

A Ph.D. Qualifier Report

by Lennart John Baals

26.01.2024

Table of Contents

1. Ph.D. Progress

- Framing the Doctoral Thesis
- Research Output
- Thesis Structure
- Ph.D. Research Events
- Timeline and Near-term Progression of Studies

2. Details on the Current Research Methodology

- Ideas from the proposal
- Notation
- α Threshold
- Range of α
- Step 2: Supervised Model

3. Appendix

- References
- Alternative Consideration: $d_{\alpha_{min}}$
- Research Question
- Network Centrality Measures
- Abstract Overview

The Big Picture of My Doctoral Thesis

Research Focus on FinTech and Digital Finance

- **Context:** With rapid technological advancements, the financial industry is experiencing a profound digital transformation (Milian et al., 2019). The emergence of digital assets like NFTs, DeFi, and the rise of novel P2P lending practices have created new research paths to investigate value creation and risk assessment within these emerging markets.
- **Objectives:** The thesis seeks to:
 - ① Analyze the price dynamics of blockchain-based digital assets.
 - ② Enhance risk assessment approaches in P2P lending with network topology and machine learning.
 - ③ Improve financial model transparency and interpretability using XAI methods.
- **Impact:** The insights and methodologies crafted in this thesis aim to empower investors, policymakers, and institutions in adapting to the digital finance era, ensuring informed and strategic decision-making.

Research Output: Published & Under Review

Published Paper:

- Wang, Y., Horky, F., **Baals, L. J.**, Lucey, B. M., & Vigne, S. A. (2022). Bubbles all the way down? Detecting and date-stamping bubble behaviours in NFT and DeFi markets. *Journal of Chinese Economic and Business Studies*, 20(4), 415-436.

Under Review:

- **Baals, L. J.**, Liu, Y., Osterrieder, J., & Hadji-Misheva, B. (2023). Leveraging Network Topology for Credit Risk Assessment in P2P Lending: A Comparative Study under the Lens of Machine Learning. Manuscript submitted for publication to *Expert Systems with Applications*.
- **Baals, L. J.**, Liu, Y., Osterrieder, J., & Hadji-Misheva, B. (2023). Network Centrality and Credit Risk: A Comprehensive Analysis of Peer-to-Peer Lending Dynamics. Manuscript submitted for publication to *Finance Research Letters*.
- **Baals, L. J.**, et al., Mitigating Digital Asset Risks (2023). Manuscript submitted for publication to *Financial Innovation*. Available at SSRN: <https://ssrn.com/abstract=4594467>

Research Output: Working Papers

Working Papers:

- **Baals, L. J.** et al. Towards a Research Agenda on the Financial Economics of NFT's (2022). Available at SSRN: <https://ssrn.com/abstract=4070710>
- **Baals, L. J.**, et al. Art and NFT. Drafting process. Planned submission to the *Journal of Corporate Finance*.
- **Baals, L. J.**, Liu, Y., Osterrieder, J., & Hadji-Misheva, B. Drafting process. Alpha Threshold Tuning: An Edge Pruning Approach to Network Simplification for Improved Default Prediction in P2P Lending.

Thesis Chapter Structure

Thesis Title: Essays in FinTech and Digital Finance

Chapter	Title	Content Description
1	Blockchain-based Finance: Pricing Digital Assets	Two studies focusing on the NFT and DeFi markets. The first study employs SADF and GSADF tests to detect speculative bubbles, revealing their frequency and correlation with market sentiments. The second study pioneers the use of the Repeat-Sales Regression (RSR) model for NFT art valuation, proposing a pricing index that factors in collection reputation, thereby offering a novel pricing benchmark for this new market domain.
2	Network-Topological Credit Risk Modeling in P2P Lending	Studies focused on the application of network topology in P2P lending credit risk assessment. The first study introduces a machine learning (ML) framework that integrates network centrality metrics, such as degree centrality, with conventional credit risk factors to refine default prediction models. The second study advances this approach by employing a supervised ML method to differentiate between defaulted and non-defaulted loan networks, proposing two centrality-based indicator variables to assess loan default likelihood.

