Research Paper

EPRA International Journal of Economic and Business Review-Peer Reviewed Journal
Volume - 13, Issue - 5, May 2025 | e-ISSN: 2347 - 9671| p- ISSN: 2349 – 0187

SJIF Impact Factor (2025): 8.862 || ISI Value: 1.433 || Journal DOI URL: https://doi.org/10.36713/epra2012

BIBLIOMETRIC ANALYSIS OF RISK ASSESSMENT IN PEER-TO-PEER LENDING

Nguyen Binh Phuong Thuy
Faculty of Finance and Accounting, Ho Chi Minh City University of Industry and Trade, Viet Nam

ABSTRACT **DOI No: 10.36713/epra21644** **Article DOI: https://doi.org/10.36713/epra21644**

P2P lending has emerged as a prominent and rapidly evolving sector within financial technology. This model allows borrowers to obtain funds quickly and easily at a rate corresponding to their risk level. Derived from these conveniences are numerous other risks, such as borrower's credit risk, platform risk, and information security. Hence, the problem of risk evaluation in the peer-to-peer lending business model is also an area of concern. This research seeks to conduct a bibliometric analysis to assess existing P2P studies and identify key trends, research groups, and impactful publications focusing on risks associated with the P2P lending model. Employing Scopus data (2010-2025), the study conducted co-citation analysis, co-occurring keywords, and bibliographic coupling. This work aims to provide enhanced coverage of the state of P2P lending research, describe the existing gaps, and propose alternative paths for future investigations. These outcomes are valuable in an information society for researchers and practitioners in the context of the increasing usage of P2P lending and FinTech globally. These findings are crucial in evaluating and developing strategies to mitigate risks, improve efficiency, and increase user-friendliness within the model.

KEYWORDS: ***Risk Assessment; Peer-To-Peer Lending; Bibliometric Analysis.***

INTRODUCTION

Fintech is a strong financial innovation marked by its rapid evolution and spurred by technological underpinnings such as the Internet, social media, artificial intelligence, and big data. These developments can reshape the financial industry through lower costs, improve the quality of financial services, and support an inclusive and stronger financial ecosystem (Mackenzie, 2015; Mehdiabadi et al., 2020). Lee and Shin (2018) indicate that fintech is developing across six main business models: payments, asset management, crowdfunding, lending, capital markets, and insurance services. The lending business model, typically peer-to-peer (P2P) lending, is a rapidly growing fintech sector compared to other sectors in many countries, such as the US, Korea, Indonesia, and China (Suryono et al., 2019). P2P lending platforms first appeared in 2005 in the UK with the launch of Zopa (Yang & Lee, 2016). In the 2010s, P2P lending platforms multiplied in the United States with the launch of platforms such as LendingClub and Prosper. This trend has also spread to Asian countries such as China, India, and Indonesia (Kusumawati et al., 2024). P2P lending is an online financial service that facilitates direct communication between borrowers and lenders through a dedicated platform (Wang et al., 2015). Owing to the streamlined framework, P2P lending can offer loans at lower interest rates while offering a quick and hassle-free borrowing experience. Unlike traditional banks, P2P lending platforms are technically not involved in lending activities; they only connect lenders to borrowers and collect user fees (Lee & Shin, 2018). These platforms facilitate the lending process by providing a digital space where borrowers can post their loan requests and lenders can choose to fund these requests based on the provided information (Kaur et al., 2024; Siering, 2023). P2P lending platforms allow borrowers with no or poor credit history to access capital more easily, thus solving the black credit problem. Along with the benefits this business model brings, it also has many potential risks, such as platform risks, legal risks, and the risk of losing capital from users.

The industry saw rapid expansion, especially from 2007 to 2016, with platforms facilitating direct transactions between lenders and borrowers without traditional banking institutions (Xu, 2022). Regarding platform risks, we can mention scandals from the Ponzi model that shocked China in 2015. Specifically, P2P Lending activities in China began in 2007, with rapid growth in the number and scale of businesses. By the end of 2015, 3,477 P2P lending platforms were operating in this country (Stern et al., 2017). However, P2P lending platforms were later discovered to operate according to the Ponzi model to defraud investors. A typical example is the peer-to-peer lending company Ezubao, which operated as a Ponzi scheme to

defraud more than 900,000 investors in 2015. During this period, Ezubao attracted approximately 50 billion yuan, equivalent to 7.6 billion US dollars (Albrecht et al., 2017). The Ezubao scandal had a contagion effect, negatively impacting other platforms and their users due to the spread of negative information (Cheng et al., 2022). Additionally, one of the significant problems in the online lending sector is information asymmetry between lenders and borrowers, which brings about moral hazard and adverse selection in peer-to-peer lending (Emekter et al., 2015). By 2019, the default rate for Chinese P2P lending platforms was 87.2%, indicating severe systemic issues (Gao et al., 2021). Factors contributing to high risks included adverse selection due to information asymmetry, lack of financial expertise among investors, and insufficient regulatory oversight. Compared to traditional financial institutions, such a system does not accommodate direct lender and borrower interaction, thereby leading to barriers in risk assessment faced by the lenders. It is, therefore, important to emphasize risk assessment in peer-to-peer lending, thereby excluding the handicap the investors face while making transactions with such platforms.

Research on peer-to-peer lending has become increasingly popular since the advent of Zopa, LendingClub, and the boom of this platform in China in the 2010s. These studies mainly focus on the peer-to-peer lending business model, default risk, and small and medium-sized companies' contributions to this platform (Kholidah et al., 2022). However, a comprehensive assessment of the risks in peer-to-peer lending is necessary, especially after China's major scandal. Applying a bibliographic approach makes it possible to map the main research topics on risk assessment in peer-to-peer lending. This helps to identify topics that have been thoroughly studied and those that have not been explored in depth, helping to determine future research directions. The authors collected data from Scopus in the period from 2010 to 2025, through a combination of 3 analytical techniques such as bibliographic coupling, co-citation, and keyword co-occurrence by using VOSviewer to meet the following research objectives: (1) Identify key research clusters, research trends related to the topic; (2) Explore the relationships between different research topics in this field; (3) Shape the research on risk assessment in peer-to-peer lending.

METHODOLOGY

Bibliometric Analysis

Bibliometric analysis is a quantitative approach used to study and analyze scientific literature such as books, articles, journals, and documents. Bibliometric analysis uses statistical methods to quantify aspects of scholarly production, such as authorship, publication patterns, and citation rates (Marvi & Foroudi, 2023; Saputro et al., 2023). These publications can be archived on data platforms such as Scopus, Web of Science, and Google Scholar (Arsenova, 2013). Bibliometric analysis includes various techniques that vary in nature and function, such as (1) bibliographic coupling analysis to identify relevant documents and map the structure of scientific fields (Alhuay-Quispe et al., 2022); (2) co-word analysis to map the thematic structure of a field and identify emerging research areas (Aleixandre-Benavent et al., 2017; Singh & Zhang, 2023); (3) citation analysis to assess the impact of individual researchers, institutions and journals (Grassia et al., 2025; Iqbal et al., 2021); and (4) co-citation evaluation to map the intellectual structure of a field and identify key research areas and influential works (Baheti, 2024). In addition, this method provides additional statistics, including authors, affiliations, and keywords. Authors often use bibliometric analysis because it allows them to analyze the impact of selected research on their field and better understand the trends in their field. This helps researchers determine which articles related to their topic have the most tremendous impact and which authors and articles are most cited (Nobanee et al., 2022).

