



A bibliometric data analysis

DISCOVERING THE LITERATURE OF FINTECH

Anette Schultz

Maarten Moes

Oliver Havasi

Shaun McFarlane



FINTECH

Natural Language Processing and Network Analysis

This report investigates the most central topics and authors within the FinTech field. It is based on 3400 titles that have been written between 2008 and 2022 regarding FinTech. There has been an increasing number of articles each year within this field, in 2008, there were 4 published FinTech articles, after 2020 there has been over 2000 articles published each year.

The most important source titles in the field of FinTech are:

Source Title	Titles
ACM International Conference Proceeding Series	93
Lecture Notes in Computer Science	53
Sustainability (Switzerland)	49
Lecture Notes in Networks and Systems	48
IEEE Access	36

Natural Language Processing

With regards to the extreme filters for creating the corpus of the tokens, we used an extreme filter that limits the words that appear, to have at least appeared half a time and max 50 times in, while the size of the dictionary had a maximum size of 1000 words.

Topic modelling of the abstract

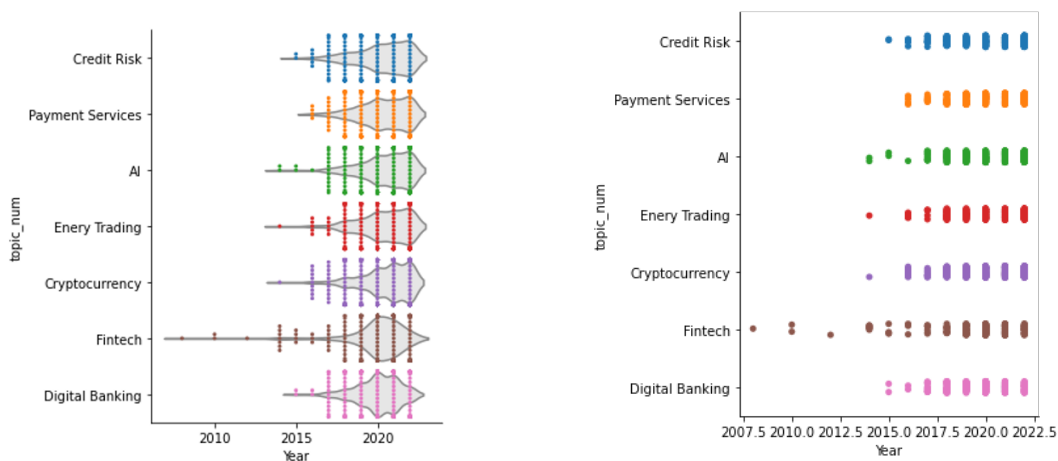
In this process we had to identify clusters within the FinTech research area to assign them labels. This resulted in seven different topics: "Innovation Management" (19,3% tokens), "Legal" (17,2% tokens), "Digital Banking" (16,9% tokens), "Empirical Finance" (13,5% tokens), "Quantitative Finance" (12,5% tokens), "Energy Trading" (11,2% tokens) and "Bank Service" (9,5% tokens).

Topic modelling of the author keywords

The same was done for Author Keywords, which also resulted in seven different labels:

“Cryptocurrency” (19,3% tokens), “AI”, “Fintech” (17,2% tokens), “Energy Trading” (16,9% tokens), “Credit Risk” (13,5% tokens), “Payment Services” (12,5% tokens) and “Digital Banking” (11,2% tokens). When visualizing the keyword topics over time, we discover that all of them have started to increase in popularity from around 2014. From the data, titles that fall under the Digital Banking category have decreased while the other categories continue to increase.

The below graphs and table illustrate the distribution of the chosen categories over time. The categories are manually assigned based on the output of the Author Keywords LDA.



Year	2008	2010	2012	2014	2015	2016	2017	2018	2019	2020	2021	2022	All
topic_num													
AI	0	0	0	2	2	1	13	28	42	92	101	121	402
Credit Risk	0	0	0	0	2	5	15	34	38	79	85	106	364
Cryptocurrency	0	0	0	1	0	9	11	43	50	124	165	177	580
Digital Banking	0	0	0	0	2	2	15	29	62	127	110	44	391
Energy Trading	0	0	0	1	0	5	5	30	35	78	77	101	332
Fintech	1	2	1	7	4	3	7	15	38	145	115	42	380
Payment Services	0	0	0	0	0	4	9	25	47	80	72	95	332
All	1	2	1	11	10	29	75	204	312	725	725	686	2781

Besides the topic of Fintech, all the topics were first introduced in 2014 and started becoming more and more popular over the following years. In addition, since 2019 all the topics within the scope of the analysis started to increase gradually.

Machine Learning

The Logistic Regression Model is used to train a machine learning model that predicts the topic category based on a given input. The possible prediction results are:

- Credit Risk
- Payment Services
- AI
- Energy Trading
- Cryptocurrency
- Fintech
- Digital Banking

This could be useful for a user making a search regarding a certain topic. Based on the user input, the model predicts a research area where the searched topic is likely to be found. This allows the user to dive into a certain topic even more, while saving valuable time in their research.

