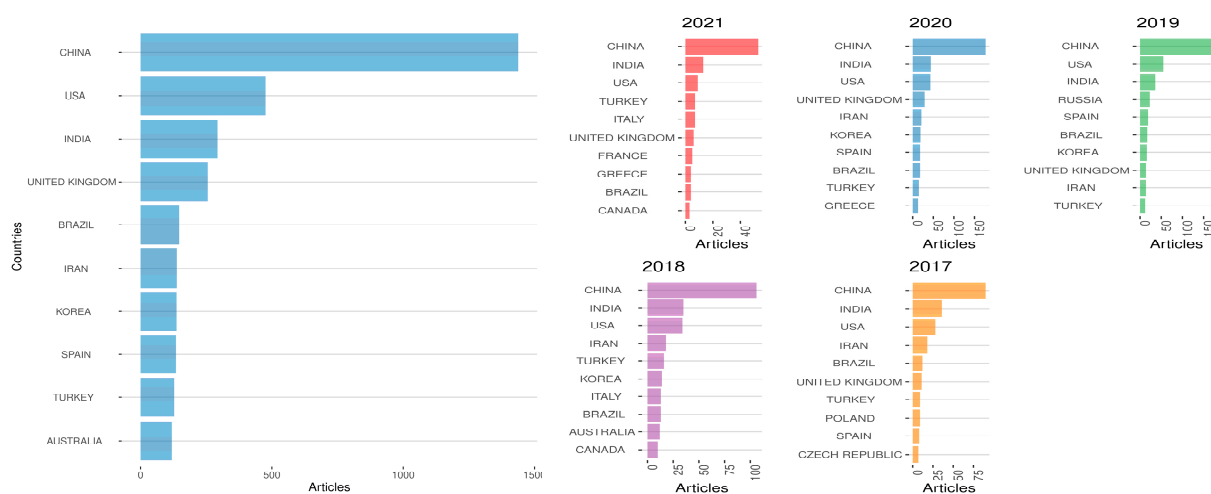
This group of statistics presented in Table 8 allows us to describe the structural properties of a network: (1) 'size': is the number of vertices composing the network; (2) 'density': is the proportion of present edges from all possible edges in the network; (3) 'transitivity' is the ratio of triangles to connected triples; (4) 'diameter' is the longest geodesic distance (length of the shortest path between two nodes) in the network; (5) 'degree distribution' is the cumulative distribution of vertex degrees, and (6) 'degree centralization' is the normalized degree of the overall network.

When it comes to countries' collaborations, China and the USA are at the center of the graph (see Figure 13). Most of the international collaborations are between China and the USA. There seems to be a slight regionalization of collaborations, China with Asian countries, though it is much less apparent in the case of the USA, which seems to be a bit more eclectic in terms of collaborations.

**Table 9.** Total citations per country, overall period, and per year.

Country	Total Citations	Average Article Citations
<b>Overall Time Period</b>		
China	17154	11.929
United States	16876	35.454
United Kingdom	4691	18.324
South Korea	4482	32.715
India	2999	10.235
<b>2017</b>		
China	1413	15.70
United States	463	16.54
India	404	11.22
Brazil	260	21.67
Germany	207	34.50
<b>2018</b>		
United States	555	16.324
China	511	4.821
Iran	285	15.833
Germany	270	54.000
India	232	6.629
<b>2019</b>		
China	607	3.529
United States	421	7.655
Brazil	165	9.706
Iran	132	9.429
South Korea	126	7.875
<b>2020</b>		
China	352	1.989
United States	127	2.953
India	107	2.432
United Kingdom	72	2.483
Australia	63	5.727
<b>2021</b>		
China	13	0.245
Portugal	12	12.000
Norway	9	3.000
India	7	0.538
Italy	6	0.857

Note: SCP = single country publications; MCP = multiple country publications; MCP\_Ratio = share of multiple country publications in the total number of publications.