Thesis Chapter Structure (Cont'd)

Chapter	Title	Content Description
3	XAI Methods for Finance	Studies focused on the integration of eXplainable Artificial Intelligence (XAI) methods in finance. The first study plans to evaluate credit risk models by introducing an 'explanatory distance measure' that seeks to measure the robustness of explanations provided by AI models in P2P lending. The second study is planned to address the applicability of XAI in the operational environment of financial institutions, highlighting constraints and proposing further solutions for model transparency.

Academic Cohesion within the Thesis

- **From Digital Assets to Credit Risk:**

- The thesis begins by exploring the speculative dynamics and systematic valuation in *blockchain-based finance*, particularly focusing on DeFi and NFTs as novel asset classes within digital markets.

- **Bridging Markets and Methodology:**

- Transitioning from asset pricing to risk assessment, the thesis then connects the speculative aspects of digital finance to more traditional financial practices, applying *network-topological models* to P2P lending. This signifies a shift from market analysis to risk management using innovative ML techniques.

- **Enhancing Interpretability and Decision-making:**

- Finally, the thesis shifts to an in-depth exploration of *XAI methods for finance*, enhancing the transparency and interpretability of complex ML models used in credit risk assessment. This chapter not only complements the technical methodologies but also addresses the need for clarity and accountability in FinTech applications.

- **Thematic Continuity:**

- Throughout the chapters, there is a thematic continuity that reflects the evolution of FinTech from a niche market to a pivotal component of modern financial systems, emphasizing the integration of innovative technologies with financial practices to improve accuracy, transparency, and trust.

Participation in Events & Conferences (1/2)

- **Crypto Assets and Digital Asset Investment Conference:**
 - Rennes Business School, April 7-8, 2022.
 - Paper entitled "Towards a Research Agenda on the Financial Economics of NFTs."
- **1st Conference on International Finance, Sustainable and Climate Finance and Growth (CINSC):**
 - Universitá degli Studi di Napoli 'Parthenope', June 12-14, 2022.
 - Presentation of "Towards a Research Agenda on the Financial Economics of NFTs."
- **COST Action FinAI: FinTech and AI in Finance - Training School:**
 - University of Twente.
 - June 12, 2023 - June 16, 2023.
 - Presented preliminary results from the first co-authored working paper with Ph.D. colleague Yiting Liu.

Participation in Events & Conferences (2/2)

- **COST Action FinAI meets Brussels: AI in Finance - Policy Implications Conference:**
 - May 15, 2023 - May 16, 2023.
- **Shenzhen Technology University - International Week:**
 - September 11 - 15, 2023; Shenzhen, China.
 - An Introduction to Asset Management.
- **8th European COST Conference on Artificial Intelligence in Finance:**
 - September 29, 2023.
 - Support in the Organisation and Conference Execution.
 - Paper entitled "Leveraging Network Topology for Credit Risk Assessment in P2P Lending: A Comparative Study under the Lens of Machine Learning" with Yiting Liu and Dr. Branka Hadji-Misheva.

Ph.D. Curriculum Milestones

- **Structural Ph.D. Component - Research Modules:**

- 50 ECTS achieved at Ph.D. level from my first year in the Ph.D. programme at Trinity College Dublin.

Degree Relevant Presentation/UT_PhD_Promoter_presentations,

Ph.D. Timeline

Timeline and plan for the forthcoming Ph.D. completion is structured as follows:

- **01.09.2021 - 01.11.2022:** Completion of 10 modules at Ph.D. level at the Doctoral programme of the University of Dublin, Trinity College. Research conducted on the first chapter of the thesis.
- **01.12.2022 - 31.12.2023:** Transition to the SNF-based research project on network-based credit risk modelling. Research conducted on the second chapter of the thesis.
- **01.01.2024 - 01.08.2024:** Commencement of work on the third working paper (α threshold network approach). Complete the draft of the first and second chapter. Moreover, conceptualisation and preliminary work on the third chapter about XAI methods for finance.
- **01.08.2024 - 28.02.2025:** Analysis and writing up of the first research paper for the third chapter on XAI methods in finance.
- **01.03.2025 - 31.08.2025:** Planned research stay at the Columbia University in the city of New York, USA, collaborating on the final chapter of the thesis with Prof. Dr. Ali Hirsa (Director of Financial Engineering, IEOR, Columbia University).
- **01.09.2025 - 31.10.2025:** Revisions and finalization of the third chapter, and initiation of the synthesis for the overarching thesis narrative.
- **01.11.2025 - 30.11.2025:** Completion of any remaining research, final thesis writing, and preparation for the oral defense.