In this study, the authors use the 4-step bibliometric analysis process proposed by Donthu et al. (2021). Specifically, the process is summarized as follows:

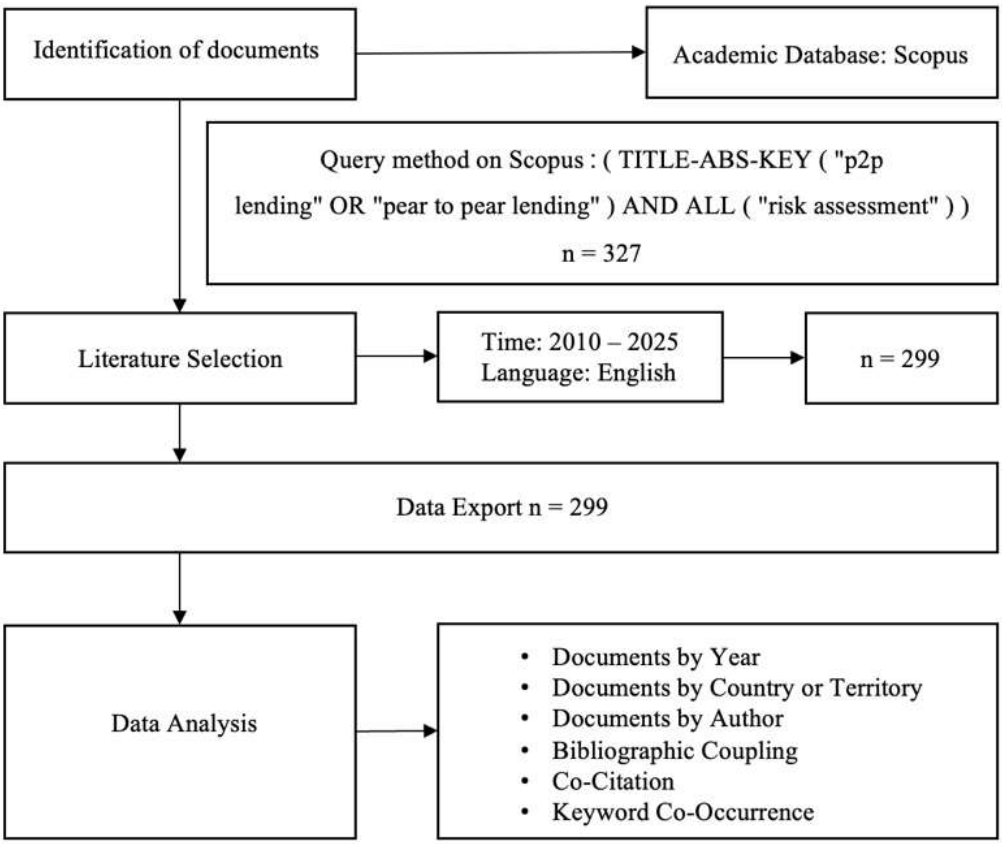
Step 1. Define the aims and scope of bibliometric study
This step involves defining the aims and scope of bibliometric studies. Accordingly, the aims of bibliometrics should relate to the retrospection of the performance and science of a research field. Meanwhile, the definition should be broad enough to warrant the use of bibliometric analysis
Step 2: Choose techniques for bibliometric analysis
In this step, it is necessary to determine the bibliometric analysis techniques appropriate for the objectives and scope of the research in Step 1.
Step 3: Collect the data for bibliometric analysis
This step requires determining search terms that yield sufficiently large, yet focused search results aligned with the research field and scope defined in Step 1. In addition, data collection in bibliometric analysis depends on the technique chosen in Step 2. Assuming that the technique chosen in Step 2 is co-word analysis, it is necessary to focus on collecting data on the title, abstract, and keywords. Furthermore, the data must be cleaned before Step 4, removing errors such as duplicates and incorrect entries.
Step 4: Run the bibliometric analysis and report the findings
After running the bibliometric analysis, it is necessary to draft a bibliographic summary and write a discussion of the findings along with their implications

Data Collection

This paper extracts the database from Scopus in the period from 2010 to 2025 because the Scopus database has proven to be more suitable for bibliometric analysis; it focuses on research articles from interdisciplinary fields, has good coverage, and fast data updates (Zhang et al., 2019). As mentioned earlier, since 2010, P2P lending has grown strongly in many countries worldwide, so the group of authors decided to collect data from 2010 to 2025.

The authors extracted a total of 299 (samples/documents) in the scope of "Article Title, Abstract, Keywords" with the search keyword "P2P lending" or "pear to pear lending" AND in the scope of "All fields" with the search keyword "risk assessment." The research time range includes articles published in the most recent decade, from 2010 to 2025. This helps the paper to be objective with updated data and an overview of P2P lending and helps to orient the research effectively. The steps taken to collect data for the final analysis are shown in Fig 1

Figure 1. The actual process of conducting bibliographic analysis.



Data Analysis Procedure

In this paper, the authors performed a bibliometric analysis using the software VOSViewer. VOSViewer is a free software tool designed for constructing and visualizing bibliometric networks. It can create maps based on various types of data, such as co-citation, co-authorship, and keyword co-occurrence, which are then represented in charts to help identify patterns and trends in research (Cárdenas-Arias et al., 2023; Li & Wei, 2022). VOSViewer excels at visualizing complex bibliometric data, helping researchers easily identify gaps, trends, and relationships in the literature (Indriati et al., 2024).

FINDINGS

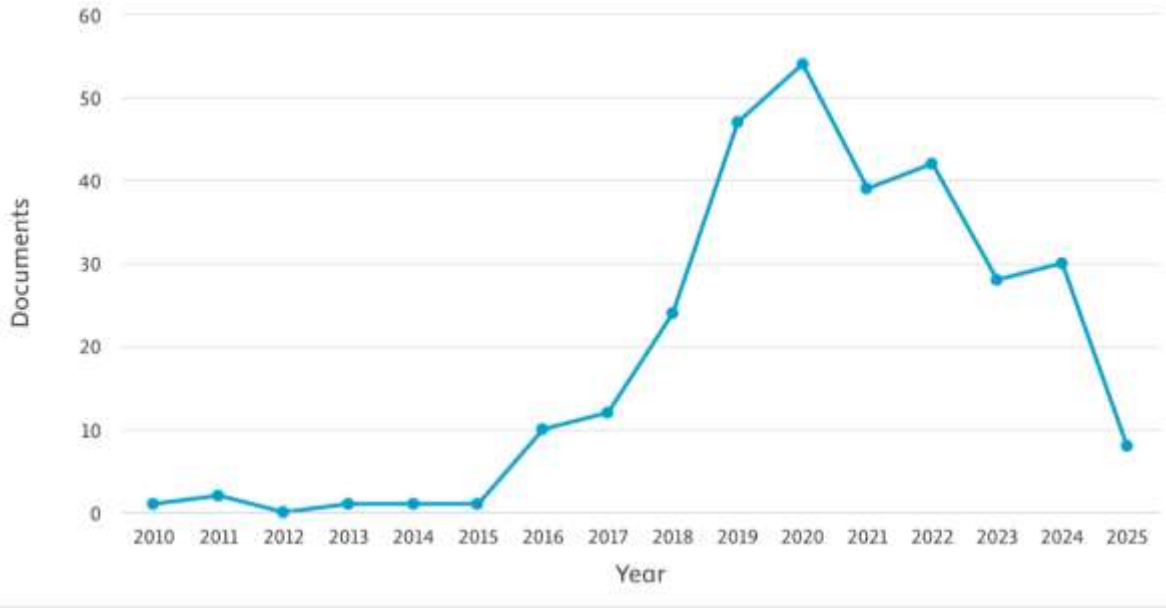
Analysis of Documents by Year

Although the first P2P lending platform was born in 2005 and exploded in the 2010s, as mentioned earlier, the statistics by number of publications by year presented in Fig. 2 do not show many studies on risk assessment in peer-to-peer lending models. Based on the statistical results, the authors divided the development stages of the research into two main stages.

Before 2015, the number of articles on risk assessment was small, with only about 1-2 articles per year. This shows that, although the field of peer-to-peer lending is growing strongly in many countries, not much attention has been paid to risk assessment in peer-to-peer lending.

After 2015,2015 marked a turning point due to legal risks and fraud related to peer-to-peer lending platforms, especially the “scandal” in China in 2015. Currently, studies are more interested in assessing risks in peer-to-peer lending operations. Specifically, the number of studies on this issue has increased from 10 in 2016 to 54 in 2020. However, after 2020, the number of studies on this topic began to gradually decrease.