**Figure 12.** The most productive countries (according to authors' residence).

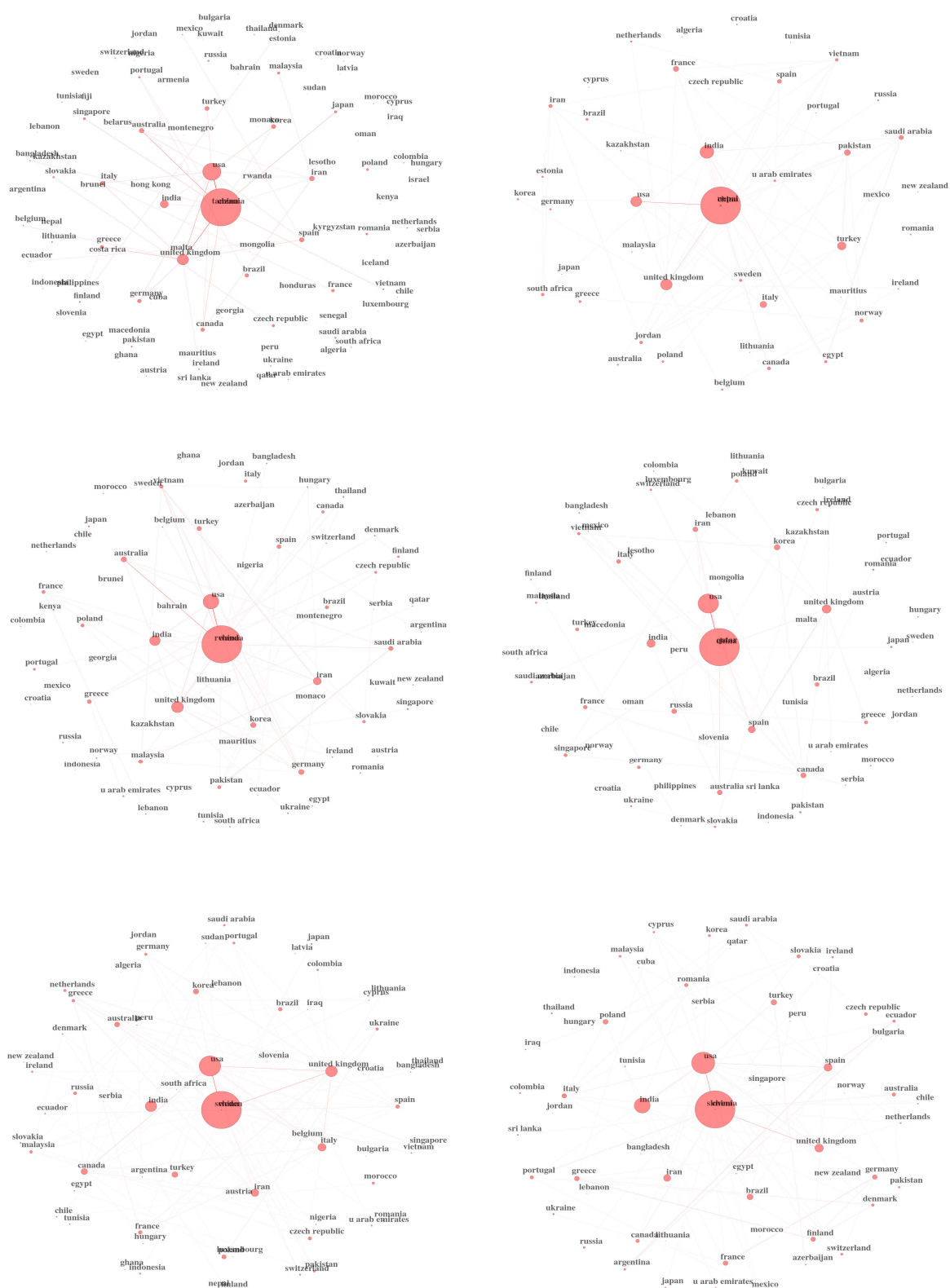


Figure 13. Country collaboration networks, overall period, and per year.

Considering the results mentioned above, it confirms that Asia and China are somehow at the forefront of the academic production on neural networks and the larger machine learning domain in finance. It is interesting to the connections with other countries, notably in Europe. Below, we will also investigate the connections at the institutional level.