Network Analysis

Top 10 Authors within Fintech

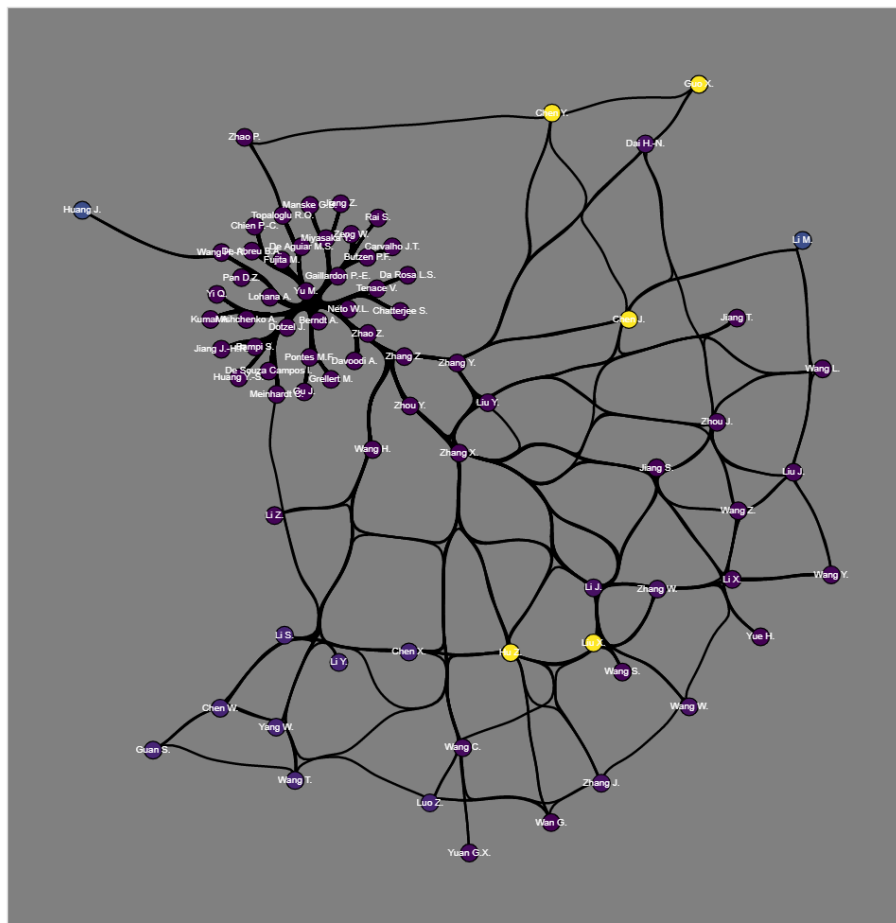
These rankings are based off the 'Eigenvector Centrality', this is a measure of how much influence the author has on the network. The purpose of the eigenvector centrality is to focus more on the quality of the author's connections rather than the quantity. The author's centrality is weighed by the centrality of other authors adjacent to them, which means an author with a high score is connected to other authors who also have a high score.

Author	University
Zhang X.	Beijing Normal University, China
Zhang Y.	Peking University, China
Zhou Y.	China University of Mining and Technology, China
Wang H.	Peking University, China
Zhang Z.	The University of Southampton, United Kingdom

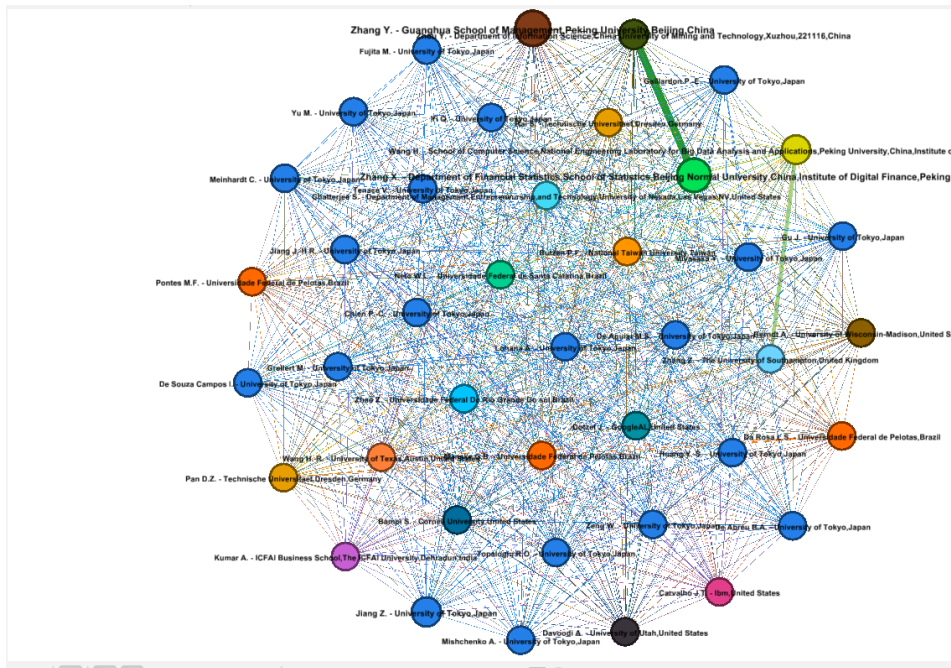
Zhao Z.	Universidade Federal Do Rio Grande Do Sul, Brazil
Chatterjee S.	University of Nevada, United States
Kumar A.	The ICFAI University, Dehradun, India
Jiang Z.	University of Tokyo, Japan
Rai S.	Technische Universitaet, Dresden, Germany

Communities within the network

In order to further analyze the network, we have grouped the authors into ‘communities’, based on the fact that authors in the same community should have a higher chance to be connected to each other. In doing this we have identified 480 different communities within FinTech. Which then allows us to identify the top authors within the top communities.



We begin to see a prominent community when we filter only on the top authors, we see that the top authors are dominated by one community. To get a better insight on the community, we filter out everything else.



This community is dominated by the University of Tokyo, which makes up 48.78% of authors, other institutions within this community are based in China, Brazil, Germany and the United States. Although most researchers are from Tokyo, the most connected people are based in China, meaning they have a lot of collaboration with Tokyo.