Figure 2. Number of publications published by year from Scopus database

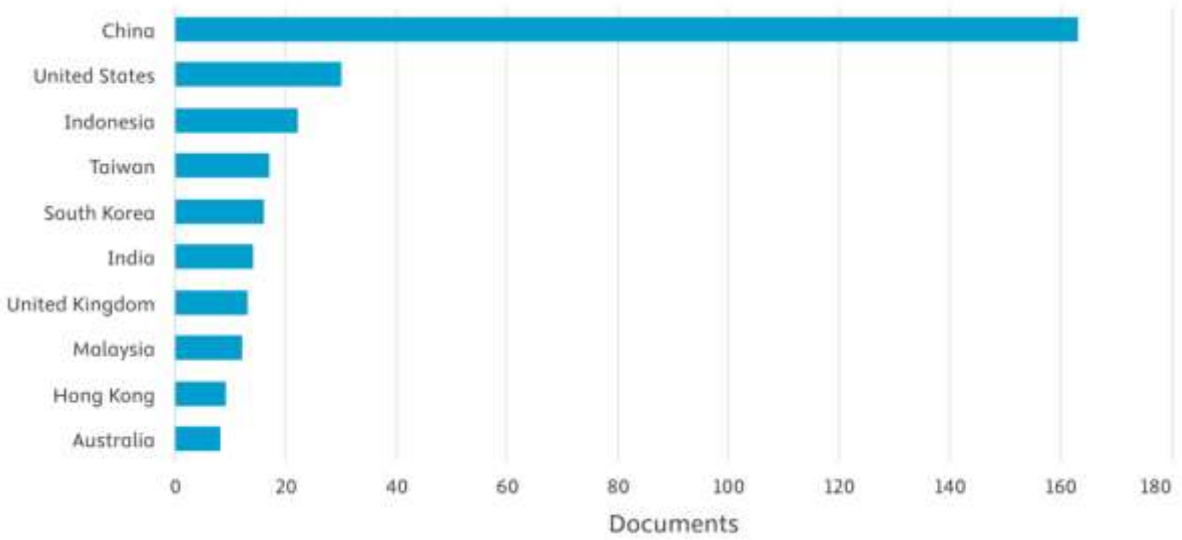


Analysis of Documents by Country or Territory

The literature statistics by region are presented in Fig. 3, which shows that China had the most significant number of publications, with 54.33% of the total sample analyzed. P2P lending has experienced a period of rapid development in China, with the world's leading peer-to-peer lending volume reaching nearly 982.3 RMB in 2015 (Wang et al., 2016).

P2P lending has become an attractive investment channel, lending in this form bringing borrowers with an average return of 21.3% in 2013, or 19.1% in 2012 (Ding et al., 2021). However, hot development is a serious problem caused by P2P lending, including illegal activities, fraud, and default. The Chinese government has shown great interest in managing and regulating this field with tight nationwide regulations. From 5,000 P2P lending platforms in early 2028, this number has decreased to 29 operating platforms in 2020 (Jiang et al., 2021). Faced with a new business model, large-scale operations, high probability of failure, and many potential risks, this country must focus more on research on risk assessment in peer-to-peer lending. In addition, research on risk assessment in peer-to-peer lending is also relatively conducted in the US, with 10%, and in Indonesia, with 7.3% of the total analysis articles.

Figure 3. Documents by country or territory by source from Scopus database.



Analysis of Documents by Author

The results of the Analysis of Documents by the Author show leading experts and influential authors on this research topic, with much research works and important contributions, thereby helping to find opportunities for future research collaboration. The results are shown in Fig. 4, with the authors featured on the topic of risk assessment in peer-to-peer lending with at least four or more publications. At the same time, the analysis results also showed close links between small groups of researchers, thereby helping to explore the collaboration network between authors.

Figure 4. Number of publications by author.

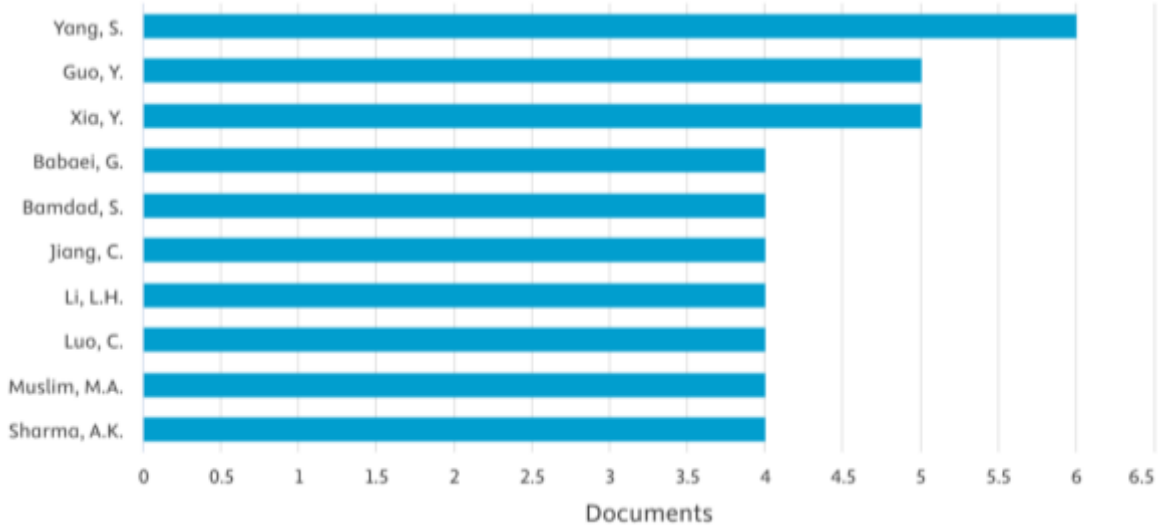


Table 1 focuses on three authors with the most publications on this topic. The authors' research focus is on applying machine learning, deep learning models, and algorithms to analyze, measure, and forecast risks in peer-to-peer lending. At the same time, the authors also propose different data treatment methods to handle bias in sampling, thereby improving the results of risk measurement and forecasting.

Table 1. Core research author on Risk Assessment in Peer-to-Peer Lending

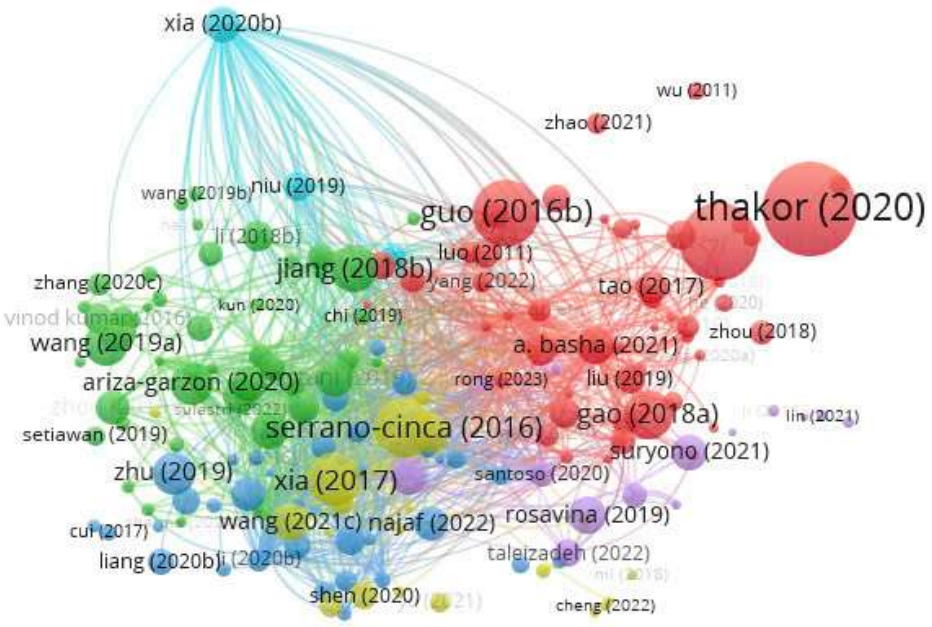
Authors		Title	General findings for the articles	No. of
Yang, S	(Wang et al., 2021)	Leveraging Multisource Heterogeneous Data for Financial Risk Prediction: A Novel Hybrid-Strategy-Based Self-Adaptive Method.	(1) Using a new learning method that allows predicting financial risks using information from various sources, from Hard features to Soft features (2) The operating ability, profitability, and security of the peer-to-peer lending platforms improve investors investment behavior (3) Proposes a new default risk prediction model based on heterogeneous ensemble learning. specifically, a combination of extreme gradient boosting (XGBoost), a deep neural network (DNN), logistic regression (LR), and a linear weight ensemble strategy. From there, the problem of missing values and hyperparameter optimization can be solved (4) The risks of China's P2P lending market can be controlled, and when this market develops sustainably, it can help improve people's quality of life. (5) The automatic bidding mechanism can undermine the herd behavior that exists in peer-to-peer lending and make herd behavior more rational. (6) Proposed a multi-round ensemble learning model based on heterogeneous ensemble frameworks to predict default risk. This model increases the accuracy of credit risk prediction compared with traditional machine learning models	6
	(Yang et al., 2020)	Identifying the influencing factors on investors' investment behavior: an empirical study focusing on the Chinese P2P lending market		
	(Li et al., 2020)	Heterogeneous ensemble learning with feature engineering for default prediction in peer-to-peer lending in China		
	(Yang et al., 2019)	Risk control of online P2P lending in China based on health investment		
	(Jiang et al., 2018)	Does automatic bidding mechanism affect herding behavior? Evidence from online P2P lending in China		
	(Li et al., 2018)	Heterogeneous ensemble for default prediction of peer-to-peer lending in China		
Guo, Y	(Guo, Jiang, Zhou, et al., 2021)	A predictive indicator using lender composition for loan evaluation in P2P lending	(1) Loan component scores based on maturity can serve as an effective indicator to determine loan quality (2) Credit risk assessment in Peer-to-peer (P2P) lending through Instance-based learning (3) An integrated loan evaluation model that exploits and fuses information from both the borrower and the investor to improve investment decisions.	5
	(Guo, Jiang, Qiao, et al., 2021)	A new integrated similarity measure for enhancing instance-based credit assessment in P2P lending		