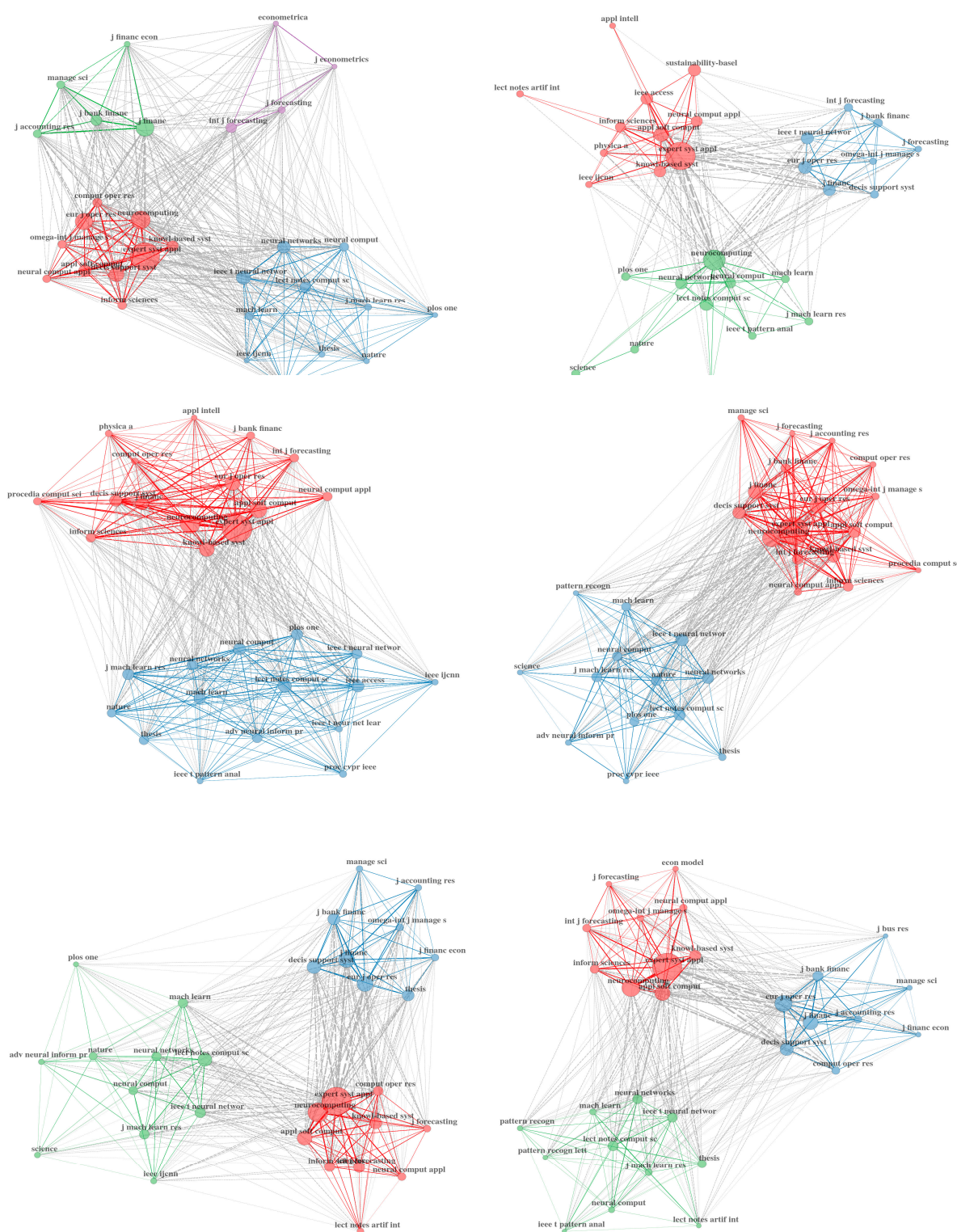
### 5.3. Co-Citations of Journals

We will look at which journals have contributed to developing the field's methodological transformation in what follows. Through time (see Table 10), we will see that it mostly started in more engineering journals to penetrate the finance field. Still, nowadays, the ranking is dominated by more engineering-oriented journals.

**Table 10.** Top journals, overall period, and per year.

Sources	Articles
<b>Overall Time Period</b>	
Expert Systems with Applications	305
Applied Soft Computing	75
Ieee Access	74
Neurocomputing	71
Neural Computing & Applications	56
<b>2017</b>	
Expert Systems with Applications	12
Applied Soft Computing	6
Physica a-Statistical Mechanics and Its Applications	5
2017 Ieee International Conference on Big Data (Big Data)	4
Agro Food Industry High-tech	4
<b>2018</b>	
Expert Systems with Applications	12
Applied Soft Computing	9
Neurocomputing	8
2018 26th Signal Processing and Communications Applications Conference (Sui)	7
2018 International Joint Conference on Neural Networks (ijcnn)	7
<b>2019</b>	
Ieee Access	24
Expert Systems with Applications	19
Physica a-Statistical Mechanics and Its Applications	11
Sustainability	11
Applied Soft Computing	9
<b>2020</b>	
Ieee Access	37
Expert Systems with Applications	17
2020 International Joint Conference on Neural Networks (ijcnn)	13
Soft Computing	13
Neural Computing & Applications	11
<b>2021</b>	
Ieee Access	10
Expert Systems with Applications	8
Computational Economics	5
Annals of Operational Research	4
Complexity	4

Figure 14 is an excellent illustration of the evolution of the knowledge map seen through journal co-citations. It is interesting to see the origin of the transformation and the pace of the penetration of machine learning in finance journals and through which channels. It is worth noticing the pivotal role played by the “*Expert Systems with Applications*” journal.



**Figure 14.** Journals source co-citation analysis, overall period, and per year.

#### 5.4. Co-Citations of Institutions

Related to Figure 13, it is interesting to study the collaborations through a different indicator: the co-citations of institutions.

The network of university collaboration is also well developed (see Figure 15), indicating a strong presence of Chinese, U.S., and Indian universities. It is interesting to notice a

[illegible]

To conclude, we visualize the main items of three fields (e.g., authors, keywords, journals) and how they are related through a so-called Sankey diagram. The three fields

plot in Figure 16 also reveals the rising importance of deep learning and neural networks in finance and its most robust channel for articulating academic contributions, the *Experts Systems with Applications Journal* for the overall period, and IEEE Access for most of the latest five years.

In the past five years, *IEEE Access* has been a prominent vehicle for developing the academic conversation on neural networks in finance and, most importantly, deep learning in finance.