Table of Contents

1. Ph.D. Progress

- Framing the Doctoral Thesis
- Research Output
- Thesis Structure
- Ph.D. Research Events
- Timeline and Near-term Progression of Studies

2. Details on the Current Research Methodology

- Ideas from the proposal
- Notation
- α Threshold
- Range of α
- Step 2: Supervised Model

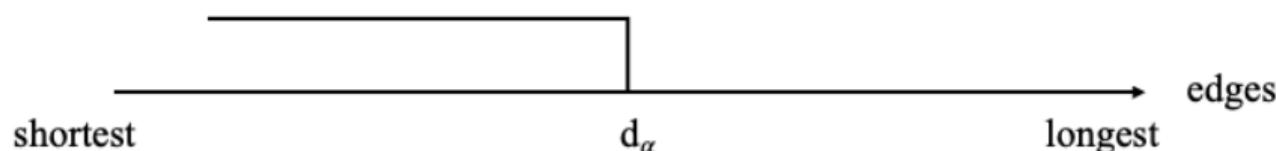
3. Appendix

- References
- Alternative Consideration: $d_{\alpha_{min}}$
- Research Question
- Network Centrality Measures
- Abstract Overview

Idea of α Threshold

Idea from proposal:

- The α of a threshold network is a hyperparameter that we plan to optimize through k -fold cross-validation.
- Various threshold networks have been used before in finance (Billio et al., 2012; Onnela et al., 2004), where the threshold is a random variable but the number of retained edges was not optimized.
- Our research idea is similar to scholarly work where network percolation is studied; i.e., the set of edges is expanded by adding edges starting with the smallest distance $d_1 < d_2 < \dots$ until a certain topological property of the network changes.



Idea of α Threshold

Research path in P2P lending:

- We could explore the α network that will lead to the highest difference between the centrality distribution of defaulted and non-defaulted loans.
- Several centrality measures, including closeness, betweenness, eigenvector centrality, and PageRank, can be explored.
- To quantify the difference in centrality distributions between defaulted and non-defaulted loans, the two-sample Kolmogorov-Smirnov (K-S) test can be employed. This non-parametric test measures the maximum distance between the empirical distribution functions of two samples.

Notation

Notation	Explanation
N_{edges}	Total number of edges in a fully connected graph.
α	$\alpha \in [\alpha_{\min}, 1]$. Keep the shortest $\alpha \cdot N_{\text{edges}}$ to build a sub-graph.
α_{\min}	The smallest α to keep the graph connected.
$\alpha_{\text{left}}, \alpha_{\text{right}}$	The left and right endpoints of interval during binary search of α_{\min} .
d_α	Length of the threshold edge. $\frac{ \{ij: d_{ij} \leq d_\alpha\} }{N} = \alpha$.

Mathematical Background

The α threshold concept is a methodology employed to extract a subgraph from a fully connected graph. In network theory, especially when considering large graphs, it's occasionally necessary to concentrate on a subset of edges more informative or relevant to a particular application.

Method: Given a fully connected graph with N_{edges} , these edges can be ranked or ordered based on a criterion, typically their weights.