	(Guo et al., 2019)	Borrower-lender Information Fusion for P2P Lending: A nonparametric approach	(4) Use models to assess the risk and return of individual loans, and then develop a method to maximize the portfolio. (5) Propose an investment decision framework based on data on past investment performance, taste, and experience of investors	
	(Guo et al., 2016)	Instance-based credit risk assessment for investment decisions in P2P lending		
	(Luo et al., 2011)	Enhancing investment decisions in P2P lending: An investor composition perspective		
Xia, Y	(Xia et al., 2021)	Forecasting loss given default for peer-to-peer loans via heterogeneous stacking ensemble approach	(1) To improve the loss given default (LGD) forecasting, a heterogeneous stacking ensemble (HSE) method was developed based on the analysis of two real-world datasets. (2) Proposing a new credit scoring model for peer-to-peer (P2P) lending by combining CatBoost algorithm and keyword clustering. (3) Combining an outlier detection technique and a state-of-the-art gradient boosting decision tree algorithm to improve the prediction performance and build a credit scoring model. (4) To solve the problem of sampling bias in credit scoring models, the authors proposed the CPLE-LightGBM method. (5) Proposing an efficient portfolio allocation model in peer-to-peer lending market and a loan evaluation model based on boosted tree	5
	(Xia et al., 2020)	Predicting loan default in peer-to-peer lending using narrative data		
	(Xia, 2019)	A Novel Reject Inference Model Using Outlier Detection and Gradient Boosting Technique in Peer-to-Peer Lending		
	(Xia et al., 2018)	A rejection inference technique based on contrastive pessimistic likelihood estimation for P2P lending		
	(Xia et al., 2017)	Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending		

Bibliographic Coupling

Bibliometric coupling is a technique for finding conceptually similar articles when citing a document. This method allows for measuring the number of references shared by articles. When two articles share a reference, they are considered bibliometrically coupled. The bibliographic coupling of two articles is stronger when they have more common references (Haghani et al., 2021). This result allows for the identification of groups of related research on the topic

Figure 5. Bibliographic coupling analysis results from VOSviewer.



The authors conducted a bibliographic coupling analysis using VOSviewer, and the results are shown in Fig. 5, six research clusters were identified.

Cluster 1: Factors Influencing Credit Risk in P2P Lending

This cluster focuses on the various factors influencing credit risk in the peer-to-peer lending business model. Studies in the

cluster indicate that these factors include borrower characteristics (credit score, income, and education level), loan characteristics (total loan amount, interest rate, and term), non-financial information (investor sentiment and social media information), and macroeconomic factors. Important studies include Guo et al. (2016), Emekter et al. (2015), and Wang et al. (2020).

Cluster 2: Using Machine Learning for Credit Risk Assessment in P2P Lending
This cluster emphasizes using machine learning algorithms and techniques to assess risk in peer-to-peer lending. Specifically, the studies here have included traditional machine learning models (logistic regression, decision tree, random forest), deep learning models (deep neural networks, convolutional neural networks, Long Short-Term Memory), ensemble models, and data-processing methods specific to P2P lending. Important studies include those by Ariza-Garzón et al. (2020), Cho et al. (2019), and Bastani et al. (2019).

Cluster 3: Modeling and predicting credit risk in P2P lending
This cluster uses machine-learning algorithms and statistical models to predict credit risk and borrower repayment ability based on borrowers' characteristics, loans, and macroeconomic factors. In addition to proposing models, the system supports investors when making investment decisions on loans within a P2P lending platform. Important studies include Babaei and Bamdad (2020), Lyócsa et al. (2022), and Li et al. (2020).

Cluster 4: Advanced Modeling and Applications in P2P Lending Risk Assessment
This cluster focuses on advanced models and specific risk-assessment applications for P2P lending. Specific models and applications include portfolio optimization models, imbalanced data processing methods, systemic risk prediction models, and applications of the IoT and blockchain in P2P lending. Important studies include those by Xia et al. (2017) and Wang et al. (2021).
Cluster 5: Factors influencing behavior and trust in peer-to-peer lending

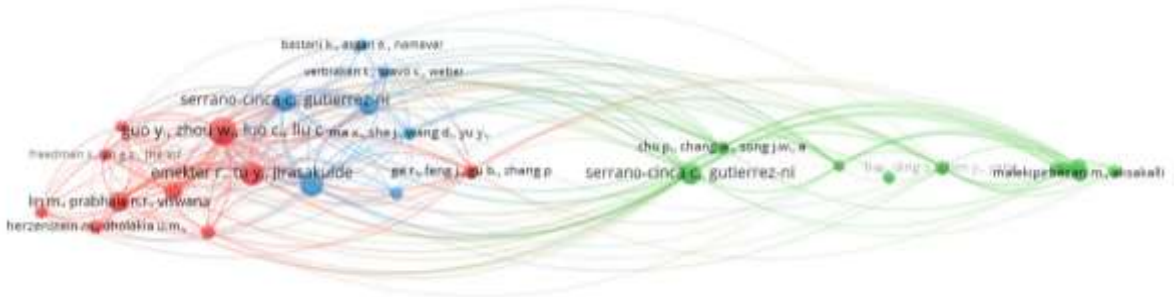
This cluster focuses on factors influencing the behavior of stakeholders (lenders, borrowers, investors) and the role of trust in peer-to-peer lending. Studies here may include psychological factors, social factors, the impact of information, and platform-related issues (perceived ease of use, security assurance, risk, trust). Important studies include Santoso et al. (2020) and Taleizadeh et al. (2022).

Cluster 6: The Impact of Information on Risk Assessment in P2P Lending
This cluster focuses on issues related to fairness in risk assessment and the impact of different types of information on P2P lending.

Co-Citation

Co-citation analysis is used in science mapping studies, assuming that cited publications tend to have similar topics (Hjørland, 2013). In a co-citation network, two publications are connected when they appear in the same reference list of another publication (Donthu et al., 2021). The primary purpose of co-citation analysis is to detect influential studies and identify major topic clusters discussed in the past (Walter & and Ribière, 2013). The authors conducted a co-citation analysis of research articles on the main topic of Risk Assessment in Peer-to-Peer Lending, intending to discover key studies and major topic clusters discussed in the past. The results of the co-citation analysis are presented in Fig. 6.