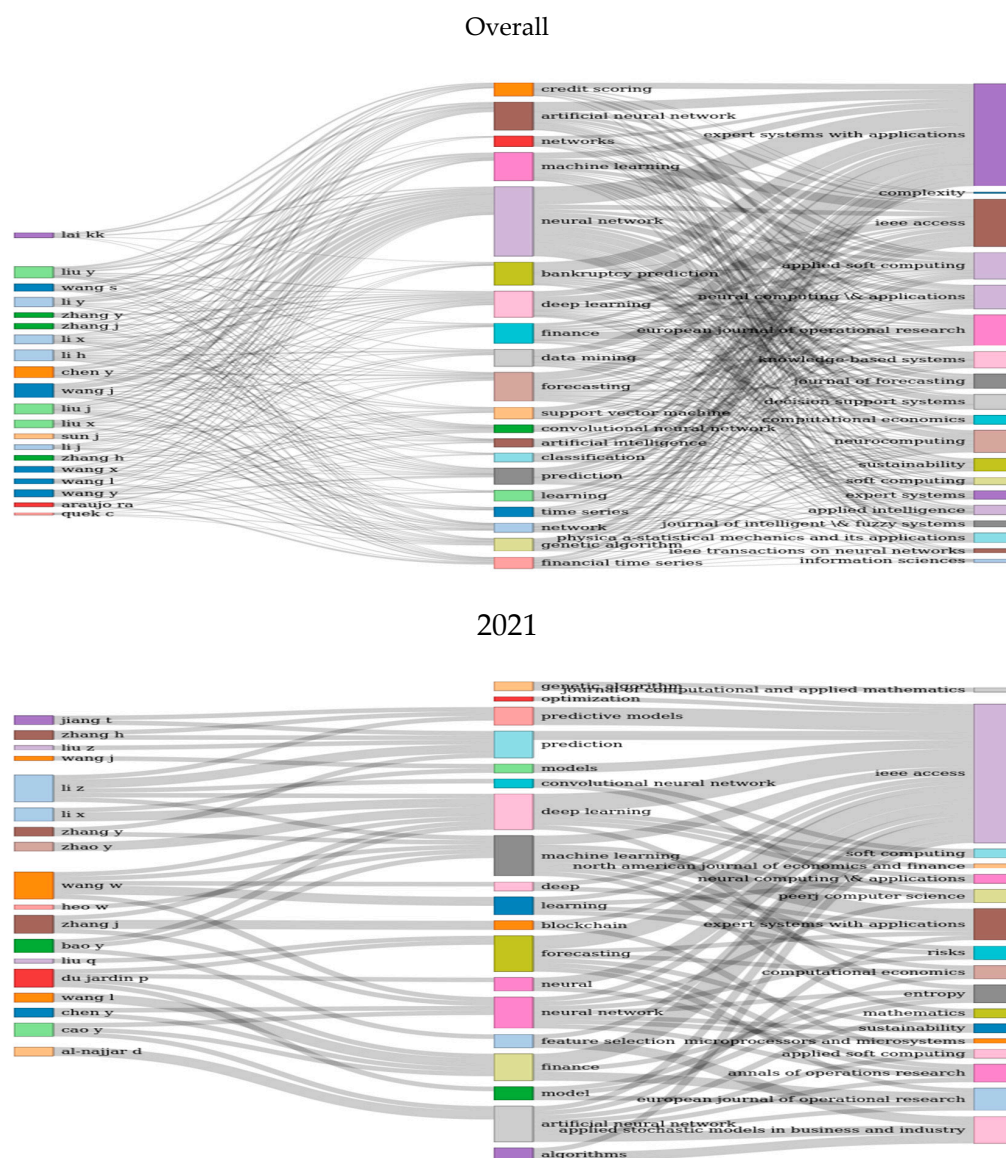
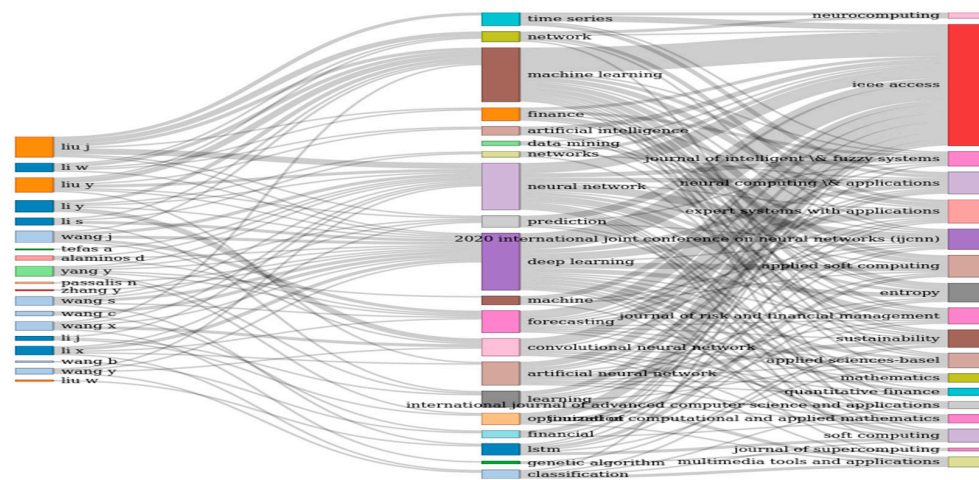


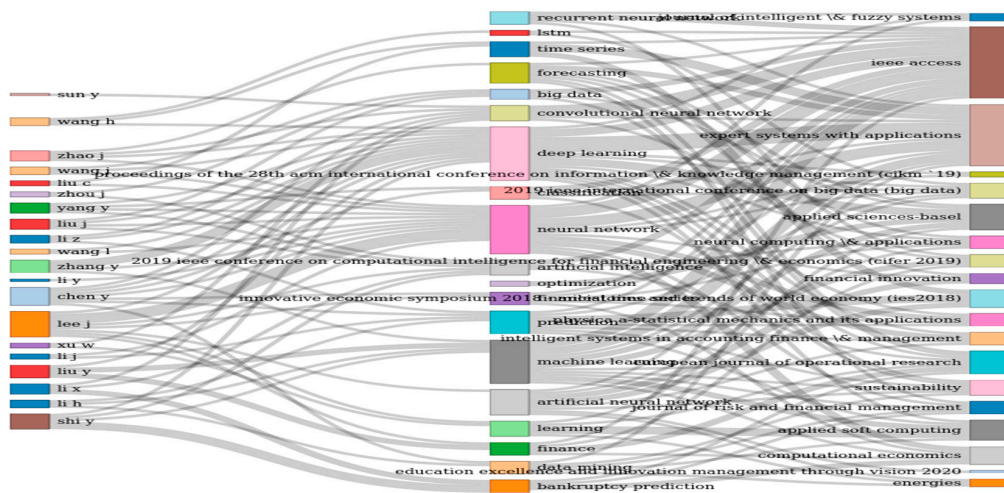
Figure 16. Cont.