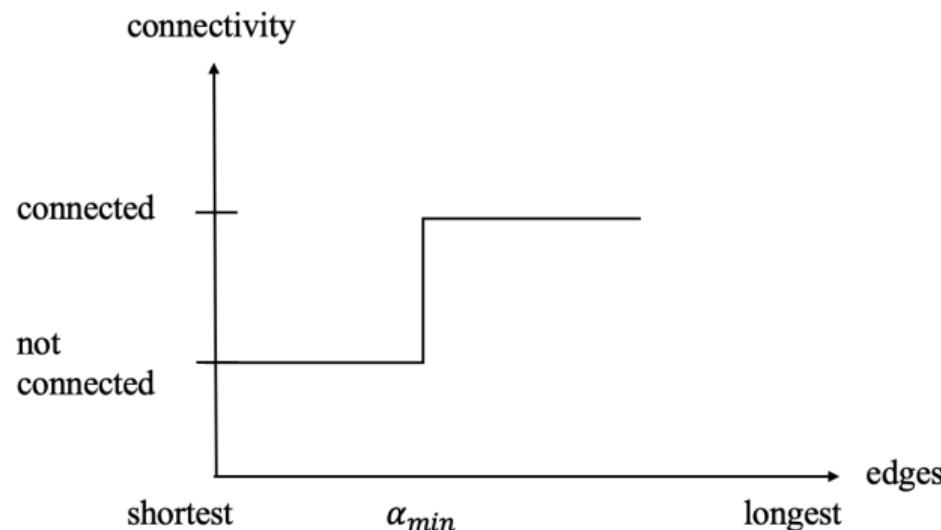
- The α threshold thus determines the proportion of the shortest edges retained to construct the subgraph.

Mathematically, if edges are ranked in ascending order by weight and if e_i represents the i-th edge, then an α threshold would retain the first $\alpha \times N_{\text{edges}}$ to form the subgraph.

- $\alpha = 1$: All edges are retained, resulting in the original fully connected graph.
- α_{\min} : The smallest threshold where the subgraph remains connected.

α_{\min}

- $\alpha \in [\alpha_{\min}, 1]$
- $\alpha = 1$ is the maximum value of α , which will result in a fully connected graph.
- α_{\min} is the minimum value of α to keep the graph connected. We use binary search to find the value of α_{\min} .



Necessity for a Connected Graph

Importance of Graph Connectivity

- Within the alpha threshold framework, a connected graph ensures that each loan is comparable to another loan, enabling a comprehensive analysis of similarity patterns.
- Graph connectivity is essential and crucial for the precise computation of centrality measures, which presuppose the existence of paths between all loan pairs in the network.

Implications for P2P Lending

- A connected graph facilitates the assessment of the network's structure and the associated information.
- It offers a holistic perspective on the interconnectedness and relative positioning of loans within the network.
- By setting α to α_{\min} , the network is pruned to its essential structure, highlighting the most significant loan similarities and disregarding weaker connections.

Finding α_{\min} : Binary Search Algorithm

Binary search, a fundamental algorithm in computer science, is efficient for finding a value in a sorted array. For determining α_{\min} , we adapt this algorithm:

- ① **Initialization:** Begin with an interval for α , where $\alpha_{\text{left}} = 0$ (no edges retained) and α_{right} includes edges up to the longest edge in the MST for efficient searching space. The rationale for this lies in the fact that by removing the MST's longest edge, the resulting graph might be disconnected, making it a suitable starting point for efficient search.
- ② **Middle Point:** Compute a midpoint value α_{mid} as the average of α_{left} and α_{right} .
- ③ **Check Connectivity:** Generate a subgraph using the α_{mid} threshold and verify its connectivity.
- ④ **Update Interval:**
 - If connected, α_{\min} might be smaller. Set α_{right} to α_{mid} .
 - If not, α_{\min} is larger than α_{mid} . Adjust α_{left} to α_{mid} .
- ⑤ **Loop:** Repeat steps 2-4.
- ⑥ **Termination:** Continue until the interval between α_{left} and α_{right} becomes negligible or the subgraph's connectivity remains unchanged.

α as a Hyperparameter

- We choose an $\alpha \in [\alpha_{\min}, 1]$. Then we eliminate edges beyond α in the fully connected graph.
- We calculate the centrality measures.
- Supervised models.
- α will be a hyperparameter to tune.