Figure 6. Co-citation analysis results on VOSviewer.



Cluster 1 is shown in red on the left side, and this cluster focuses on the influence of "social factors and information asymmetry" on risk assessment in peer-to-peer lending. Specifically, studies in this cluster study the role of social networks, relationships in lending, and credit quality assessment, such as Guo et al. (2016) and Freedman and Jin (2017). At the same time, studies also consider the issue of information asymmetry between borrowers and lenders in peer-to-peer lending activities, such as Yum et al. (2012).
Cluster 2, shown in the green section, focuses on modeling and predicting risk in peer-to-peer lending. The studies in this cluster mainly cite the study of using the random forest (RF) model to predict the status of lenders (Malekipirbazari & Aksakalli, 2015), followed by co-integration from Serrano-Cinca and Gutiérrez-Nieto (2016) on applying credit scoring systems to estimate the probability of default of loans; or the study of Xia et al. (2017) applying a model that incorporates cost-sensitive learning and extreme gradient boosting (XGBoost) to find a way to distinguish borrowers with high probability of default, or applying machine learning and deep learning models according to the study of Byanjankar et al. (2015) and Cho et al. (2019).

Cluster 3 shows the blue part in the middle, focusing on advanced techniques for evaluating the profitability of a loan in peer-to-peer lending. This cluster focuses on the research of

Bastani et al. (2019), which uses wide and deep learning to build predictive models for credit scoring and profit prediction, a decision support system proposed by Xia et al. (2017).

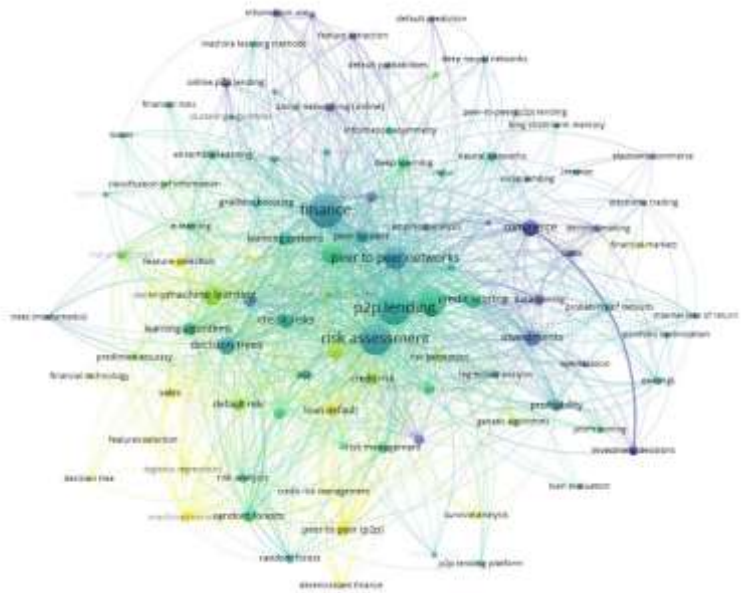
Keyword Co-Occurrence

By examining the links between keywords in documents, keyword co-occurrence allows for exploring research topics and their relationships in the author's field of interest (Radhakrishnan et al., 2017). Regarding risk assessment in peer-to-peer lending, 1821 keywords appeared in 299 articles related to the selected topic. The authors selected a minimum frequency of keyword occurrence in an article of 5 times to reinforce the research topic as having a high concentration. The results showed that 99 keywords that satisfied the criteria were included in the analysis. Table 2 summarizes the 15 keywords that appeared most frequently in articles on risk assessment in peer-to-peer lending.

Table 2. Frequency of co-occurrence of keywords.

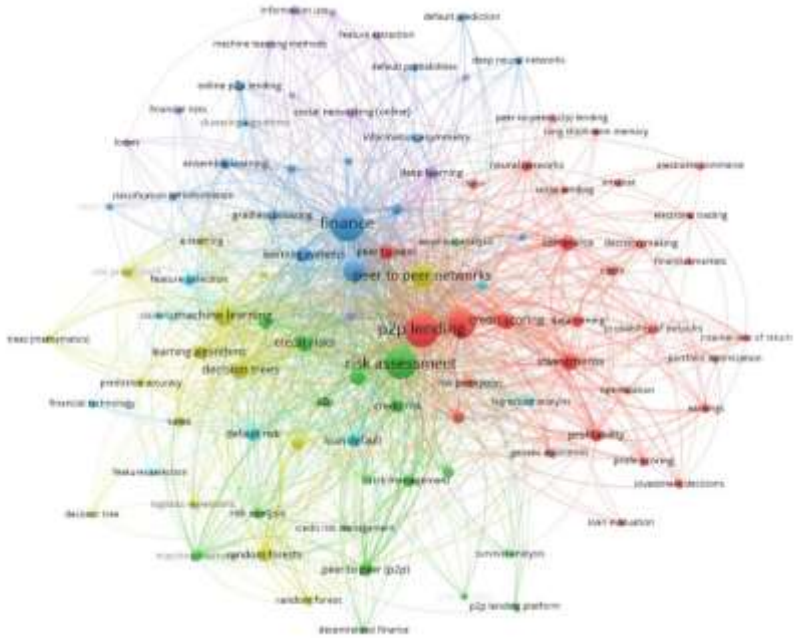
Rank	Keywords	Frequency
1	Finance	138
2	Risk Assessment	116
3	P2P Lending	134
4	Peer-To-Peer Lending	85
5	Peer To Peer Networks	63
6	Machine Learning	45
7	Forecasting	49
8	Decision Trees	37
9	Investments	31
10	Credit Scoring	31
11	Credit Risks	28
12	Learning Systems	26
13	Behavioral Research	22
14	Peer To Peer	22
15	Learning Algorithms	21

Figure 7. Keyword overlap visualization results from VOSviewer.



The keyword clusters are closely related, showing the risk assessment problem in peer-to-peer lending, as shown in Fig. 7. For example, "Machine learning models, risk prediction" (cluster 2), "Technology, analysis, and credit risk management" (cluster 3) relate to "P2P Lending and traditional financial factors" (cluster 1), showing the application of machine learning algorithms in risk assessment. The yellow cluster presented in Fig. 7 shows a new direction for this topic in recent years. The authors focus mainly on applying machine learning models such as logistic regression, decision trees, random forests… in predicting default risk for loans. While previous studies on this topic often focus on core concepts of the field, such as "peer-to-peer lending," "investment," "peer-to-peer networks," "risk assessment," or focus on more traditional aspects of risk assessment, management, and profitability assessment in peer-to-peer lending. Researching machine learning algorithms in risk assessment is necessary to ensure this business model is truly efficient for borrowers and lenders.

Figure 8. Keyword co-occurrence analysis results from VOSviewer.



The Keyword co-occurrence analysis results in Fig. 8 indicate that the keyword clusters are closely linked. Specifically, cluster 1 in red, which concerns P2P Lending and traditional financial factors, is linked to clusters showing the application of technology and machine learning algorithms in credit risk analysis and management in Cluster 2 (green), Cluster 3 (blue), and Cluster 4 (yellow).

Cluster 1 (Red): "P2P Lending and traditional financial factors". Top keywords: "P2P lending," "credit scoring," "investment," "probability," "credit risk assessment," "data mining," "decision making," "earnings," "internal rate of return," "internet decisions," "long short-term memory," "portfolio optimization," "probability of defaults," "risk perception," "social lending," "loan evaluation," "neural networks," "optimization," "peer to peer," "peer-to-peer (P2P) lending," "peer-to-peer lending." This cluster focuses on combining traditional finance elements, technology, and peer-to-peer lending models to ensure the effectiveness of this business model. Traditional finance elements, such as credit risk assessment, portfolio optimization, and effective risk management, are necessary to operate peer-to-peer lending platforms.

Cluster 2 (Green): "Technology, credit risk analysis, and management." Keywords: "artificial intelligence," "big data," "credit assessment," "credit risk," "credit risk management," "decentralized finance," "empirical analysis," "fintech," "interest rates," "machine-learning," "P2P," "P2P lending platform," "peer to peer (P2P)," "risk analysis," "risk assessment," "risk management," "survival analysis." Cluster 2 applies technologies such as AI and big data to assess, analyze, and manage credit risks in peer-to-peer lending.