2020



2019



2018

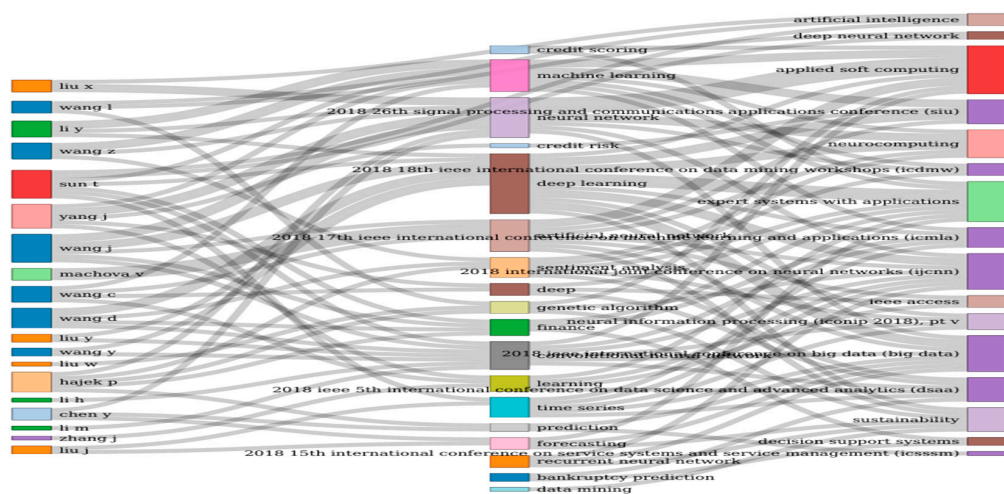


Figure 16. Cont.

2017

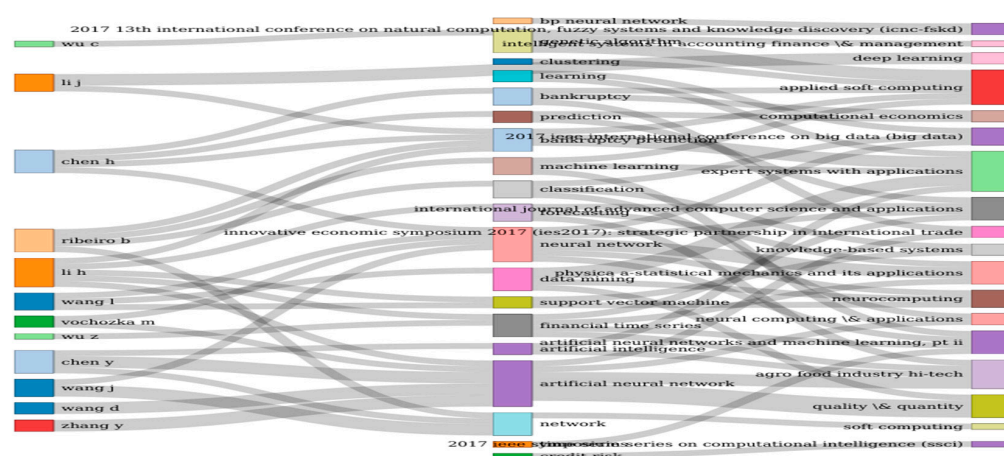


Figure 16. Three fields plot, overall period and per year.

## 6. Conclusions

Neural networks in finance are becoming increasingly popular tools to analyze financial market trends based on preprocessing and transforming a large amount of information into machine-readable data. It would be a mistake to attribute this development solely to the outstanding computing power and storage capacity growth.

ML can make essential contributions to the technical analysis of financial market trends. It has a wide variety of applications: supervised, unsupervised, and semi-supervised learning; reinforcement learning; inverse reinforcement learning; imitation learning; self-learning; feature learning; sparse dictionary learning; anomaly detection, etc. A subfield at the intersection of linguistics, computer science, and artificial intelligence—Natural Language Processing (NLP)—has found numerous applications in finance.

This article demonstrated the basic steps required to conduct a metadata-based SLR in the finance field.

The method can help generate topic-specific existing knowledge, trends, and gaps observed and the derivation of a conclusion suitable for policymakers and the scientific community.

Indeed, in this article, we conducted a metadata-based systematic review of the academic contributions to finance between 1990 and 2021. A metadata-based systematic literature review complements more conventional approaches to systematic literature reviews. It allows to collect more significant amounts of documents and then analyze the current dynamics within the collected documents. This article leverages the text information found in this dataset. Titles, abstracts, keywords, authors' names, institutions, and references are transformed into quantitative indicators. From there, using text-as-data techniques such as NLP as well as graph theory, we could provide a mapping capturing multiple dimensions. In particular, we used a theoretical framework that organizes the literature's mapping through three dimensions: conceptual, intellectual, and social. Beyond this mapping, we also used two techniques to deal with the data: NLP and graph theory.

The results are a mapping of the literature through these three dimensions. Researchers can use this mapping to select a sub-sample to perform the systematic literature review of their choice.

This mapping is helpful for researchers, university administrators willing to understand the evolution of the finance field, and policymakers. Concerning the latter, the conversation in academic circles about machine learning in finance finds its parallel in the financial industry with the development of the so-called fintech. It is relevant to map col-



laboration networks both at the authors' level and the institutional level for policymakers. It is also relevant to be able to visualize the knowledge maps.