Table of Contents

1. Ph.D. Progress

- Framing the Doctoral Thesis
- Research Output
- Thesis Structure
- Ph.D. Research Events
- Timeline and Near-term Progression of Studies

2. Details on the Current Research Methodology

- Ideas from the proposal
- Notation
- α Threshold
- Range of α
- Step 2: Supervised Model

3. Appendix

- References
- Alternative Consideration: $d_{\alpha_{min}}$
- Research Question
- Network Centrality Measures
- Abstract Overview

References

-  Billio, M., Getmansky, M., Lo, A. W., & Pelizzon, L. (2012).
Econometric measures of connectedness and systemic risk in the finance and insurance sectors.
Journal of financial economics, 104(3), 535-559.
-  Kruskal, J. B. (1956).
On the shortest spanning subtree of a graph and the traveling salesman problem.
Proceedings of the American Mathematical society, 7(1), 48-50.
-  Milian, E. Z., Spinola, M. D. M., & de Carvalho, M. M. (2019).
Fintechs: A literature review and research agenda.
Electronic Commerce Research and Applications, 34, 100833.
-  Onnela, J. P., Kaski, K., & Kertész, J. (2004).
Clustering and information in correlation based financial networks.
The European Physical Journal B, 38(2), 353-362.

Alternative Consideration: $d_{\alpha_{min}}$ can be no shorter than the longest edge in the MST

Proof:

- Consider the Kruskal's Algorithm (Kruskal, 1956), which grows a MST by arranging edges from shortest to longest. Edges are added to the tree if they are acyclic with the existing nodes in the tree (two nodes connected by this edge are not in the same disjoint set).
- Initially, there are n disjoint sets for all nodes. The edges are sorted from the shortest to the longest edge. For the i -th edge, there are two situations:
 - Two nodes connected by the i -th edge are not in the same disjoint sets. Then, both algorithms add this edge. The number of disjoint sets decreases by 1.
 - Two nodes connected by the i -th edge are in the same disjoint set. Then, Kruskal's Algorithm will not add this edge, while the α threshold method will add.
- In both cases, both algorithms will make the same changes to the number of disjoint sets for all nodes (-1 or unchanged).
- This process continues until the number of disjoint sets decreases to 1, i.e., all nodes are in the same disjoint set, and the MST is found. Thus, both algorithms will stop at the same edge, which is the longest edge in the MST.
- Therefore, the α threshold method will have edges equal or more to the MST. The longest edge in graphs generated by both algorithms will have the same length.**

Overarching Research Question

Research Question: *"How are emerging FinTech innovations reshaping financial market dynamics and risk assessment strategies?"*

This study investigates the impact of blockchain-based digital assets and network-based credit risk models on the financial industry. It aims to bridge the gap in literature by connecting technological advancements with their financial implications, serving the interests of stakeholders and informing policy and strategy in the digital economy.

Hypotheses

- **H1: Digital Asset Dynamics** - Blockchain assets, like NFTs and DeFi, create new market behaviors, challenging established financial theories.
- **H2: Network-based Credit Risk** - Network topology improves predictive accuracy and risk stratification in P2P lending over traditional methods.
- **H3: FinTech Integration** - FinTech's integration into finance will fundamentally shift risk management and asset valuation strategies.

Centrality Measures

- **Degree Centrality**

$(C_D(i) = \sum_{j=1, j \neq i}^n I[i \text{ and } j \text{ are connected}])$ quantifies the number of direct connections a node has in the network.

- **Closeness Centrality**

$(C(i) = \frac{1}{\sum_{j=1, j \neq i}^n d(i,j)}$, where $d(i,j)$ is the shortest-path distance between i and j) captures how quickly information can propagate from a given node to others.

- **Betweenness Centrality**

$(C_B(i) = \sum_{j,k \in \{1,2,\dots,n\}} \frac{\sigma(j,k|i)}{\sigma(j,k)}$, where $\sigma(j, k)$ is the total number of shortest paths from node j to node k , and $\sigma(j, k|i)$ is the number of those paths passing through node i) captures the influence of a node over the flow of information between other nodes in the network.

- **PageRank**

$(PR(i) = (1 - d) + d \sum_{j \in M(i)} \frac{PR(j)}{L(j)}$, where $M(i)$ is the set of pages that link to i , $L(j)$ is the number of outbound links on page j , and d is a damping factor) evaluates the importance of nodes based on the quality of incoming links.