Cluster 3 (Blue): "Machine learning models, risk prediction". Keywords: "adaptive boosting," "classification (of inform)," "clustering algorithms," "deep neural networks," "default prediction," "default probabilities," "empirical studies," "ensemble learning," "finance," "financial institution," "forecasting," "gradient boosting," "information asymmetry," "learning systems," "online P2P lending," "reject inference," "soft information." Cluster 3 focuses on applying machine learning and deep learning models to assess default risk in lending activities.

Cluster 4 (Yellow): "Basic machine learning algorithms, applications". Keywords: "convolutional neural ne," "decision tree," "decision trees," "e-learning," "learning algorithms," "logistic regression," "logistics regressions," "machine learning," "peer-to-peer networks," "prediction accuracy," "random forest," "random forests," "risk predictions," "sales," "trees (mathematics)." Cluster 4 focuses on fundamental machine learning algorithms, such as decision trees, logistic regression, random forests, and their applications in performance evaluation and risk prediction.

Cluster 5 (Purple): "Machine learning and information processing". Keywords: "credit scoring model", "deep learning", "feature extraction", "financial risks", "information systems", "information use", "losses", "machine learning method, "machine learning mode, "natural language process", "social networking (online)". Cluster 5 mainly concerns applying machine learning to solve financial problems, such as credit scoring models, financial risks, losses, and applications in information processing.

Cluster 6 (Turquoise): "Risk analysis and modeling." Keywords: "default risk," "feature selection," "features selection," "financial technology," "loan default," "regression analysis," "stackings," "support vector machine." Cluster 6 deals with the analysis and modeling of risks through statistical and machine-learning techniques.

DISCUSSION

In this study, the authors performed a bibliometric analysis using VOSviewer to answer the research questions posed previously. They used analytical techniques such as

bibliographic coupling, co-citation, and keyword co-occurrence.

Research question 1 (RQ1) aims to identify key research clusters and related trends.
Bibliographic coupling analysis allows the identification of groups (clusters) related to the topic of risk assessment in peer-to-peer lending and, at the same time, shows the trend of further research. The results showed the formation of six document clusters with different aspects related to the topic. For example, Cluster 1 focuses on the factors that affect credit risk in peer-to-peer lending. Meanwhile, clusters 2, 3, and 4 focused on applying machine learning and statistical models in risk assessment and prediction. The remaining clusters focus on factors affecting behavior, trust, and information asymmetry in risk assessment. In addition, a literature review shows that recent studies have focused on using machine learning models in risk prediction. Implementing bibliographic coupling has also provided much-needed overview information related to risk assessment in peer-to-peer lending, paving the way for further relevant research on this topic.

In addition, The Keyword Co-Occurrence analysis results show that there are six clusters of topics that are closely linked together, such as the connection between traditional financial factors in peer-to-peer lending (Cluster 1) and the application of technology such as artificial intelligence, machine learning models, deep learning, and big data to assess and predict risks based on traditional financial factors. Simultaneously, when considered over time, the results show that the application of new technologies is receiving considerable attention, especially in the current explosion of artificial intelligence. This study provides an overview of the current research on risk assessment in peer-to-peer lending. Create a premise for researchers, developers, and fintech startups to focus more deeply on risk assessment in this business model. Contributing to helping the peer-to-peer lending business model develop sustainably and safely for borrowers and lenders.

Research question 2 (RQ2) aimed to explore the relationships between different research topics in this field.
The results of the co-citation analysis allowed for the identification of influential studies, thematic clusters that have been discussed, and the overlapping relationship of these topics in the past. Accordingly, three thematic clusters are discussed, focusing on the influence of "social factors and information asymmetry" on risk assessment in peer-to-peer lending (cluster 1), risk modeling and prediction (cluster 2), and advanced techniques for assessing loan profitability (cluster 3). This shows the developments in risk assessment research with many topics, methods, assessment models, risk prediction, and profits in peer-to-peer lending.
Research question 3 (RQ3) aims to shape the research on risk assessment in peer-to-peer lending.

An analysis of Documents by Year shows that studies on risk assessment in peer-to-peer lending received little attention before 2015. Meanwhile, after 2015, studies on this topic began to increase dramatically. This stems from the issues of legal risks and fraud related to peer-to-peer lending platforms, especially the "scandal" in China in 2015. At the same time, the Documents by Country or Territory analysis shows that China has the most significant number of publications, with 54.33% of the total sample analyzed. In addition, research on risk assessment in peer-to-peer lending is also relatively conducted in the US, with 10%, and in Indonesia, with 7.3% of the total number of analysis articles.

CONCLUSIONS AND FURTHER RESEARCH

This research gives an overview of the state of the art, detects gaps in knowledge, and indicates avenues for future research. Given the increasing popularity of P2P lending and fintech, these results are relevant to researchers and practitioners. This research nevertheless had certain limitations. The first limitation is that the data were collected only through the Scopus database. Although Scopus is a large data repository, it cannot include all the relevant academic literature. Therefore, the study may have missed research articles related to topics not in the Scopus database and publications not in English. Future studies can collect data from various databases, such as Web of Science, Dimensions, PubMed, and Google Scholar, to overcome this limitation.

In addition, Bibliometric Analysis allows the identification of topic clusters and research trends; however, it does not provide a detailed analysis of the content or quality of each research article. Future research should combine bibliometric analysis with other qualitative methods, such as systematic reviews or meta-analyses, to gain deeper insights into the research in the sample.

REFERENCES

1. *Albrecht, C., Morales, V., Baldwin, J. K., & Scott, S. D. (2017). Ezubao: a Chinese Ponzi scheme with a twist. Journal of Financial Crime, 24(2), 256-259. https://doi.org/10.1108/JFC-04-2016-0026*
2. *Aleixandre-Benavent, R., González De Dios, J., Castelló Cogollos, L., Navarro Molina, C., Alonso-Arroyo, A., Vidal-Lnfer, A., & Lucas-Domínguez, R. (2017). Bibliometrics and indicators of scientific activity (1). the evaluation of research and scientific activity in pediatrics through bibliometrics [Article]. Acta Pediatrica Espanola, 75(1-2), 18-25. https://www.scopus.com/inward/record.uri?eid=2-s2.085018553703&partnerID=40&md5=3102d2e7d8723935a584ba0679ac20a3*
3. *Alhuay-Quispe, J., Estrada-Cuzcano, A., & Bautista-Ynofuente, L. (2022). Analysis and data visualization in bibliometric studies [Article]. JLIS.it, 13(2), 58-73. https://doi.org/10.36253/jlis.it-461*
4. *Ariza-Garzón, M. J., Arroyo, J., Caparrini, A., & Segovia-Vargas, M. J. (2020). Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending. IEEE Access,8,64873-64890. https://doi.org/10.1109/ACCESS.2020.2984412*
5. *Arsenova, I. (2013). New Application of Bibliometrics. Procedia - Social and Behavioral Sciences, 73, 678–682. https://doi.org/10.1016/j.sbspro.2013.02.105*
6. *Babaei, G., & Bamdad, S. (2020). A multi-objective instance-based decision support system for investment recommendation in peer-to-peer lending. Expert Systems withApplications,150,113278. https://doi.org/https://doi.org/10.1016/j.eswa.2020.113278*

