BLOCKCHAIN BEYOND BORDERS: INTERDISCIPLINARY PERSPECTIVES AND EVOLVING LEGITIMACY IN ICO-DRIVEN ENTREPRENEURSHIP

by

MUHAMMAD NAUMAN SHAHID

(M.S. Comp. Eng., Saint Petersburg Polytechnic University)

A DISSERTATION SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

INFORMATION SYSTEMS AND ANALYTICS

in the

GRADUATE DIVISION

of the

SCHOOL OF COMPUTING

NATIONAL UNIVERSITY OF SINGAPORE

2024

Supervisor:
Professor HAHN Jungpil

Examiners:
Associate Professor TAN Chuan Hoo
Assistant Professor QIAO Dandan

Declaration

I affirm that this thesis is my original work, written and compiled independently. All sources, as well as digital and software tools used in the thesis, have been carefully cited and acknowledged.

I further confirm that this work has not been submitted for a degree or other academic qualification.

Muhammad Nauman Shahid

13 June 2024

To my mentor, friends, and family

Acknowledgments

It is with deep gratitude that I acknowledge the invaluable support of those who have contributed to this journey. Professor Jungpil Hahn, my advisor, to whom I owe my deepest gratitude for his unwavering support, insightful guidance, and constructive feedback throughout my research. His expertise and mentorship have shaped this work in many ways.

I would like to express my sincere thanks to Professor Chuan Hoo Tan and Professor Danadan Qiao for their invaluable guidance and support. I would also like to thank the Department of Information Systems and Analytics at the National University of Singapore for providing me with a stimulating environment focused on Information Systems research, which nurtured my academic growth and fuelled my research endeavours.

Special thanks go to the members of the Garbage Can Lab for their time, suggestions, and insightful feedback that has enriched this work in many ways. I also thank NUS for its generous financial support through NUS Research Scholarship, which helped me greatly in this research.

My heartfelt gratitude goes to my colleagues and friends at NUS for their constant encouragement, camaraderie, and unwavering belief in me. I am also deeply grateful to my family for their steadfast love and support, which have been my pillars of strength throughout this journey.

Finally, I would like to thank all those who were indirectly involved in this project. Their contributions, however small, have not gone unnoticed. Every helping hand and every encouraging word has helped to shape this journey. Thank you to all.

Contents

Abstract

Blockchain Beyond Borders: Interdisciplinary Perspectives and Evolving Legitimacy in ICO-driven Entrepreneurship

by

Muhammad Nauman Shahid

Doctor of Philosophy in Information Systems and Analytics

National University of Singapore

This dissertation comprises two essays that examined the interdisciplinary landscape of blockchain research and shed light on the paradigm shift in organisational legitimacy in the context of Initial Coin Offerings (ICOs).

The first essay[1] maps blockchain literature across disciplines. Using topic modelling and statistical analysis on 35,604 academic articles, it found a growing trend of interdisciplinarity. This challenges the conventional wisdom that blockchain belongs to a single discipline while uncovering new research themes and trends. Furthermore, it introduces the IDEA framework to capture ten different aspects of research across 15 disciplines, providing a foundation for future studies to measure the interdisciplinary nature of emerging technologies.

The second essay[2] examines cultural entrepreneurship theory through the lens of organisational identity building. Organisational legitimacy, crucial for attracting resources, is traditionally signalled through methods like corporate social responsibility and media management. However, these methods are often costly and biased, making them unsuitable for early-stage entrepreneurial ventures, that lack significant financial resources and face challenges in overcoming established biases. With the rise

[1]An earlier version of this study, titled "A Cross-Disciplinary Review of Blockchain Research Trends and Methodologies: Topic Modeling Approach," was presented and published at the 53rd *Hawaii International Conference on System Sciences* (HICSS), in Maui, Hawaii, United States.

[2]An earlier version of this study, titled "Pick The Odd-Ones Out! Conferring Legitimacy Of Initial Coin Offerings By Distinction," was presented and published at the 42nd *International Conference on Information Systems* (ICIS) in Austin, Texas, United States.

of Initial Coin Offerings (ICOs), a blockchain-based financing method, traditional signalling methods have become less effective, leading early-stage entrepreneurial ventures to use entrepreneurial storytelling as a signalling method and to seek alternative sources of legitimacy.

This essay argues that optimal distinctiveness in entrepreneurial storytelling can effectively transmit organisational legitimacy to resource-providing audiences, challenging the notion that optimal distinctiveness and legitimacy are incompatible. It suggests that optimal distinctiveness can serve as a signalling source for legitimacy in early-stage entrepreneurial ventures seeking ICO funding. Furthermore, the essay explores the role of online media as a tool that can influence the distinctiveness-legitimacy relationship. Strategically managed online media can have a similar effect as traditional media; however, unlike traditional media, online media can shift organisational legitimacy away from an organisation's strategic monopoly. Analysing data from 306 ICOs in 29 market categories, the results confirm that optimal distinctiveness of entrepreneurial storytelling in ICOs can legitimise early-stage entrepreneurial ventures and that online media perceptions play a significant role in this process.

Keywords: Blockchain, literature review, topic modelling, latent dirichlet allocation, entrepreneurship, organisational legitimacy, optimal distinctiveness, online media, initial coin offerings.

List of Figures

List of Tables

Chapter 1

Introduction

The widespread impact of technology on contemporary society has become increasingly evident. It transforms numerous aspects of our lives, from the way we learn and work to the way we conduct digital transactions. A prime example of this technological revolution is the rise of blockchain technology (BCT). This innovation is growing at an exponential rate, alongside other rapidly advancing technologies such as Artificial Intelligence (AI), the Internet of Things (IoT), 3D printing, and Computer Vision (Kurubacak et al., 2022).

BCT has found significant applications in non-academic settings, particularly in industries and organisations that offer various products and services (Swan, 2015). Alongside these developments, Initial Coin Offerings (ICOs) platforms are reshaping the financial landscape, complementing the transformative influence of BCT and other digital innovations in various industries and organisations. This surge in practical implementation has spurred robust academic interest, and scholarly work on crowdfunding platforms is accumulating at a rapid pace.

The application contexts of blockchain technology (BCT) are diverse and require the integration of knowledge from a wide range of disciplines, including computer science, economics, social sciences, and law (Shahid and Hahn, 2020; Janssen et al., 2020). The multifaceted nature of BCT underscores the need for an understanding that transcends the boundaries of individual academic disciplines, highlighting its inherent complexities. Complex technologies such as BCT are fundamentally interdisciplinary and encompass different fields of study. In this context, interdisciplinary refers to the integration of concepts, theories, or methods from two or more distinct disciplines to create new knowledge or solve complex problems. An isolated understanding of these technologies, confined to the propositions of a single discipline, may inadvertently impede their growth. Therefore, it is of utmost importance to determine whether BCT research is truly interdisciplinary, to what extent, and whether existing scientific work on BCT reflects this interdisciplinary nature.

Recognising that an isolated understanding of these technologies, confined to the parameters of a single discipline, may inadvertently impede their growth, assessing the extent of interdisciplinarity in current BCT research and how this translates into practical applications and shifts in traditional paradigms is imperative. BCT's rapid evolution necessitates a comprehensive, interdisciplinary understanding, a theme central to my first essay. Here, I explore the breadth of interdisciplinarity in BCT research, addressing a critical gap in understanding its multifaceted implications. Concurrently, BCT challenges existing paradigms, particularly in reshaping the legitimacy of new entrepreneurial ventures, a topic that my second essay delves into. These essays offer a dual perspective on BCT, revealing not only its diverse implications but also its role as an effective catalyst for interdisciplinarity. The first essay maps the web of research across disciplines, while the second essay illuminates how this network manifests itself in tangible consequences, challenging and reshaping the legitimacy landscape for emerging entrepreneurial ventures.

In my first essay (Essay 1), I conducted a comprehensive examination of 35,604 academic articles and provided a detailed overview of blockchain research across various academic disciplines. I employed a novel approach that combines topic modelling using Latent Dirichlet Allocation (LDA) with a framework I developed, which incorporates four different metrics. This methodology allowed me to assess the interdisciplinary nature of blockchain research in 109 academic disciplines, further categorised into 15 broader discipline clusters for a more streamlined analysis.

My approach provides a thorough analysis of the nature and extent of interdisciplinarity in BCT research, incorporating both qualitative and quantitative perspectives. This is a unique contribution, as existing literature reviews often overlook this aspect. It is important to recognise that, while some topics are highly interdisciplinary, others are narrowly confined to specific academic fields. My analysis of emerging topics from 2013 to 2022 indicates a positive trend towards greater interdisciplinarity. This development suggests a gradual breakdown of disciplinary barriers and the promotion of a more unified blockchain research landscape.

The findings of this study can be a valuable resource for academic and non-academic stakeholders, especially those interested in the overall development of technology and the field.