- **Katz Centrality**

$(C_{Katz}(i) = \sum_{j=1}^n \beta A_{ij} C_{Katz}(j) + \alpha$, where A_{ij} denotes the adjacency matrix element, β is a scaling factor, and α is a constant term representing the node's intrinsic centrality) considers both direct and indirect influence of a node's neighbors.

- **Authority Score and Hub Score**

$(a(i) = \sum_{j \in M(i)} h(j)$, where $M(i)$ is the set of nodes that point to i ; $h(i) = \sum_{j \in N(i)} a(j)$, where $N(i)$ is the set of nodes that i points to).

Abstract Overview

- **Title:**

- *Bubbles all the way down? Detecting and date-stamping bubble behaviours in NFT and DeFi markets*

- **Abstract:**

- Amid rising market values and regulatory debates, NFT and DeFi markets are often viewed as mere speculative domains.
 - Using Supremum Augmented Dickey-Fuller (SADF) and Generalized Supremum Augmented Dickey-Fuller (GSADF) tests, we identify and date price bubbles in these markets.
 - Both markets show speculative bubbles, but NFTs display more frequent and intense bursts.
 - These bubbles correlate strongly with market buzz and broader cryptocurrency volatility.
 - However, bubble-free periods indicate underlying intrinsic value in these markets, opposing the view of them as purely speculative instruments.

Abstract Overview

- **Title:**

- *Mitigating Digital Asset Risks*

- **Abstract:**

- Digital assets, encompassing cryptocurrencies, tokenized securities, stablecoins, NFTs, and CBDCs, promise to revolutionize financial markets with new business models and efficient transactions.
 - However, they bring challenges like fraud, market manipulation, and regulatory ambiguities.
 - This paper delves into the digital asset landscape, discussing their classifications, technological underpinnings like blockchain and DeFi, and regulatory scenarios.
 - Key recommendations include crafting regulations that balance innovation with consumer protection, fostering global regulatory alignment, and utilizing regulatory sandboxes for continuous adaptation.
 - Effective regulation, in collaboration with stakeholders, can optimize the potential of digital assets while minimizing risks.

Abstract Overview

- **Title:**

- *Leveraging Network Topology for Credit Risk Assessment in P2P Lending: A Comparative Study under the Lens of Machine Learning*

- **Abstract:**

- Peer-to-Peer (P2P) lending markets have witnessed remarkable growth, changing how borrowers and lenders interact.
 - Despite its popularity, P2P lending presents challenges in credit risk assessment and default prediction.
 - Traditional credit risk models, used in P2P lending, might not capture the loan networks complexity and borrower behavior nuances.
 - This study introduces a two-step machine learning approach, which begins with network analysis insights.
 - It combines network centrality metrics with traditional credit risk factors, aiming to enhance prediction accuracy in credit risk modelling.

Abstract Overview

- **Title:**

- *Network Centrality and Credit Risk: A Comprehensive Analysis of Peer-to-Peer Lending Dynamics*

- **Abstract:**

- This study explores credit risk assessment in the Peer-to-Peer (P2P) lending field using a dataset from Bondora, a top European P2P platform.
 - Combining traditional credit features with network topological features, especially degree centrality, reveals the importance of a borrower's position in the P2P network for loan default probabilities.
 - Robustness checks with shuffled centrality features emphasize the value of integrating financial and network traits in credit risk assessments.
 - The findings provide fresh insights into credit risk determinants in P2P lending and assist investors in gleaning valuable information from P2P loan networks.

Abstract Overview

- **Title:**

- *Art and NFT*

- **Abstract:**

- As digital art gains prominence with the rise of Non-Fungible Tokens (NFTs), the need for systematic valuation methods becomes imperative.
 - This study introduces the application of the Repeat-Sales Regression (RSR) model, typically used in physical asset markets, to the digital world of NFTs.
 - Through this approach, we aim to construct a pricing index for digital art NFTs, shedding light on the interplay between collection reputation, valuation, and evidence of a "masterpiece" effect in NFT art collections.
 - This paper delves into the dynamics of various NFT art genres, providing insights for artists, investors, and the wider NFT community.
 - By leveraging the RSR model, the research strives to set a benchmark for future pricing methodologies in the rapidly evolving NFT art market.