7. *Baheti, S. S. (2024). Exploring evolution, development, and contribution of International Journal of Industrial and Systems Engineering (2005–2022): a bibliometric study [Article]. International Journal of Industrial and Systems Engineering,48(4),441-469. https://doi.org/10.1504/IJISE.2024.143711*
8. *Bastani, K., Asgari, E., & Namavari, H. (2019). Wide and deep learning for peer-to-peer lending. Expert Systems with Applications,134,209-224. https://doi.org/https://doi.org/10.1016/j.eswa.2019.05.042*
9. *Byanjankar, A., Heikkilä, M., & Mezei, J. (2015, 7-10 Dec. 2015). Predicting Credit Risk in Peer-to-Peer Lending: A Neural Network Approach. 2015 IEEE Symposium Series on Computational Intelligence,*
10. *Cárdenas-Arias, C. G., Zuza-Hernández, E., Quiroga-Méndez, J. E., Lengerke-Perez, O., & Acosta-Cárdenas, O. A. (2023). Lactic acid production a bibliometric study [Article]. Periodicals of Engineering and Natural Sciences, 11(6),27-35. https://doi.org/10.21533/pen.v11i6.3892.g1345*
11. *Cheng, W. G., Leite, R. D. O., & Caldieraro, F. (2022). Financial contagion in internet lending platforms: Who pays the price? [Article]. Finance Research Letters, 45, Article 102187. https://doi.org/10.1016/j.frl.2021.102187*
12. *Cho, P., Chang, W., & Song, J. W. (2019). Application of instance-based entropy fuzzy support vector machine in peer-to-peer lending investment decision. IEEE Access, 7, 16925-16939. https://doi.org/10.1109/ACCESS.2019.2896474*
13. *Ding, C., Kavuri, A. S., & Milne, A. (2021). Correction to: Lessons from the rise and fall of Chinese peer-to-peer lending. Journal of Banking Regulation, 22(2), 144-144. https://doi.org/10.1057/s41261-020-00142-0*
14. *Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 133, 285-296. https://doi.org/https://doi.org/10.1016/j.jbusres.2021.04.070*
15. *Emekter, R., Yanbin, T., Benjamas, J., & and Lu, M. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. Applied Economics, 47(1), 54-70. https://doi.org/10.1080/00036846.2014.962222*
16. *Freedman, S., & Jin, G. Z. (2017). The information value of online social networks: Lessons from peer-to-peer lending. International Journal of Industrial Organization, 51, 185-222. https://doi.org/https://doi.org/10.1016/j.ijindorg.2016.09.002*
17. *Gao, M., Yen, J., & Liu, M. (2021). Determinants of defaults on P2P lending platforms in China [Article]. International Review of Economics and Finance, 72, 334-348. https://doi.org/10.1016/j.iref.2020.11.012*
18. *Grassia, V., Fiori, A., Diodati, F., Sayahpour, B., Jamilian, A., Armogida, N. G., D'apuzzo, F., & Nucci, L. (2025). Clear aligners: A network and bibliometric analysis of 50 pivotal articles [Review]. Dental and Medical Problems, 62(1), 161-171. https://doi.org/10.17219/dmp/188319*
19. *Guo, Y., Jiang, S., Chen, F., Li, Y., & Luo, C. (2019). Borrower-lender Information Fusion for P2P Lending: A nonparametric approach [Article]. Ingenierie des Systemes d'Information,24(3),269-279. https://doi.org/10.18280/isi.240307*
20. *Guo, Y., Jiang, S., Qiao, H., Chen, F., & Li, Y. (2021). A new integrated similarity measure for enhancing instance-based credit assessment in P2P lending. Expert Systems with Applications, 175, 114798. https://doi.org/https://doi.org/10.1016/j.eswa.2021.114798*
21. *Guo, Y., Jiang, S., Zhou, W., Luo, C., & Xiong, H. (2021). A predictive indicator using lender composition for loan evaluation in P2P lending. Financial Innovation, 7(1), 49. https://doi.org/10.1186/s40854-021-00261-1*
22. *Guo, Y., Zhou, W., Luo, C., Liu, C., & Xiong, H. (2016). Instance-based credit risk assessment for investment decisions in P2P lending. European Journal of Operational Research, 249(2), 417-426. https://doi.org/https://doi.org/10.1016/j.ejor.2015.05.050*
23. *Haghani, M., Bliemer, M. C. J., & Hensher, D. A. (2021). The landscape of econometric discrete choice modelling research. Journal of Choice Modelling, 40, 100303. https://doi.org/https://doi.org/10.1016/j.jocm.2021.100303*
24. *Hjørland, B. (2013). Citation analysis: A social and dynamic approach to knowledge organization. Information Processing & Management, 49, 1313–1325. https://doi.org/10.1016/j.ipm.2013.07.001*
25. *Indriati, S. P., Saputro, D. R. S., & Widyaningsih, P. (2024). THE BIBLIOMETRIC NETWORK TO IDENTIFY RESEARCH TRENDS IN MULTI-INPUT TRANSFER FUNCTION [Article]. Barekeng, 18(3), 1931-1938. https://doi.org/10.30598/barekengvol18iss3pp1931-1938*
26. *Iqbal, S., Hassan, S. U., Aljohani, N. R., Alelyani, S., Nawaz, R., & Bornmann, L. (2021). A decade of in-text citation analysis based on natural language processing and machine learning techniques: an overview of empirical studies [Article]. Scientometrics, 126(8), 6551-6599. https://doi.org/10.1007/s11192-021-04055-1*
27. *Jiang, C., Xu, Q., Zhang, W., Li, M., & Yang, S. (2018). Does automatic bidding mechanism affect herding behavior? Evidence from online P2P lending in China. Journal of Behavioral and Experimental Finance, 20, 39-44. https://doi.org/https://doi.org/10.1016/j.jbef.2018.07.001*
28. *Jiang, J., Liao, L., Wang, Z., & Zhang, X. (2021). Government Affiliation and Peer-To-Peer Lending Platforms in China. Journal of Empirical Finance, 62, 87-106. https://doi.org/https://doi.org/10.1016/j.jempfin.2021.02.004*
29. *Kaur, A., Bhatnagar, D., & Singh, G. (2024). Exploring the machine learning-based decision tree technique in assessing credit risk likelihood of P2P Fintech lending for effective marketing. In Neuromarketing's Role in Sustainable Finance (pp. 397-421). https://doi.org/10.4018/979-8-3693-9117-4.ch019*
30. *Kholidah, H., Hijriah, H. Y., Mawardi, I., Huda, N., Herianingrum, S., & Alkausar, B. (2022). A Bibliometric mapping of peer-to-peer lending research based on economic and business perspective. Heliyon, 8(11). https://doi.org/10.1016/j.heliyon.2022.e11512*
31. *Kusumawati, R., Hanafi, M. M., Surwanti, A., & Rahmawati, A. (2024). Peer-to-peer lending: An analytical review of research trends and future prospects [Article]. Edelweiss Applied Science and Technology, 8(6), 5336-5348. https://doi.org/10.55214/25768484.v8i6.3190*
32. *Lee, I., & Shin, Y. J. (2018). Fintech: Ecosystem, business models, investment decisions, and challenges [Article]. Business Horizons, 61(1), 35-46.*