In Essay 2, an analysis of 306 Initial Coin Offerings (ICOs) explores how blockchain-based funding options, like ICOs, disrupt traditional ideas of organisational legitimacy in early-stage entrepreneurial ventures. This study discusses the concept of distinctive entrepreneurial storytelling to highlight the increasing importance of organisational identity management in the context of ICOs. The findings suggest that unique entrepreneurial stories can act as costly signals to resource-providing audiences, improving the legitimacy of new entrepreneurial ventures. The online media also plays a role in mediating the relationship between legitimacy and entrepreneurial storytelling. This examination implies that traditional assumptions about the conferment of legitimacy may have changed in the landscape of ICOs. Therefore, new ICOs may struggle to attract investors unless they reevaluate the notions of legitimacy. The study highlights the importance of distinctiveness in conveying legitimacy and emphasises the significant role that online media play in influencing perceptions of legitimacy.

Overall, the use of two interconnected lenses – the interdisciplinary nature of blockchain within academic research and its impact on the legitimacy of new entrepreneurial ventures – makes significant and unique theoretical and practical contributions to both interdisciplinary research and cultural entrepreneurship.

To delve into the interdisciplinary aspect, the first study introduces a novel framework to map interdisciplinary blockchain research, uncovering hidden collaboration opportunities and potential areas for future research previously overlooked by traditional methods. This framework serves as a valuable tool for researchers seeking potential collaborators, funding agencies evaluating the merit of projects, and the broader blockchain community in identifying gaps and bottlenecks that may be hindering the technology's advancement. The innovative combination of topic modelling (LDA) and robust statistical metrics offers a methodological advancement applicable beyond blockchain research.

The second study makes a substantial contribution to cultural entrepreneurship theory. It establishes a direct link between distinctiveness in entrepreneurial storytelling and the perceived legitimacy of new ventures, highlighting the amplifying role of online media and the moderating influence of market category viability. These findings provide entrepreneurs with actionable insights to effectively manage their identity in today's competitive and media-saturated landscape. Moreover, the study's insights into investor behaviour in the context of ICOs contribute to a deeper understanding of how new ventures are evaluated in the digital age, with potential implications for policy and regulation.

Chapter 2

Which Block Are You From? An Interdisciplinary Review of Blockchain Technology Research

2.1 Introduction

Blockchain technology (BCT) has generated various responses, ranging from enthusiastic support to sceptical criticism from academia, industry, and the general public. Advocates of BCT praise its ability to facilitate economic transactions, as emphasised by (Glaser, 2017). In contrast, critics question the practical utility of blockchain implementations, expressing doubts about their efficiency and versatility in different applications (Maurer et al., 2013).

The division between these perspectives in the field of BCT is not just the result of differing opinions but is rooted in the fragmentation of knowledge (Janssen et al., 2020). This fragmentation is evident in the diverse knowledgebase required for the effective development, deployment, and use of blockchain technology. For instance, consider the development of decentralised finance (DeFi) applications built on blockchain technology. This emerging field requires the interplay of computer science to design secure smart contracts and scalable infrastructure, economics to understand tokenomics and financial impacts, law to navigate the evolving regulatory landscape, and social sciences to assess the societal implications of financial inclusion and access. This example highlights the multifaceted complexity of BCT and the need for a comprehensive, interdisciplinary approach (Tapscott and Tapscott, 2016).

The rapid technological advances in the blockchain sector and its global implications introduce additional complexity, including diverse international regulatory and cultural contexts, as well as significant ethical considerations, particularly in the areas of data security and user privacy. The transformative potential of blockchain technology in sectors such as finance, supply chain, and logistics requires a forward-looking and interdisciplinary approach. Addressing these diverse and evolving challenges in the field requires mapping knowledge and integrating it across multiple disciplines. Collaboration and innovation are essential for the development of ethically sound and practical solutions in this dynamic field.

However, despite recognising the need for interdisciplinary expertise, a critical question arises: *Does the existing body of research on BCT adequately reflect interdisciplinarity, and more importantly, does it effectively map the diverse knowledge between disciplines?* Addressing this question is the focus of this study, and it is precisely this gap that our study seeks to fill. Using a descriptive approach, we employ topic modelling and statistical analysis in innovative ways to decipher research patterns, macro-level themes, and temporal trends in BCT-related research publications across various disciplines. The goal is not only to delineate BCT's interdisciplinary nature but also to systematically map the evolving themes and contributions from different fields. This mapping is crucial to developing a comprehensive understanding of BCT, as it reveals how disparate strands of research interconnect and contribute to our overall understanding of the technology. The study's findings aim to provide critical insight into the following research questions.

RQ1: What is the current state of BCT research in various research disciplines?

RQ2: To what extent does the BCT research reflect interdisciplinarity?

We used Latent Dirichlet Allocation (LDA) and a statistical framework incorporating various metrics on 35,604 articles published between 2013 and 2022 to answer these questions. This approach allowed us to paint a comprehensive picture of the interdisciplinary landscape of BCT research. Our analysis revealed that many topics exhibited high interdisciplinarity, while others remained confined to specific academic disciplines. Furthermore, we observed a promising and healthy trend towards greater interdisciplinarity from 2013 to 2022. This trend indicates a gradual dismantling of disciplinary silos and a shift towards a more integrated and collaborative BCT research landscape.

This study offers several important implications for research and practice. First, it improves the understanding of researchers and organisations about interdisciplinary topics in BCT, which is vital for continuous technological progress. It also sheds light on the intricate nature of BCT, highlighting the need for interdisciplinary collaboration in specific research areas for unimpeded advancement. Second, this study introduces a novel approach to conducting, mapping, and interpreting research from the perspective of interdisciplinarity. This methodology could serve as a template for exploring the interdisciplinarity of research discourse related to other complex technologies, potentially influencing future research directions. Finally, the results of this investigation are valuable for both academia and practice, particularly for those engaged in the development of BCT and its applications. Our findings underscore the importance of interdisciplinary collaboration to unlock the full potential of complex technologies such as BCT, paving the way for future developments in this rapidly evolving field.

The remainder of this chapter is organised as follows: Section 2.2 introduces the theoretical framework, reviews the existing literature, and outlines the research gaps. Section 2.3 describes the research approach and data collection methods. Section 2.4 presents an analysis of research trends and discusses the key findings. Finally, Section 2.5 highlights contributions and identifies opportunities for future research.

2.2 Theoretical Background

The emergence of blockchain technology (BCT) as the foundation for many cryptocurrencies marks a new era of disruptive technological innovation, as it integrates into a wide range of use cases across various industry sectors (Benz et al., 2022). Many entrepreneurial organisations are exploring its potential, and its proponents argue that its revolutionary features, including decentralisation, transparency, immutability, security, efficiency, and speed, offer advantages over traditional systems (e.g., Xu et al., 2019; Kumar Bhardwaj et al., 2021; Chaturvedi and Kumaran, 2019). However, critics contend that the technology is susceptible to scalability issues (Khan et al., 2021), environmental impacts (Islam et al., 2022), regulatory challenges (Yeoh, 2017), lack of user knowledge and understanding (Xu et al., 2020), and potential misuse (Boucher et al., 2017; Kuperberg, 2020). Given this landscape, in which the revolutionary potential of BCT is both celebrated and criticised, a nuanced interdisciplinary approach is not only beneficial but essential. Such an approach can harmonise these different perspectives and provide a more balanced and comprehensive understanding of BCT. This is critical to address the concerns of its critics while capitalising on the opportunities identified by its proponents to ultimately promote broader utility and societal acceptance of the technology.

Building on the identified need for a nuanced, interdisciplinary perspective, our research seeks to operationalise this approach through the mapping and synthesis of knowledge from various areas of study relevant to BCT. This integrative approach can bridge the gaps between different viewpoints and reconcile conflicting perspectives, promoting an enlightened understanding of the technology.

Extending our pursuit of an integrative approach to BCT, we evaluated three different research approaches: multidisciplinary, transdisciplinary, and interdisciplinary. In a multidisciplinary approach, a phenomenon is explored from various academic lenses, but integrating these diverse perspectives often remains a challenge due to the lack of integrative mechanisms necessary to bridge disciplinary divides and synthesise different insights (Repko and Rick, 2020, p.75). In contrast, a transdisciplinary approach transcends disciplinary barriers, incorporating insights from academic and non-academic entities, such as industry professionals (Repko and Rick, 2020, p.75), to create a more holistic understanding. While valuable, transdisciplinary approach may dilute the academic rigour due to the broad spectrum of inputs. This study adopts an interdisciplinary approach that aims not only to map insights from various academic disciplines but also to synthesise them. This synthesis is highly relevant to our research because it focuses on multiple disciplines, allowing for a comprehensive exploration that aligns seamlessly with our research objectives.