For further research, the appearance of artificial intelligence and machine learning, in particular in finance, is quite attractive in the context of the old-time debate between the theorists and the chartists. While the opposing theorists and chartists debate is still relevant, we conjecture that ML techniques could shed some new light on theoretical advancement. MLAs are not an atheoretical approach, as it is premised on inductive reasoning, which generates causal relationships based on the state of information at the moment of estimation. The main advantage of ML is the ability to process vast information, simultaneously ignoring ideological standpoints or inclinations to a particular school of thought.

**Author Contributions:** Conceptualization, T.W. and A.S.; methodology, T.W.; software, T.W.; validation, T.W. and A.S.; formal analysis, T.W. and A.S.; investigation, T.W.; resources, T.W.; data curation, T.W.; writing—original draft preparation, A.S.; writing—review and editing, A.S. and T.W.; visualization, T.W.; supervision, T.W.; project administration, A.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** The authors express their deep gratitude to CIRANO (Montreal, Canada), Martin Paquette (CIRANO), Marine Leroi (CIRANO), and Aïchata Kone (HEC Montréal) for their excellent support. The usual caveats apply.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

### References

- Adcock, R., and N. Gradojevic. 2019. Non-fundamental, non-parametric Bitcoin forecasting. *Physica A: Statistical Mechanics and Its Applications* 531:121727.
- Atsalakis, G. S., I. G. Atsalaki, F. Pasiouras, and C. Zopounidis. 2019. Bitcoin price forecasting with neuro-fuzzy techniques. *European Journal of Operational Research* 276: 770–80.
- Bekiros, S. D., and D. A. Georgoutsos. 2008. Direction-of-change forecasting using a volatility-based recurrent neural network. *Journal of Forecasting* 27: 407–17.
- Blake, A. P., and G. Kapetanios. 2000. A radial basis function artificial neural network test for ARCH. *Economics Letters* 69: 15–23.
- Chung-Ming, Kuan, and Halbert White. 1994. Artificial neural networks: An econometric perspective, *Econometric Reviews* 13: 1–91.
- Cui, Herui, Ruoyao Wang, and Haoran Wang. 2020. An Evolutionary Analysis of Green Finance Sustainability Based on Multi-Agent Game. *Journal of Cleaner Production* 269: 121799.
- Donaldson, R., and M. Kamstra. 1997. An artificial neural network-garch model for international stock return volatility. *Journal of Empirical Finance* 4: 17–46.
- Falcone, Pasquale Marcello. 2020. Environmental Regulation and Green Investments: The Role of Green Finance. *International Journal of Green Economics* 14: 159–73.
- Fernandez-Rodriguez, F., C. Gonzalez-Martel, and S. Sosvilla-Rivero. 2000. On the profitability of technical trading rules based on artificial neural networks: Evidence from the Madrid stock market. *Economics Letters* 69: 89–94.
- Fischer, T., and C. Krauss. 2018. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research* 270: 654–69.

## References

- Garcia, R., and R. Gencay. 2000. Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics* 94: 93–115.
- Gencay, R., and T. Stengos. 1998. Moving average rules, volume and the predictability of security returns with feedforward networks. *Journal of Forecasting* 17: 401–14.
- Gerritsen, D. F., E. Bouri, E. Ramezanifar, and D. Roubaud. 2020. The profitability of technical trading rules in the Bitcoin market. *Finance Research Letters* 34: 101263.
- Gradojevic, N., and R. Gencay. 2013. Fuzzy logic, trading uncertainty and technical trading. *Journal of Banking and Finance* 37: 578–86.
- Gu, S., B. Kelly, and D. Xiu. 2020. Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies* 33: 2223–73.
- Hans, F. P., and van Griensven Kasper. 1998. Forecasting Exchange Rates Using Neural Networks for Technical Trading Rules. *Studies in Nonlinear Dynamics & Econometrics* 2: 1–8.
- Hsieh, D. A. 1989. Testing for nonlinear dependence in daily foreign exchange rates. *The Journal of Business* 62: 339–68.
- Huang, J.-Z., W. Huang, and J. Ni. 2019. Predicting Bitcoin returns using high-dimensional technical indicators. *The Journal of Finance and Data Science* 5: 140–55.
- Hutchinson, J. M., A. W. Lo, and T. Poggio. 1994. A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance* 49: 851–89.
- Kaucic, M. 2010. Investment using evolutionary learning methods and technical rules. *European Journal of Operational Research* 207: 1717–27.
- Krauss, C., X. A. Do, and N. Huck. 2017. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research* 259: 689–702.
- Kristjanpoller, W., and M. C. Minutolo. 2018. A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis. *Expert Systems with Applications* 109: 1–11.
- Lo, A.W. 2004. The adaptive markets hypothesis. *The Journal of Portfolio Management* 30: 15–29.
- Lo, A.W., H. Mamaysky, and J. Wang. 2000. Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation. *The Journal of Finance* 55: 1705–65.
- Menkhoff, L. 1997. Examining the use of technical currency analysis. *International Journal of Finance & Economics* 2: 307–18.
- Menkhoff, L. 2010. The use of technical analysis by fund managers: International evidence. *Journal of Banking & Finance* 34: 2573–86.
- Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou. 2014. Forecasting the equity risk premium: The role of technical indicators. *Management Science* 60: 1772–91.
- Neely, C., P. Weller, and J. Ulrich. 2009. The Adaptive Markets Hypothesis: Evidence from the Foreign Exchange Market. *Journal of Financial and Quantitative Analysis* 44: 467–88.
- Taylor, M. P., and H. Allen. 1992. The use of technical analysis in the foreign exchange market. *Journal of International Money and Finance* 11: 304–14.