https://doi.org/10.1016/j.bushor.2017.09.003

33. *Li, J., & Wei, R. (2022). VOSviewer Application Status and Its Knowledge Base [Article]. Journal of Library and Information Science in Agriculture, 34(6), 61-71. https://doi.org/10.13998/j.cnki.issn1002-1248.21-0843*
34. *Li, W., Ding, S., Chen, Y., & Yang, S. (2018). Heterogeneous ensemble for default prediction of peer-to-peer lending in China [Article]. IEEE Access, 6, 54396-54406, Article 8305471. https://doi.org/10.1109/ACCESS.2018.2810864*
35. *Li, W., Ding, S., Wang, H., Chen, Y., & Yang, S. (2020). Heterogeneous ensemble learning with feature engineering for default prediction in peer-to-peer lending in China. World Wide Web, 23(1), 23-45. https://doi.org/10.1007/s11280-019-00676-y*
36. *Luo, C., Xiong, H., Zhou, W., Guo, Y., & Deng, G. (2011). Enhancing investment decisions in P2P lending: An investor composition perspective. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,*
37. *Lyócsa, Š., Vašaničová, P., Hadji Misheva, B., & Vateha, M. D. (2022). Default or profit scoring credit systems? Evidence from European and US peer-to-peer lending markets. Financial Innovation, 8(1), 32. https://doi.org/10.1186/s40854-022-00338-5*
38. *Mackenzie, A. (2015). THE FINTECH REVOLUTION. London Business School Review, 26, 50-53. https://doi.org/10.1111/2057-1615.12059*
39. *Malekipirbazari, M., & Aksakalli, V. (2015). Risk assessment in social lending via random forests. Expert Systems with Applications, 42(10), 4621-4631. https://doi.org/https://doi.org/10.1016/j.eswa.2015.02.001*
40. *Marvi, R., & Foroudi, M. M. (2023). Bibliometric analysis: Main procedure and guidelines. In Researching and Analysing Business: Research Methods in Practice (pp. 43-54). https://doi.org/10.4324/9781003107774-4*
41. *Mehdiabadi, A., Tabatabeinasab, M., Spulbar, C., Karbassi Yazdi, A., & Birau, R. (2020). Are We Ready for the Challenge of Banks 4.0? Designing a Roadmap for Banking Systems in Industry 4.0. International Journal of Financial Studies, 8(2).*
42. *Nobanee, H., Shanti, H. Z., Dilshad, M. N., Alzaabi, F., Alkindi, S., Alhammadi, J., & Alnaqbi, M. (2022). BIBLIOMETRIC ANALYSIS OF FOREIGN EXCHANGE RISK [Article]. Journal of Governance and Regulation, 11(1), 86-99. https://doi.org/10.22495/jgrv11i1art9*
43. *Radhakrishnan, S., Erbis, S., Isaacs, J. A., & Kamarthi, S. (2017). Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature [Retracted]. PLoS ONE, 12(3), Article e0172778. https://doi.org/10.1371/journal.pone.0172778*
44. *Santoso, W., Irwan, T., & and Risfandy, T. (2020). What Determine Loan Rate and Default Status in Financial Technology Online Direct Lending? Evidence from Indonesia. Emerging Markets Finance and Trade, 56(2), 351-369. https://doi.org/10.1080/1540496X.2019.1605595*
45. *Saputro, D., Prasetyo, H., Wibowo, A., Khairina, F., Sidiq, K., & Wibowo, G. (2023). BIBLIOMETRIC ANALYSIS OF NEURAL BASIS EXPANSION ANALYSIS FOR INTERPRETABLE TIME SERIES (N-BEATS) FOR RESEARCH TREND MAPPING. BAREKENG: Jurnal Ilmu Matematika dan Terapan, 17, 1103-1112. https://doi.org/10.30598/barekengvol17iss2pp1103-1112*
46. *Serrano-Cinca, C., & Gutiérrez-Nieto, B. (2016). The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. Decision Support Systems, 89, 113-122. https://doi.org/https://doi.org/10.1016/j.dss.2016.06.014*
47. *Siering, M. (2023). Peer-To-Peer (P2P) Lending Risk Management: Assessing Credit Risk on Social Lending Platforms Using Textual Factors [Article]. ACM Transactions on Management Information Systems, 14(3), Article 25. https://doi.org/10.1145/3589003*
48. *Singh, A. K., & Zhang, Y. (2023). Understanding the Evolution of Environment, Social and Governance Research: Novel Implications From Bibliometric and Network Analysis [Review]. Evaluation Review, 47(2), 350-386. https://doi.org/10.1177/0193841X221121244*
49. *Stern, C., Makinen, M., & Qian, Z. (2017). FinTechs in China – with a special focus on peer to peer lending. Journal of Chinese Economic and Foreign Trade Studies, 10(3), 215-228. https://doi.org/10.1108/JCEFTS-06-2017-0015*
50. *Suryono, R. R., Purwandari, B., & Budi, I. (2019). Peer to Peer (P2P) Lending Problems and Potential Solutions: A Systematic Literature Review. Procedia Computer Science, 161, 204-214. https://doi.org/https://doi.org/10.1016/j.procs.2019.11.116*
51. *Taleizadeh, A. A., Safaei, A. Z., Bhattacharya, A., & Amjadian, A. (2022). Online peer-to-peer lending platform and supply chain finance decisions and strategies. Annals of OperationsResearch,315(1),397-427. https://doi.org/10.1007/s10479-022-04648-w*
52. *Walter, C., & and Ribière, V. (2013). A citation and co-citation analysis of 10 years of KM theory and practices. Knowledge Management Research & Practice, 11(3), 221-229. https://doi.org/10.1057/kmrp.2013.25*
53. *Wang, G., Chen, G., Zhao, H., Zhang, F., Yang, S., & Lu, T. (2021). LEVERAGING MULTISOURCE HETEROGENEOUS DATA FOR FINANCIAL RISK PREDICTION: A NOVEL HYBRID-STRATEGY-BASED SELF-ADAPTIVE METHOD [Article]. MIS Quarterly: Management Information Systems, 45(4), 1949-1998. https://doi.org/10.25300/MISQ/2021/16118*
54. *Wang, H., Chen, K., Zhu, W., & Song, Z. (2015). A process model on P2P lending. Financial Innovation, 1(1), 3. https://doi.org/10.1186/s40854-015-0002-9*
55. *Wang, J., Yan, S., & and Huang, Y. (2016). Evaluating the regulatory scheme for internet finance in China: the case of peer-to-peer lending. China Economic Journal, 9(3), 272-287. https://doi.org/10.1080/17538963.2016.1214381*
56. *Wang, Z., Cuiqing, J., Huimin, Z., & and Ding, Y. (2020). Mining Semantic Soft Factors for Credit Risk Evaluation in Peer-to-Peer Lending. Journal of Management Information Systems, 37(1), 282-308. https://doi.org/10.1080/07421222.2019.1705513*
57. *Xia, Y. (2019). A Novel Reject Inference Model Using Outlier Detection and Gradient Boosting Technique in Peer-to-Peer Lending [Article]. IEEE Access, 7, 92893-92907, Article 8758218. https://doi.org/10.1109/ACCESS.2019.2927602*
58. *Xia, Y., He, L., Li, Y., Liu, N., & Ding, Y. (2020). Predicting loan default in peer-to-peer lending using narrative data [Article]. Journal of Forecasting, 39(2), 260-280. https://doi.org/10.1002/for.2625*
59. *Xia, Y., Liu, C., & Liu, N. (2017). Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending. Electronic*

Commerce Research and Applications, 24, 30-49. https://doi.org/https://doi.org/10.1016/j.elerap.2017.06.004

60. *Xia, Y., Yang, X., & Zhang, Y. (2018). A rejection inference technique based on contrastive pessimistic likelihood estimation for P2P lending. Electronic Commerce Research and Applications, 30, 111-124. https://doi.org/https://doi.org/10.1016/j.elerap.2018.05.011*
61. *Xia, Y., Zhao, J., He, L., Li, Y., & Yang, X. (2021). Forecasting loss given default for peer-to-peer loans via heterogeneous stacking ensemble approach. International Journal of Forecasting, 37(4), 1590-1613. https://doi.org/https://doi.org/10.1016/j.ijforecast.2021.03.002*
62. *Xu, H. (2022). Angel or Demon? The Ethics of Online Peer-to-Peer Lending Firms*. In Dialogue with China: Opportunities and Risks (pp. 103-113). https://doi.org/10.1142/9789811250231_0010*
63. *Yang, Q., & Lee, Y.-C. (2016). Influencing factors on the lending intention of online peer-to-peer lending: Lessons from renrendai. com. The Journal of Information Systems, 25(2), 79-110. https://doi.org/https://doi.org/10.5859/KAIS.2016.25.2.79*
64. *Yang, X., Fan, W., Wang, L., Yang, S., & Wang, W. (2019). Risk control of online P2P lending in China based on health investment [Article]. Ekoloji, 28(107), 2013-2022, Article e107232. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85063878650&partnerID=40&md5=6730448c46c5eb5c6cf98fcf02e7b15a*
65. *Yang, X., Fan, W., & Yang, S. (2020). Identifying the Influencing Factors on Investors' Investment Behavior: An Empirical Study Focusing on the Chinese P2P Lending Market. Sustainability (Switzerland), 12(13).*
66. *Yum, H., Lee, B., & Chae, M. (2012). From the wisdom of crowds to my own judgment in microfinance through online peer-to-peer lending platforms. Electronic Commerce Research and Applications, 11(5), 469-483. https://doi.org/https://doi.org/10.1016/j.elerap.2012.05.003*
67. *Zhang, D., Zhang, Z., & Managi, S. (2019). A bibliometric analysis on green finance: Current status, development, and future directions. Finance Research Letters, 29, 425-430. https://doi.org/https://doi.org/10.1016/j.frl.2019.02.003*