Building on our interdisciplinary approach to BCT, it becomes imperative to reassess traditional analytical methodologies, such as citation and collaboration analyses, when applied to such complex contexts (Bergstrom et al., 2008). Limitations of existing citation practices include citation biases, negative citations, inefficiency in capturing inherent complexity, and inadequate citation granularity (Worrall and Cohn, 2023; MacRoberts and MacRoberts, 1989). These factors can potentially obscure the interdisciplinarity and dynamism of a field and lead to distorted or biased understanding. Given the breadth of this review, it is necessary to employ more advanced and comprehensive methodologies.

Advanced methodologies such as topic modelling (Blei et al., 2003) combined with appropriate statistical summarisation indicators (e.g., Steele et al., 2022) have the potential to be valuable in this context. The use of advanced methodologies can provide a fresh and nuanced understanding of the interdisciplinarity of the literature, facilitating a holistic investigation of the research literature on BCT. The use of these techniques, along with the associated methodological considerations and identified gaps, is discussed in a subsequent section.

2.2.1 Interdisciplinary Dynamics in Blockchain Research

Blockchain technology, a complex and innovative fusion of diverse disciplines, from cryptography to distributed systems, is a prime example of a crossover technology (National Academy Of Sciences, 2010) with transformative potential in numerous sectors. Current research efforts often focus narrowly on specific disciplines, overlooking the vast potential at the intersection of these fields.

The crossover nature pertains to its ability to emerge from the convergence of multiple disciplines (National Academy Of Sciences, 2010), allowing it to be applied and to influence various industries, domains, or disciplines significantly. BCT is a crossover technology as it emerges from the combination of diverse fields, such as cryptography from mathematics, distributed systems from computer science and engineering, incentive structures and mechanisms from economics, and legal frameworks from law. This interdisciplinary blend parallels the development of the "World Wide Web", which combined computers, communication, and software technologies (National Academy Of Sciences, 2010). BCT applications are as diverse as its foundations, ranging from finance with cryptocurrencies such as Bitcoin and Ethereum (Khan et al., 2021), to healthcare for securing patient records and medication tracking (Hölbl et al., 2018; Chukwu and Garg, 2020), to supply chain management for greater transparency and efficiency (Calzadilla and Villa, 2017; Wang et al., 2019; Dutta et al., 2020), and to digital governance systems exemplified by Estonia's adoption for electronic government services (Tapscott and Tapscott, 2016).

The differences in technological readiness across research disciplines serves as a differentiator to influence BCT's adoption and research. Research disciplines such as computer science and business sciences are naturally aligned with digital innovations, tend to have a more advanced state of blockchain research compared to disciplines such as social sciences and natural sciences, that may not prioritise technologies to the same extent. This difference highlights how BCT's integration depends on each field's familiarity with digital tools and innovations.

Furthermore, each research discipline has unique research objectives and methodologies that shape its approach to blockchain technology. For instance, in finance, blockchain is often explored for its potential to revolutionise transactions and asset management, while in healthcare, the focus might be on secure data sharing and patient privacy. These divergent goals demonstrate the different research trajectories and levels of interdisciplinary collaboration driven by BCT.

BCT serves as a good tool for interdisciplinary research because its disruptive nature challenges established paradigms, creating fertile ground for diverse perspectives to converge. The shared technical foundation and socio-technical complexity of

BCT systems necessitate collaboration across disciplines. Moreover, the open-source nature of many projects and dedicated funding mechanisms further incentivise interdisciplinary research.

Despite its multidisciplinary foundation and transformative potential (Casino et al., 2019; Shen and Pena-Mora, 2018; Yin and Ran, 2021), current research efforts in BCT still seem to focus narrowly on topics bounded by specific disciplines (Shen and Pena-Mora, 2018; Lach, 2014; Lynn et al., 2018; Bammer, 2012; Kher et al., 2020; Risius and Spohrer, 2017), indicating a fragmented exploration (Wastl and Fane, 2024; Balietti et al., 2015) of the wide-ranging capabilities of BCT. This fragmentation may be due to several factors. First, the novelty of blockchain technology led early studies to focus on domain-specific inquiries (Liu et al., 2019). Second, its foundational aspects require specialised analysis (Zaghloul et al., 2020). Third, the initial traction for blockchain was its application in creating cryptocurrencies, such as Bitcoin (Jabbar and Dani, 2020). Therefore, the immediate academic interest revolved around its potential and challenges in the financial sector through the lens of financial and monetary economics. Finally, educational structures (e.g., universities) organised around established disciplines can also pose challenges due to institutional barriers (Glied and Bakken, 2007; Larson et al., 2011), funding structures (Bromham et al., 2016), or peer-review processes (McLeish and Strang, 2016), which are often narrowly confined to specific disciplines. Deeply ingrained disciplinary silos, technical jargon, and the hype surrounding BCT can hinder effective collaboration. Researchers from different fields may struggle to find common ground and communicate effectively due to these barriers.

The inherent characteristics of BCT as a crossover technology necessitate the adoption of a comprehensive and interdisciplinary approach. Researchers are encouraged to analyse the implications of blockchain across various sectors and explore the intersections and interactions of its applications and impacts in different domains. By overcoming disciplinary boundaries, researchers can leverage broader knowledge, diverse perspectives, and varied methodologies, potentially leading to groundbreaking discoveries, more effective and holistic policies, and more robust collaborative ventures.

Disciplinary silos[1] in research not only limit its potential, but also introduce a multitude of problems. First, the evolving nature of research themes increases the risk of overlooking or disregarding valuable knowledge (De Domenico et al., 2016). Second, disciplinary silos can result in redundancy because researchers in different fields may not know about the same challenges or duplicate their efforts (Legendre et al., 2017), further exacerbating the problem of overlooking new knowledge while focusing on already explored areas (Hofer and Pintrich, 1997). Conducting research in isolated domains hinders innovation and prevents potential breakthroughs that could result from collaborative and interdisciplinary approaches (Kreitzer and Saper, 2015). This narrow focus also creates barriers to the broader adoption of blockchain solutions, as it restricts the perspective to specific applications instead of recognising the full potential of the technology. Third, disciplinary silos complicate the crafting of informed decisions from a policy point of view (Smith, 2013). Policymakers may introduce oversight or misjudgements if they lack a comprehensive and holistic understanding of the capabilities and implications of a technology (Talukder and Kuzma, 2008; Dignam, 2020; Marchant, 2011). Furthermore, a narrow perspective inhibits interdisciplinary collaboration, restricting the range of insights and innovative points of view. Fourth, emerging interdisciplinary themes that could drive innovation may face significant challenges, particularly in acquiring the required resources. This challenge can arise due to the prevailing disciplinary perspective, which often prioritises traditional or mainstream research themes. Consequently, these main topics can overshadow important interdisciplinary topics, even if the latter have transformative potential, as resources may be skewed toward more established research areas focusing on disciplines (e.g., Gray, 2008; Obradović, 2019). Finally, the segmented approach presents significant challenges to garner support for large-scale projects. As perceptions remain limited to specific domains, bringing together diverse stakeholders, including researchers, investors, and industry leaders, can be challenging.

Given the challenges posed by disciplinary silos in the field of BCT research, the pri-

[1] Disciplinary silos is defined as the separation of individuals or groups within a specific field or discipline caused by organisational structures, limited communication, or systemic obstacles (Jacobs, 2015, p.18). This state of isolation can impede interdisciplinary collaboration, restrict innovation, and limit knowledge and resource sharing.

mary objective of this review of the literature is to systematically analyse, synthesise, and map the existing body of blockchain research in diverse disciplines. This review aims to bridge these gaps and identify intersections between different areas of study, with the ultimate goal of promoting a more cohesive and integrated understanding of the transformative potential of BCT.

2.2.2 Limitations of Past Literature Reviews on BCT

Several reviews of the literature on BCT have been conducted, each with a different focus and approach, but many do not capture the complexity and interdisciplinary nature of BCT, which we attribute to disciplinary focus and methodological limitations to conduct an interdisciplinary review of past research. For example, there exist biases, quantity over quality, in citation-based and collaboration-based analysis that demand a new approach like topic modelling for a holistic and objective analysis.

Yli-Huumo et al. (2016) undertook a systematic review of 41 peer-reviewed articles on blockchain technology, focusing primarily on technical issues related to Bitcoin. Although the review offers valuable technical insights, it fails to explore the interdisciplinary connections between blockchain and formal research disciplines such as computer science. This narrow focus overlooks the rich interplay between the technical, social, and economic domains, a deficiency that our research, through the use of Latent Dirichlet Allocation (LDA), seeks to address by uncovering hidden thematic structures.

Risius and Spohrer (2017) analysed 69 non-technical articles on blockchain, focusing on its application, development, usage and impact. Although its analysis presents a broader perspective on BCT compared to Yli-Huumo et al. (2016), it does so mainly from the point of view of computer experts who interpret nontechnical domains. This method undermines the linkages to technical domains of practice and carries a strong bias, failing to capture the multifaceted aspects of the blockchain. Our research aims to overcome this limitation by including a variety of articles from different fields.