## References

- Aria, Massimo, and Corrado Cuccurullo. 2017. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics* 11: 959–75. [\[CrossRef\]](#)
- Aria, Massimo, V. Della Corte, and A. Piscitelli. 2017. Business Orientation and Governance Choices in Cultural Firms: A Survey Research in Area of Naples. *Italian Journal of Applied Statistics* 29. [\[CrossRef\]](#)
- Atsalakis, George S., Ioanna G. Atsalaki, Fotios Pasiouras, and Constantin Zopounidis. 2019. Bitcoin price forecasting with neuro-fuzzy techniques. *European Journal of Operational Research* 276: 770–80. [\[CrossRef\]](#)
- Bekiros, Stelios D., and Dimitris Georgoutsos. 2008. Direction-of-change forecasting using a volatility-based recurrent neural network. *Journal of Forecasting* 27: 407–17. [\[CrossRef\]](#)
- Berardi, Michele. 2011. Fundamentalists vs. chartists: Learning and predictor choice dynamics. *Journal of Economic Dynamic Control* 35: 776–92. [\[CrossRef\]](#)
- Blake, Andrew P., and George Kapetanios. 2000. A radial basis function artificial neural network test for ARCH. *Economics Letters* 69: 15–23. [\[CrossRef\]](#)
- Chung-Ming, Kuan, and Halbert White. 1994. Artificial neural networks: An econometric perspective. *Econometric Reviews* 13: 1–91.
- Day, Richard H., and Weihong Huang. 1990. Bulls, bears and market sheep. *Journal of Economic Behavior & Organization* 14: 299–329.

- De Spiegeleer, Jan, Dilip B. Madan, Sofie Reyners, and Wim Schoutens. 2018. Machine learning for quantitative finance: Fast derivative pricing, hedging and fitting. *Quantitative Finance* 18: 1635–43. [CrossRef]
- del Amo, Iñigo Fernández, John Ahmet Erkoyuncu, Rajkumar Roy, Riccardo Palmarini, and Demetrius Onoufriou. 2018. A systematic review of Augmented Reality content-related techniques for knowledge transfer in maintenance applications. *Computers in Industry* 103: 47–71. [CrossRef]
- Dixon, Matthew F., Igor Halperin, and Paul Bilokon. 2020. *Machine Learning in Finance: From Theory to Practice*. Cham: Springer.
- Donaldson, Glen R., and Mark Kamstra. 1997. An artificial neural network-GARCH model for international stock return volatility. *Journal of Empirical Finance* 4: 17–46. [CrossRef]
- Fernandez-Rodriguez, F., C. Gonzalez-Martel, and S. Sosvilla-Rivero. 2000. On the profitability of technical trading rules based on artificial neural networks: Evidence from the Madrid stock market. *Economics Letters* 69: 89–94. [CrossRef]
- Fischer, Thomas, and Christopher Krauss. 2018. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research* 270: 654–69. [CrossRef]
- Frankel, Jeffrey A., and Kenneth A. Froot. 1990. Chartists, Fundamentalists, and Trading in the Foreign Exchange Market. *The American Economic Review* 80: 181–85.
- Garcia, Rene, and Ramazan Gencay. 2000. Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics* 94: 93–115. [CrossRef]
- Gavrishchaka, Valeriy, and Supriya Banerjee. 2006. Support Vector Machine as an Efficient Framework for Stock Market Volatility Forecasting. *Computational Management Science* 3: 147–60. [CrossRef]
- Gencay, Ramazan, and Thanasis Stengos. 1998. Moving average rules, volume, and the predictability of security returns with feedforward networks. *Journal of Forecasting* 17: 401–14. [CrossRef]
- Gerritsen, Dirk F., Elie Bouri, Ehsan Ramezanifar, and David Roubaud. 2020. The profitability of technical trading rules in the Bitcoin market. *Finance Research Letters* 34: 101263. [CrossRef]
- Glasserman, Paul, Kriste Krstovski, Paul Laliberte, and Harry Mamaysky. 2020. Choosing News Topics to Explain Stock Market Returns. In Proceedings of the ACM International Conference on A.I. in Finance (ICAIF' 20), New York, NY, USA, October 15–16; New York: ACM. [CrossRef]
- Gradojevic, Nikola, and Ramazan Gencay. 2013. Fuzzy logic, trading uncertainty and technical trading. *Journal of Banking and Finance* 37: 578–86. [CrossRef]
- Grant, Maria J., and Andrew Booth. 2009. A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal* 26: 91–108. [CrossRef]
- Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies* 33: 2223–73. [CrossRef]
- Hans, Franses P., and Kasper van Griensven. 1998. Forecasting Exchange Rates Using Neural Networks for Technical Trading Rules. *Studies in Nonlinear Dynamics & Econometrics* 2: 1–8.
- Hsieh, David A. 1989. Testing for nonlinear dependence in daily foreign exchange rates. *The Journal of Business* 62: 339–68. [CrossRef]
- Huang, Jian, Junyi Chai, and Stella Cho. 2020. Deep learning in finance and banking: A literature review and classification. *Frontiers of Business Research in China* 14: 1–24. [CrossRef]
- Hutchinson, James M., Andrew W. Lo, and Tomaso Poggio. 1994. A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance* 49: 851–89. [CrossRef]
- Jiang, Xinxin, Shirui Pan, Jing Jiang, and Guodong Long. 2018. Cross-domain deep learning approach for multiple financial market predictions. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, July 8–13; pp. 1–8.
- Kaucic, Massimiliano. 2010. Investment using evolutionary learning methods and technical rules. *European Journal of Operational Research* 207: 1717–27. [CrossRef]
- Kraus, Mathias, and Stefan Feuerriegel. 2017. Decision Support from Financial Disclosures with Deep Neural Networks and Transfer Learning. Available online: <https://arxiv.org/pdf/1710.03954.pdf> (accessed on 18 March 2021).
- Krauss, Christopher, Xuan Anh Do, and Nicolas Huck. 2017. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research* 259: 689–702.
- Kristjanpoller, Werner, and Marcel C. Minutolo. 2018. A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis, and principal components analysis. *Expert Systems with Applications* 109: 1–11. [CrossRef]
- Lee, Cheng-Fee, and John Lee, eds. 2010. *Handbook of Quantitative Finance and Risk Management*. New York: Springer.
- Lo, Andrew W. 2004. The adaptive markets hypothesis. *The Journal of Portfolio Management* 30: 15–29. [CrossRef]
- Lo, Andrew W., Harry Mamaysky, and Jiang Wang. 2000. Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation. *The Journal of Finance* 55: 1705–65. [CrossRef]
- Loughran, Tim, and Bill McDonald. 2011. When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance* 66: 35–65. [CrossRef]
- Malinauskaitė, Laura, David Cook, Brynhildur Davíðsdóttir, Helga Ögmundardóttir, and Joe Roman. 2019. Ecosystem services in the Arctic: A thematic review. *Ecosystem Services* 36: 100898. [CrossRef]
- Markowitz, Harry M. 1952. Portfolio Selection. *The Journal of Finance* 7: 77–91.