Hawlitschek et al. (2018) and Holub and Johnson (2018) conducted literature studies focusing on business, the sharing economy, and academic research on Bitcoin. Their analysis, while insightful, suffers from limitations, such as overlooking complex interdependencies and relying on preprinted manuscripts. These constraints hinder a comprehensive understanding of blockchain's complexities and restrict the analysis from a siloed perspective. Our research, through the use of advanced techniques such as LDA, aims to provide a more nuanced and interconnected view of the blockchain.

In addition, Dutta et al. (2020) analysed 178 articles on blockchain technology in supply chain operations, but did not have an in-depth examination of the technical features and interdisciplinary dynamics of the technology. Our research addresses this gap by applying diversity indices, such as the Herfindahl-Hirschman Index (HHI), to provide a more nuanced and comprehensive analysis of blockchain's diffusion across different sectors.

Most existing studies rely on manual literature search strategies, which are resource intensive and time consuming. In contrast, our research leverages the computationally intensive technique of topic modelling, a method that facilitates interpretation and decreases the need for subject-matter experts. Coupled with LDA and appropriate statistical summarisation techniques (e.g., HHI), our approach offers a deeper understanding of the interdisciplinary nature of blockchains, efficiently exploring their complex facets, and enhancing our ability to interpret this rapidly evolving field by analysing 35,604 research articles on BCT.

2.2.3 Overcoming Analytical Limitations with Topic Modelling

Several factors limit the analytical methods used in the current research. Two main methods in existing research are citation-based and collaboration-based analyses. Citation-based analyses have limitations such as citation biases (e.g., Okamura, 2019), misrepresentation of influence or Matthew effect (Shema, 2013), and negative citations (Worrall and Cohn, 2023). On the other hand, collaboration-based analyses have limitations, such as overemphasis on quantity rather than quality (Vieira, 2023), inadequate capture of interdisciplinarity (Maz-Machado and Jiménez-Fanjul, 2018), and potential data fragmentation (Ding, 2011). Both methods can also have a narrow scope and limited exploration of emerging trends (Hicks, 1987; Worrall and Cohn, 2023). Therefore, there is an evident need for literature review approaches that transcend these limitations. Consequently, our study uses topic modelling (LDA), which offers a probabilistic and systematic framework to identify hidden thematic structures within large-scale text datasets. Unlike traditional methods, LDA operates without relying on predetermined categories, facilitating a more objective and scalable literature analysis. The following sections will provide a detailed critique of the limitations of current analytical approaches and justify the adoption of topic modelling in our study.

2.2.3.1 Citation-based Analyses

Citation-based analyses involves a variety of analytical methods that focus on the examination of citations of scholarly publications (Osareh, 1996). These methods provide insight into the influence, interconnections, development, and evolution of a particular field (Maz-Machado and Jiménez-Fanjul, 2018). This category includes various methods such as citation (Golosovsky, 2019), cocitation analysis (Eom, 2010; Surwase et al., 2011), and bibliographic coupling (Weinberg, 1974).

Citation and cocitation analyses evaluate the impact and interconnectedness of academic publications (Goodwin and Garfield, 1980) by their citation frequency, identifying influential works and authors (Zhu et al., 2015), detect research trends (Shahid and Hahn, 2020), and reveal emerging research areas (Culnan, 1986; Shahid and Hahn, 2020). It also assesses the societal influences on research practices, promoting knowledge propagation and interdisciplinary connections.

Cocitation analysis examines the relationship between publications based on shared citations (González-Teruel et al., 2015). This technique maps the intellectual structure of a field (Culnan, 1986), uncovers themes and trends, and illuminates indirect connections between ideas, thereby proving particularly useful in understanding complex subjects with non-obvious relationships (González-Teruel et al., 2015; Hou et al., 2018). Finally, bibliographic coupling examines the similarity between reference lists of multiple publications (Most et al., 2018). The analysis of shared references determines the strength of connections based on standard intellectual foundations. This analysis can reveal collaborative opportunities, uncover research directions, and provide an additional layer of understanding, thus enhancing the insights gained from citation and cocitation analysis.

However, citation-based analyses techniques have limitations, such as citation biases, negative citations, and the inability to distinguish between pivotal and tangential references. It becomes challenging as each citation is equally weighted, obscuring the discernment of their varying contributions to intellectual content. For example, citation bias in blockchain research can lead to a limited and skewed representation of the literature (Bornmann and Daniel, 2008), which often focuses on specific technical aspects and neglects intricate connections and interdisciplinary insights (e.g., Yu and Pan, 2021). This bias can hinder the exploration of emerging trends and constrain the comprehensiveness of the research, especially when there are skewed distributions in the number of papers on a topic across disciplines. However, LDA offers a method to fill this gap. By applying LDA, researchers can uncover hidden thematic structures within blockchain literature and identify topics representing different facets of the technology. Unlike citation-based analysis, which may overlook less-cited or non-cited sources, LDA considers the entire text corpus, ensuring a more balanced representation, allowing the detection of nuanced relationships and emergent properties that characterise blockchain's complexity.

Additionally, LDA's ability to group related terms into coherent topics can facilitate the integration of interdisciplinary insights, thereby capturing the multifaceted nature of blockchain without the constraints of citation bias. Thus, by applying LDA, we can access a more comprehensive and diverse array of themes within the blockchain literature. This approach allows for the identification of under-represented and emerging topics that contribute to a more nuanced exploration of the field. While LDA does not directly interpret the complexity of blockchain technology, it aids in constructing a more informed and thorough analysis, thereby supporting a more comprehensive and multi-dimensional view of the literature.

Negative citations, which indicate limitations, inconsistencies, or flaws in previous work, can misrepresent influence, leading to a distorted view of a field. The misrepresentation of influence is particularly critical for understanding BCT, where divergent perspectives and critiques are vital in its continuous development and refinement. The limitations of negative citations include potential bias, as they can be influenced by personal biases or agendas, leading to unfair criticism or overlooking the essential aspects of previous work. Negative citations may not fully represent the literature as they may only focus on specific limitations or flaws in previous works. This limited representation can hinder a nuanced understanding of blockchain's complexity, as it may emphasise particular critiques while neglecting the broader context and interconnectedness of technology. Recognising the limitations of negative citations is essential to ensure a balanced and thorough examination of the subject, encompassing its strengths and improvement areas.

In conclusion, while citation-based analyses offers valuable insights into a field's influence and development, its limitations, such as citation biases, misinterpretations of negative citations, and the uniform weight of citations, can obscure the complete picture. Advanced techniques such as LDA present a promising alternative by revealing hidden thematic structures within the literature, considering all sources, and enabling nuanced exploration of interdisciplinary dynamics. Using LDA, researchers can develop a more comprehensive understanding of complex fields such as blockchain technology, promoting broader collaboration and deeper insights into its evolving landscape.

2.2.3.2 Collaboration-based Analyses

Collaboration-based analyses involves authorship and journal analysis that examines collaboration patterns among researchers, institutions, or countries (e.g., Gazni et al., 2012; Cheng et al., 2013). Although these methods provide insight into research collaboration patterns and their influence on productivity, they also present significant restrictions, especially when applied in the context of BCT.

The authorship analysis examines the publication trends of individual authors and evaluates their productivity and collaboration patterns within a specific field. However, its emphasis on co-authorship may overlook the intricate network of relationships between various disciplines and real-world applications that define the blockchain. Moreover, discrepancies in author names and irregularities in handling multiple authorships can cause data fragmentation, concealing vital connections, and hindering the ability to capture blockchain's interdisciplinary essence accurately.

On the other hand, journal analysis, which evaluates scholarly journals based on metrics such as citation counts, impact factors, and the h-index can help understand influence and reputation. However, this reliance on traditional metrics and prestige associated with particular publication venues can inadvertently lead to an overemphasis on well-established ideas and institutions. Consequently, this approach might overshadow less recognised, but equally important, contributions from emerging scholars or journals that explore innovative and unconventional perspectives on blockchain technology. Such biases can constrain a thorough and nuanced exploration of the complex and interdisciplinary nature of blockchain research.

Moreover, collaboration-based analyses emphasis on co-authorship patterns may neglect the intricate interactions between diverse disciplines that shape BCT. LDA has the potential to overcome these limitations by uncovering hidden thematic structures within the literature. Unlike collaboration-based analyses, which may overlook nuanced connections, LDA can expose underlying themes and trends and highlight indirect or subtle relationships between ideas or schools of thought. This approach offers a more comprehensive understanding of the complexity of blockchain, including the interaction between technology, governance, security, economics, and other dimensions. Additionally, the HHI provides a quantitative measure of diversity within a field, which can uncover shifts in focus and the convergence of interdisciplinary efforts. The application of HHI serves as a tool for identifying changes in collaboration and diversification over time, as a decrease in HHI could indicate increased diversification and collaboration between disparate domains, while an increase in HHI could reveal consolidation or narrowing of focus within specific areas.

LDA provides a more nuanced approach to understanding thematic diversity and interdisciplinarity of BCT research. Meanwhile, statistical metrics such as HHI complement this analysis by quantifying the concentration or diversity of research contributions and collaborations in the field. Although HHI does not directly unravel the complexity of blockchain technology, it offers valuable metrics for understanding the evolving landscape of research activity, indicating trends toward diversification or consolidation in blockchain scholarship. Therefore, LDA and HHI provide a multifaceted view of blockchain research, combining thematic depth with quantitative analysis of research patterns.

2.3 Research Methodology

This study uses Latent Dirichlet Allocation (LDA) (Blei et al., 2003), a probabilistic topic modelling technique, and four analytical metrics to identify themes or research topics in our corpus. We propose the IDEA framework to map and evaluate the interdisciplinary nature of BCT research. The following subsections provide a detailed explanation of data collection and our methodology.

2.3.1 Data Collection

The literature dataset was created by combining articles from six well-known academic databases, including IEEE Xplore, JSTOR, PubMed, ScienceDirect, Scopus, and Web of Science. A search was conducted using the keyword "blockchain." The decision to use the term "blockchain" or its variations as "block chain" or "block-chain" can be attributed to two main factors. First, our research is aimed to understand blockchain as a whole rather than a niche technology for a specific domain (e.g., Vacca et al., 2021), so researchers are highly likely to use the keyword "blockchain" or its combinations in their articles, to represent BCT as the underlying technology. Second, the use of search keywords such as "bitcoin", "smart contract", or "digital ledger technology" may introduce noise into the dataset due to their broader or sometimes overlapping meanings. Our approach aligns with several influential review papers that also utilised "blockchain" as their primary keyword for identifying relevant publications. For example, Casino et al. (2019) employed a similar approach, demonstrating its effectiveness in capturing the core body of blockchain research.

To further validate the robustness of our keyword "blockchain", we conducted a sensitivity analysis, incorporating additional keywords (e.g., "cryptocurrency", "distributed ledger technology"). The analysis revealed that the overall patterns and trends remained consistent, suggesting that the choice of a single keyword did not significantly bias our results.

Our database search found 109,369 publications from 01 January 2013 to 31 December 2022. After excluding 73,765 publications that met specific criteria (such as duplication, absence of an abstract, non-English text, publication outside the specified range, and entries from sources other than conferences or journal publications), a final dataset of 35,604 entries was retained for further analysis.

The main variables in the final dataset are listed in Table 2.1. Additional information, such as the Research Discipline and the Discipline Cluster (refer to Section A.1 to see a preview of data table), was added to the final dataset using text classification and hierarchical clustering techniques. This process resulted in all publications distributed in 109 research disciplines that were further grouped into 15 high-level discipline clusters for analysis.[2]

A new dataset instance, called the "corpus", was created specifically for text processing. The corpus consists of the abstract and title of the publication. Each abstract was combined with its corresponding title to provide context (see Table 2.2 for an illustration) for the content. Each row in the corpus represents a "document".

Figure 2.1 illustrates the overall distribution of 35,604 documents from 2013 to 2022.

[2]We used these 15 discipline clusters for analysis and refer to them interchangeably as discipline or discipline clusters throughout this study.

Table 2.1: Main variables

Category	Variables
Publication	Year, Title, Abstract, Author, Research Discipline[1] Discipline Cluster[2]

[1] Because our corpus consists of six major databases, of which Scopus and Web of Science are the largest, each document is assigned a research discipline at the source. Each document also contains keywords that represent the assigned discipline. We used these keywords to create a dictionary using a two-fold strategy: clustering and classification. First, we trained a classifier using a keyword dictionary which yielded an $F-score$ of 92%, demonstrating the precision of successful classification. Second, we use our classifier to determine the discipline of each publication on IEEE Xplore, JSTOR, PubMed, and ScienceDirect.
[2] The discipline clusters were determined using hierarchical clustering.

Table 2.2: Illustration of a document's metadata for text processing

	Title	Abstract	Concatenated Title and Abstract
Document	This is the title of the publication.	This is the abstract of the publication.	This is the title of the publication: This is the abstract of the publication.

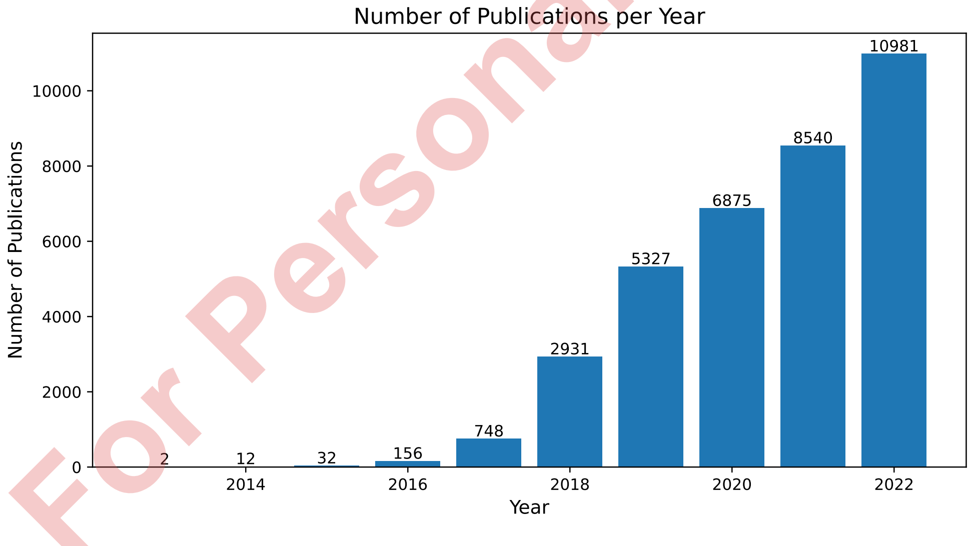

Figure 2.1: Distribution of research publications from 2013 to 2022

2.3.2 Methodology

The research methodology used in this study consists of two main components, as shown in Figure 2.2 and Figure 2.3. First, we extracted LDA topics from the corpus. Second, we analysed the extracted LDA topics using four analysis metrics.

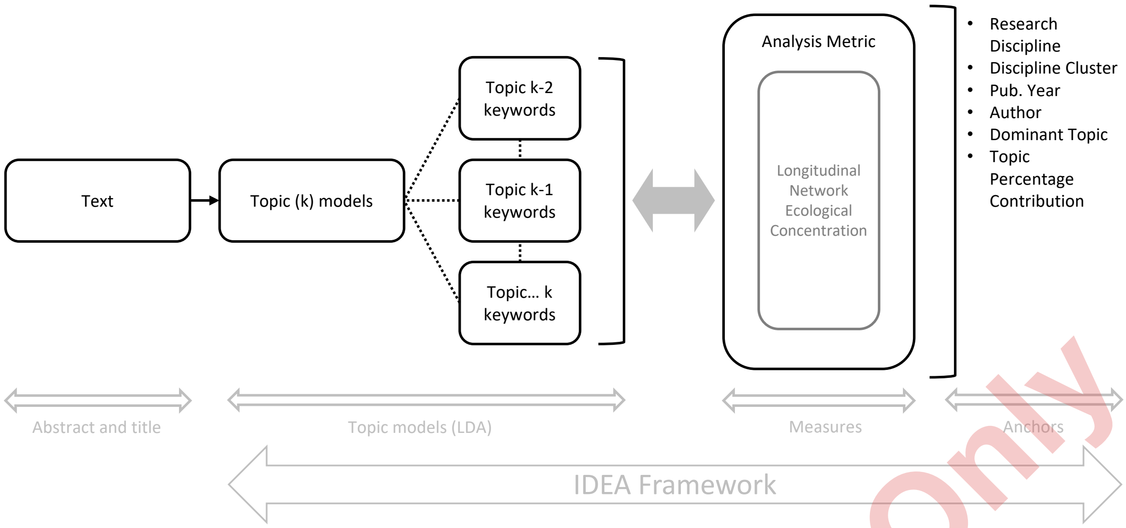

Figure 2.2: Overview of research method

Figure 2.2 shows overview of the research model. It has three main components: text, topic models, and measures. First, the text was prepared by including the abstract and title of the publication. Second, topic models were generated using LDA. Third, analysis metrics were applied to topic models to draw inferences. The analysis metrics also form the basis of our innovative IDEA framework. Figure 2.3 illustrates a three-stage process model that describes the sequential steps used to determine LDA topics and enrich the corpus data.