- Matsubara, Takashi, Ryo Akita, and Kuniaki Uehara. 2018. Stock price prediction by deep neural generative model of news articles. *IEICE Transactions on Information and Systems* 4: 901–8. [\[CrossRef\]](#)
- Mengist, Wondimagegn, Teshome Soromessa, and Gudina Legese. 2020. Method for conducting systematic literature review and meta-analysis for environmental science research. *MethodsX* 7: 11. [\[CrossRef\]](#) [\[PubMed\]](#)
- Menkhoff, Lukas. 1997. Examining the use of technical currency analysis. *International Journal of Finance & Economics* 2: 307–18.
- Menkhoff, Lukas. 2010. The use of technical analysis by fund managers: International evidence. *Journal of Banking & Finance* 34: 2573–86.
- Minh, Dang, Abolghasem Sadeghi-Niaraki, Huy Huynh, Kyungbok Min, and Hyeonjoon Moon. 2017. Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network. *IEEE Access* 6: 55392–404. [\[CrossRef\]](#)
- Mitra, Gautam, and Xiang Yu, eds. 2016. *The Handbook of Sentiment Analysis in Finance*. New York: Albury Books.
- Mitra, Leela, and Gautam Mitra. 2011. *The Handbook of News Analytics in Finance*. Hoboken: John Wiley & Sons.
- Neely, Christopher, David E. Rapach, Jun Tu, and Guofu Zhou. 2014. Forecasting the equity risk premium: The role of technical indicators. *Management Science* 60: 1772–1791. [\[CrossRef\]](#)
- Neely, Christopher, Paul Weller, and Joshua Ulrich. 2009. The Adaptive Markets Hypothesis: Evidence from the Foreign Exchange Market. *Journal of Financial and Quantitative Analysis* 44: 467–88. [\[CrossRef\]](#)
- Perevochtchikova, Maria, José Álvaro Hernández Flores, Wilmer Marín, Alfonso Langle Flores, Arturo Ramos Bueno, and Iskra Alejandra Rojo Negrete. 2019. Systematic review of integrated studies on functional and thematic ecosystem services in Latin America, 1992–2017. *Ecosystem Services* 36: 100900. [\[CrossRef\]](#)
- Popper, Karl Raimund. 1962. *Conjectures and Refutations; The Growth of Scientific Knowledge*. New York: Basic Books.
- Prasch, Robert, and Thierry Warin. 2016. Systemic Risk and Financial Regulations: A Theoretical Perspective. *Journal of Banking Regulation* 17: 188–99. [\[CrossRef\]](#)
- Schumaker, Robert P., and Hsinchun Chen. 2010. A discrete stock price prediction engine based on financial news. *Computer* 43: 51–56. [\[CrossRef\]](#)
- Sharpe, William F. 1963. A Simplified Model for Portfolio Analysis. *Management Science* 9: 277–93. [\[CrossRef\]](#)
- Sharpe, William F. 1964. Capital Asset Prices—A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance* 19: 425–42.
- Taylor, Mark, and Hellen Allen. 1992. The use of technical analysis in the foreign exchange market. *Journal of International Money and Finance* 11: 304–14. [\[CrossRef\]](#)
- Warin, Thierry. 2005. Popper's Falsifiability and Mises' a-Priorism: Is Dogmatism Everywhere? *Epistemologia* 28: 121–38.