2.3.2.1 Dataset Structure and LDA Output Variables

The dataset used in our study consists of 17 attributes (refer to Section A.1 to see a preview of data table), including document number, year, title, abstract, text, discipline, discipline cluster, authors, topic contributions, and dominant topics. The text in the abstract column was converted into stem words in the “Text” column. The year and author columns in this study were obtained from the original data in the databases. In addition, research disciplines were assigned to each document in the Scopus and Web of Science databases. To assign a research discipline to IEEE Xplore, JSTOR, PubMed, and ScienceDirect documents, we used a two-step approach. First, we created a dictionary of keywords from documents in the Scopus and Web of Science databases, and trained a classifier, that yielded a $F-score$ of 92%. This outcome provides evidence of reliable classification. Second, we used the

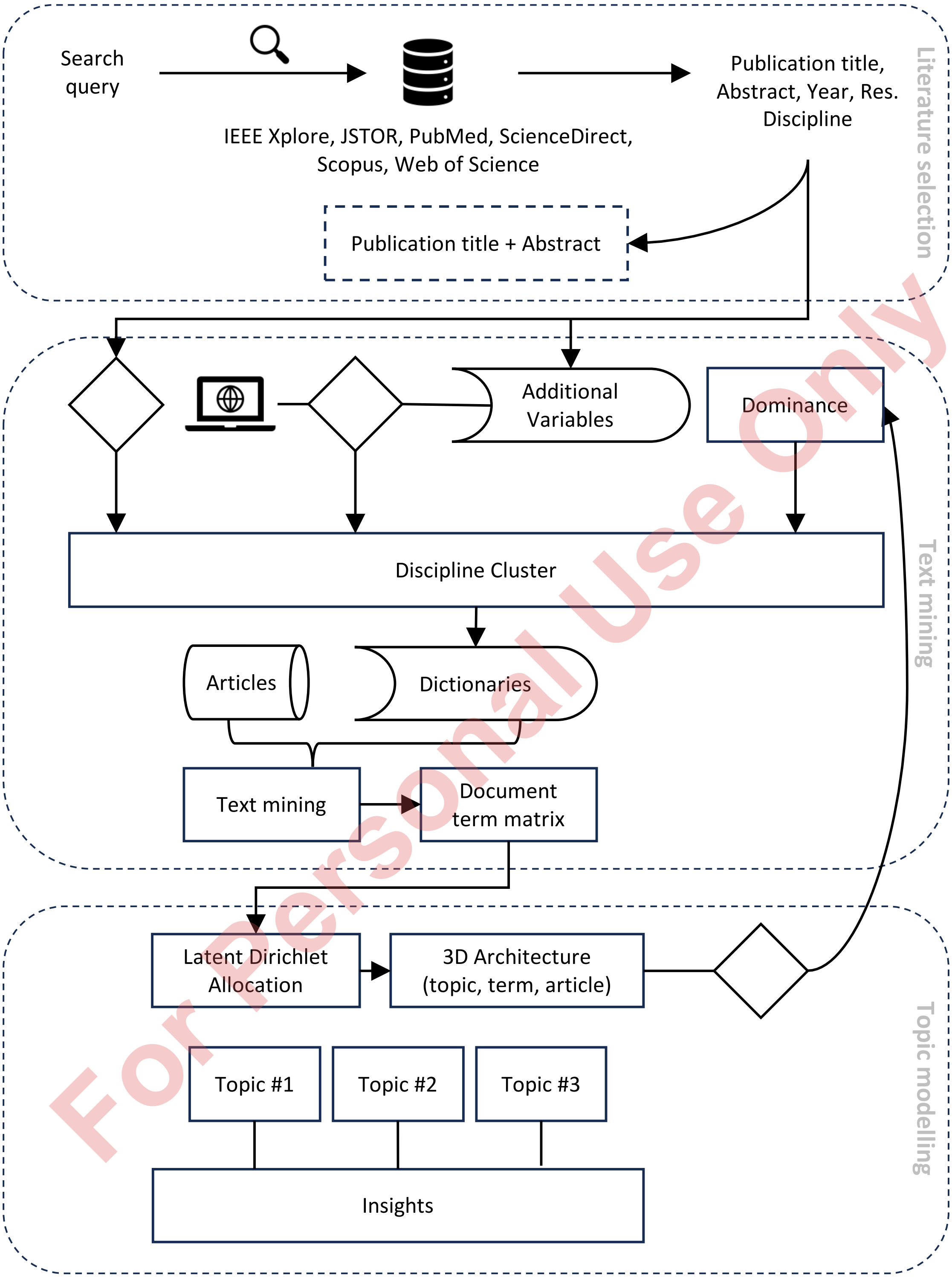

Figure 2.3: A process model for determining topic models and supplementing the dataset with new data

Notes: The figure illustrates a three-stage model for generating LDA topics and incorporating additional data.

classifier to interpret the research disciplines of documents in IEEE Xplore, JSTOR, PubMed, and ScienceDirect databases. We then used hierarchical clustering (refer to Section A.2 and Section A.3) to categorise disciplines into higher-level discipline clusters.

The LDA algorithm analyses the composition of each document to determine the contributions of different topics and identify the dominant topic. It represents each topic as a distribution of words. The iterative process used in this study calculates the distribution of document-topic (θ) and topic-word (β) by observing specific words. The dominant topic ($T = \arg\max_k \theta_{d,k}$) was determined, where k represents potential topics. In this context, $\theta_{d,k}$ represents the proportion of topic k in document d, and $\arg\max_k$ denotes the argument that maximises the function. The predominant subject in document d is denoted by k, representing the highest proportion $\theta_{d,k}$.

2.3.2.2 The IDEA Framework for Assessment of the Research Landscape

To analyse the research landscape of BCT, we propose and use the IDEA framework described in Table 2.3. The acronym "IDEA" stands for "Integrated Dynamics, Evolution, and Analysis". It encompasses ten aspects examined using four analysis categories: longitudinal, network, ecological, and concentration. The aspects we aim to investigate in our research represent the overarching dimensions. The analysis categories were chosen based on their proven effectiveness in literature reviews and their direct relevance to the research questions of our study. We categorised all aspects into four broader categories: temporal and evolutionary analysis, interdisciplinary dynamics, research complexity and behaviour, and research focus and concentration. These more general categories offer a distinct perspective on BCT research.

The analysis category and subcategory specify the general type of analysis and the specific techniques used to explore each aspect, providing the methodological context for our study. The sub-question column serves as a practical guide, breaking down our main research questions into specific, measurable queries. These subquestions guided our data analysis and provided a framework for discussing the interdisciplinary

Table 2.3: IDEA framework

S.No.	Aspect to Explore	Analysis Category	Analysis Sub-category	Sub-Questions	Implications for Academic Inquiry
			Temporal and Evolutionary Analysis		
1.	Research Focus Evolution	Longitudinal Analysis	Time-Series Analysis	How have dominant research topics evolved?	Shifting or stable themes over time.
2.	Dynamic Research Interests	Longitudinal Analysis	Time-Series Analysis	Topics gaining or losing momentum over time?	Reveals topics of increasing or decreasing interest.
			Interdisciplinary Dynamics		
3.	Synergistic Knowledge Areas	Network Analysis	Association Analysis	Which entities are frequently discussed together?	Reveals connection or separation between study areas.
4.	Academic Convergence or Divergence	Ecological Metrics	Richness (Shannon Diversity) and Evenness (Pielou's Index)	Are topics confined to specific disciplines?	Indicates if research is integrated or compartmentalised[1]
5.	Thematic Overlaps	Ecological Metrics	Overlap Analysis	Are there common research themes that span multiple disciplines?	Identifies unifying themes that could be central to the interdisciplinary nature of blockchain.
			Research Complexity and Behaviour		
6.	Evolution of Research Complexity	Longitudinal Analysis	Trend Analysis	How has the diversity of topics evolved and unique disciplines added over time?	Shows if the field is diversifying or narrowing and implications for this for academic research.
7.	Research Landscape Breadth	Ecological Metrics	Diversity Analysis	What is the range of topics covered?	Reveals the breadth and depth of the research landscape by showing the variety of topics covered[2]
8.	Scholarly Versatility	Ecological Metrics	Versatility Score (Entropy-based Analysis)	How do the contributions of first authors evolve across disciplines over time and what is the versatility of all authors across topics within these disciplines?	Indicates versatility of the author's expertise across topics.
9.	Collaborative Trends	Ecological Metrics	Diversity Analysis	How does the nature of collaboration vary by discipline and topic?	Insights into the structure and evolution of the academic community.
			Research Focus and Concentration		
10.	Research Focus Concentration	Concentration Metrics	Herfindahl - Hirschman Index (HHI) Analysis	How concentrated is research within specific topics across disciplines?	Indicates if research is highly focused or spread across multiple themes.

[1] The focus is on the distribution of topics across disciplines, not the topics themselves.
[2] The focus is on the variety of topics themselves.

nature of blockchain. The implications for academic inquiry column provides a broader understanding of the academic significance of each aspect explored, placing our research findings within the larger academic discourse.

Building upon the foundational elements described in our methodology, we discuss each aspect in detail in the following subsections.

2.3.2.2.1 Temporal and Evolutionary Analysis: This broad category has two aspects: Research Focus Evolution and Dynamic Research Interests.

Research Focus Evolution: This aspect is intended to provide a temporal perspective on the landscape of blockchain research, using longitudinal analysis, a method particularly popular in the social sciences to track changes over time (Elder et al., 2003). In our study, this aspect examined how dominant research topics have evolved.

The rationale behind using longitudinal analysis lies in its ability to capture temporal dynamics, which is crucial to understanding the evolution of blockchain research. This method can reveal whether specific research topics have remained stable, demonstrating a sustained focus, or if new topics have emerged, indicating changes in research priorities.

Therefore, Research Focus Evolution is a crucial aspect of our framework, providing a time-critical lens to understand the dynamics and assess the evolution of the interdisciplinarity of blockchain research.

We calculated Research Focus Evolution using three measures: Frequency Count $f_{t,y}$, Yearly Growth Rate $g_{t,y}$, and Proportional Representation $p_{t,y}$. The Frequency Count $f_{t,y}$ tracks the prevalence of each dominant topic per year, the Yearly Growth Rate $g_{t,y}$ measures the rate of change in the prevalence of the topic from one year to the next, providing information on emerging or declining topics, and the Proportional Representation $p_{t,y}$ offers a perspective on the relative importance of each topic within a given year.

Mathematically,

$$f_{t,y} = \text{Number of occurrences of dominant topic } t \text{ in year } y \tag{2.1}$$

where, $f_{t,y}$ is the frequency count of a dominant topic in a year y.

$$g_{t,y} = \frac{f_{t,y+1} - f_{t,y}}{f_{t,y}} \tag{2.2}$$

where, $g_{t,y}$ represents the growth rate of the dominant topic t from year y to $y + 1$.

$$p_{t,y} = \frac{f_{t,y}}{\sum_{t'} f_{t',y}} \tag{2.3}$$

where, $p_{t,y}$ calculates the proportion of dominant topic t in year y, with t' covering all topics in that year.

Dynamic Research Interests: This aspect uses longitudinal analysis, particularly Time-Series analysis to shed light on evolving patterns in BCT research. Time series can serve as a valuable tool for tracking the history of variables over different periods. In the context of our research, a time series can help to understand the rate at which certain topics gain or lose academic attention.

In contrast to Research Focus Evolution which uses dominant topics, Dynamic Research Interests uses topic contributions within individual papers to ascertain the gain or loss of momentum over time. This approach not only enables the identification of current hot topics but also facilitates the observation of historical shifts in topic popularity. Furthermore, it allows for differentiation between topics gaining newfound significance and foundational topics experiencing a decline in prominence. Such nuanced insights can inform future research endeavours by highlighting areas ripe for exploration and those potentially approaching saturation.

Our dataset consists of columns for each topic's percentage contribution to a document for different years. We denote these by $Topic1_{P}erc_{C}ontrib$ to $Topic8_{P}erc_{C}ontrib$, respectively, and by a year column Y.

For each given year y in Y, we calculated the average contribution of each topic T_i in all records for that year. The formula for the average contribution of topic T_i in year y is:

$$\text{Average Contribution of } T_i \text{ in year } y = \frac{\sum_{j=1}^{n_y} T_{i,j}}{n_y} \tag{2.4}$$

where, $T_{i,j}$ is the contribution of topic T_i in the j^{th} record (document or row) of year y, and n_y is the number of records in the dataset for year y.

2.3.2.2.2 Interdisciplinary Dynamics: This broad category has three aspects: Synergistic Knowledge Areas, Academic Convergence/Divergence, and Thematic Overlaps.

Synergistic Knowledge Areas: This aspect examines the interconnections between topics or keywords. Using Network Analysis, a method commonly used in bibliometric studies to map relationships between different nodes (Newman, 2004), our objective was to determine whether blockchain research is compartmentalised within specific research areas or whether there is cross-pollination of ideas between research themes. We determined synergistic knowledge areas at two different levels: within topic keywords, and between topic keywords.

To determine the synergy within and between topic keywords, we used association analysis (co-occurrence), which refers to the frequency with which two keywords are associated in research articles. Such co-occurrence of keywords can serve as an indicator of interdisciplinary synergy. For example, a keyword pair "network" and "data", having a co-occurrence scale ranging between 0 (low) and 1(high), a value close to 1, will indicate a high co-occurrence compared to a keyword pair "network" and "medical" with a co-occurrence value close to 0.

Academic Convergence/Divergence: In contrast to Synergistic Knowledge Areas, which uses the co-occurrence of keywords to examine synergy, Academic Convergence/Divergence examines whether topics are limited to particular disciplines by using Richness and Evenness. Richness refers to the diversity of disciplines that contribute to each dominant topic and serves as an indicator of academic convergence. Therefore, a high value of richness means that a topic is not confined to a specific discipline.

Evenness, as measured by Pielou's Evenness Index, indicates the distribution of research contributions across different disciplines within each dominant topic. The concept of Evenness is based on its ability to quantify consistency in the distribution of contributions across various disciplines. High evenness values indicate a balanced and equitable contribution from various disciplines, suggesting harmonious integration and academic convergence. Lower values of Evenness indicate a skewed distribution where one or a few disciplines dominate, suggesting academic divergence. This measure provides insight into whether a topic's interdisciplinary nature is collaborative or dominated by specific fields.

We calculated Richness as the "count of the number of different disciplines contributing to each dominant topic." This measure is a preliminary indicator of interdisciplinary involvement, suggesting that a higher count indicates broader academic interest across various fields. Mathematically:

$$\text{Richness (Topic } i) = \text{Count of Unique Discipline contributing to Topic } i \tag{2.5}$$

We used Pielou's Evenness Index to calculate Evenness:

$$\text{Evenness (Topic } i) = \frac{\text{Shannon Diversity Index (Topic } i)}{\ln(\text{Richness (Topic } i))} \tag{2.6}$$

where, the Shannon Diversity Index is calculated as:

$$\text{Shannon Diversity Index (Topic } i) = -\sum_{j=1}^{n} p_{ij} \ln(p_{ij})$$

The value p_{ij} shows the proportion of publications from Discipline j contributing to Topic i, and n is the number of unique disciplines contributing to Topic i.

In existing research, Richness and Evenness have been used by Porter and Rafols (2009), who have demonstrated its usefulness in interdisciplinary studies. They showed that this metric can indicate whether a research discipline has a high degree of academic convergence, where many different disciplines contribute to common topics, or whether there is academic divergence, where each discipline focuses on a narrow set of topics. Therefore, this metric is useful for understanding the level of academic integration or compartmentalisation.

Thematic Overlaps: This aspect highlights the interdisciplinary nature of BCT research by identifying how dominant topics are shared across different disciplines. We used this aspect to understand the extent to which certain topics were common in different research disciplines. We were able to quantify and visualise the extent to which different dominant topics were prevalent across disciplines. Our approach draws inspiration from Rafols and Meyer (2010), who used overlap analysis to identify common themes that unify disciplines. We intend to identify these unifying themes in the BCT research landscape, thereby shedding light on the field's intellectual realm and providing valuable insights for disparate audiences.

The three aspects of Interdisciplinary Dynamics may seem similar at first glance but may serve a different purpose. To further elaborate on the differences between the three, we offer the following explanation for the example illustrated in Figure 2.4.

Suppose that we have three research disciplines: Computer Science, Economics, and Law, and three research topics: Blockchain Security, Cryptocurrency Regulation, and Smart Contracts.

In the Synergistic Knowledge Areas aspect, research papers that discuss "Blockchain Security" may also discuss "Smart Contracts", within the discipline of "Computer Science". This indicates synergy between these two topics, showing that understanding one topic could be beneficial to understanding the other topic.

In the Academic Convergence and Divergence aspect, "Distributed Systems" is discussed in the "Computer Science" discipline, "Cryptocurrency Regulation" is discussed in the "Law" discipline, and "Smart Contracts" is discussed in the three research disciplines. In this setting, "Smart Contracts" are a converging topic between

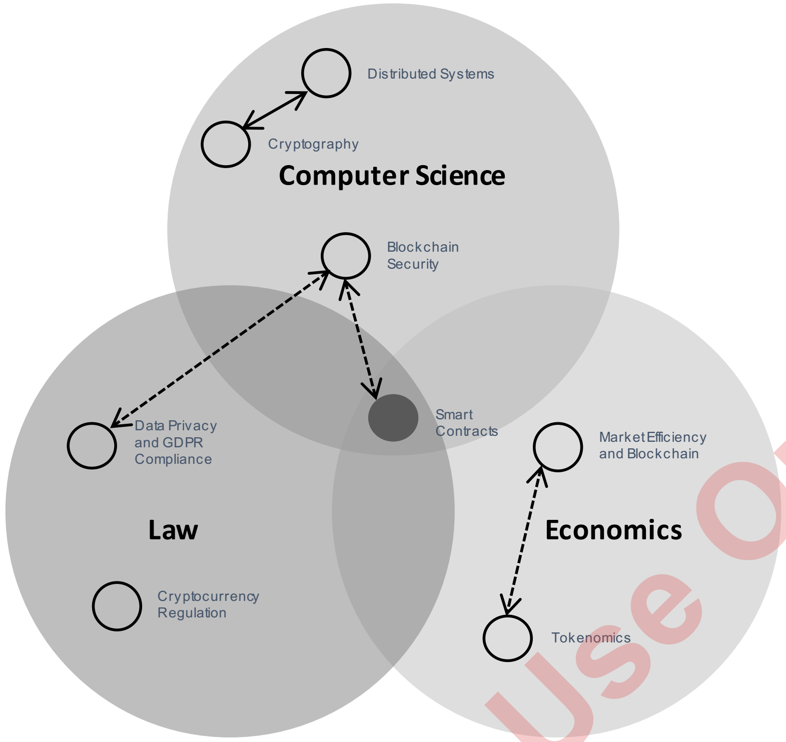

Figure 2.4: Venn diagram illustrating the interdisciplinary nature of BCT research topics across Computer Science, Law, and Economics.

Notes: The Venn diagram represents the distribution of blockchain topics in "Computer Science", "Law", and "Economics". The colour of the topic circles within the disciplinary circles represents affiliation with the discipline. The solid coloured arrows represent strong affiliation within the topics, and the dotted arrows represent the co-occurrence of topics between disciplines. The topic area at the interface of disciplinary circles represents a common topic across all disciplines.

disciplines, whereas "Distributed Systems" and "Cryptocurrency Regulation" are divergent topics due to their affiliation only with specific disciplines.

Finally, in the aspect of Thematic Overlaps, "Smart Contracts" are discussed in the three research disciplines: "Computer Science", "Economics", and "Law". This may indicate that "Smart Contracts" are a unifying topic that spans multiple disciplines, making them an ideal example of thematic overlap.

2.3.2.2.3 Research Complexity and Behaviour: This broad category has four aspects: Evolution of Research Complexity, Research Landscape Breadth, Scholarly Versatility, and Collaborative Trends.

Evolution of Research Complexity: In this aspect, we used longitudinal analysis to examine how the complexity of blockchain research has evolved. Studies in science and technology use a similar method to show whether a field is diversifying or narrowing (e.g., Rafols et al., 2010). Research complexity typically involves several disciplines or a range of topics covered in research.

The addition of unique disciplines into blockchain's research landscape adds a new dimension or layer of complexity and can provide a macro-level view of the evolution of complexity. In contrast, topic contributions in each paper provide a micro-level understanding of complexity by revealing whether the research is delving deeply into specific themes or spanning multiple topics within a paper.

To capture a multidimensional view of the evolution of research complexity, we employed both types of analysis. We assess complexity from the perspective of topic contributions in each paper and how unique disciplines are added to the blockchain research landscape over time. The main goal is to understand whether the field of blockchain research is diversifying, indicating an expansion of topic contributions and sub-disciplines, or narrowing, which indicates a focus on limited topics or themes.

Research Landscape Breadth: In this aspect, our aim was to understand the breadth of the blockchain research landscape and we used diversity analysis, which is often used in ecological studies and serves as a model to indicate the breadth and depth of the research landscape (e.g., Jost, 2006).

Diversity refers to the variety of topics explored in different disciplines. Therefore, our methodology involved counting the number of documents grouped by dominant topics and disciplines. This could serve as an indicator of Research Landscape Breadth: a high count of documents spanning multiple disciplines and topics suggests a broad research landscape, while a lower number might imply a more confined or specialised research scope.

Scholarly Versatility: In this aspect, we employ versatility analysis, inspired by biodiversity indices in ecology (e.g., Magurran, 2021), to assess the versatility of researchers in blockchain research. Versatility is conceptualised in two different dimensions: versatility across disciplines and versatility across topics.

Versatility across disciplines focuses on authors who contribute to unique disciplines as first authors (also known as polymaths), while versatility across topics focuses on those authors who contribute to publications of different topics. Versatility can signify the breadth of expertise and the ability to contribute to different aspects of blockchain. We choose the first author of a publication to measure Scholarly Versatility, because it is a common practice in academia, since the first authors typically signify expertise and significant relevance to the scientific field.

To determine the authors for the two dimensions, we extract all the first authors of publications and then associate them with the set of disciplines and topics to which they have contributed as the first author. Next, we counted the number of first authors who contributed to publications in more than one discipline or topic, thus quantifying their "versatility across disciplines."

We aggregated topic contribution $\mathrm{Agg(Topic)}_i$ for topic i for a specific author in a given discipline as follows:

$$\mathrm{Agg(Topic)}_{i,a} = \sum_{j=1}^{n_a} \mathrm{Topic}_{i,j,a} \tag{2.7}$$

where, $\mathrm{Agg(Topic)}_{i,a}$ represents the aggregated contribution to topic i by author a, $\mathrm{Topic}_{i,j,a}$ is the contribution of author a to topic i in paper j, and n_a is the total number of documents to which author a has contributed as first author.

To normalise the topics' contributions $\mathrm{Norm(Topic)}_i$, we use the following equation.

$$\mathrm{Norm(Topic)}_{i,a} = \frac{\mathrm{Agg(Topic)}_{i,a}}{\sum_{i=1}^{m} \mathrm{Agg(Topic)}_{i,a}} \tag{2.8}$$

where, m is the total number of topics, the denominator sums the aggregated topic contributions of the author a in all topics to normalise the contributions.

The Versatility Score V_a for an author a was calculated using Shannon's entropy:

$$V_a = -\sum_{i=1}^{m} \left[\text{Norm(Topic)}_{i,a} \times \log_2(\text{Norm(Topic)}_{i,a}) \right] \tag{2.9}$$

Collaborative Trends: This aspect uses diversity analysis to examine the diversity of disciplinary background of authors who co-author or collaborate in topics. Therefore, collaboration is explored in the context of interdisciplinary contributions. We used Shannon Diversity Index to assess the diversity of the authors.

To calculate Shannon Diversity Index, we used a five-step process. First, extract the first author from each document. Second, calculate the number of first authors in each discipline ($n_{d,t}$). Here, d indicates the discipline within the topic t. Third, calculate the total number of first authors in topic t and denote it as N_t. Fourth, calculate the proportion of first authors for each discipline d in topic t. The proportion $p_{d,t}$ is given by $\frac{n_{d,t}}{N_t}$. Finally, Shannon Diversity Index is calculated as follows.

$$SDI(t) = -\sum_{d \in D_t} p_{d,t} \cdot \ln(p_{d,t}) \tag{2.10}$$

where, D_t represents the set of all disciplines in the topic t. The summation runs through all these disciplines, multiplying the proportion of each discipline $p_{d,t}$ by the natural logarithm of itself and summing these products.

2.3.2.2.4 Research Focus and Concentration: This broad category has one aspect: Research Focus Concentration.

Research Focus Concentration: This aspect uses concentration metrics, specifically the HHI (Hirschman, 1964), to determine the concentration of research within specific topics or disciplines. Each dominant topic is analogous to an "industry", and research disciplines are considered "firms" within these industries. The number of research papers of a particular discipline on a given topic in a specific year is treated as the "market share" of that "firm" in the "industry".

Since the HHI is sensitive to the distribution of papers across disciplines, an even distribution (papers spread across multiple disciplines) results in a lower HHI, indicating high interdisciplinarity. However, a higher HHI suggests that most of the articles come from one or a few disciplines, indicating low interdisciplinarity.

To calculate the HHI for each dominant topic ("industry") in disciplines, we counted the number of documents belonging to each discipline ("firm") for each dominant topic. These counts were then normalised with the total number of documents for each topic to derive the proportion of documents per topic for each discipline. The HHI for each dominant topic was then calculated by squaring these proportions and summing them across all disciplines. This process yields a measure that reflects the level of document distribution between disciplines for each dominant topic and, by extension, the degree of interdisciplinarity within these topics.

Mathematically, the HHI for a given topic in a given year is:

$$\text{HHItopic, year} = \sum i = 1^N (p_{i,\text{topic, year}})^2 \tag{2.11}$$

where, N is the total number of disciplines, $p_{i,\text{topic, year}}$ is the proportion of papers from discipline i for a specific topic in a given year.

p_i is calculated as follows:

$$p_{i,\text{topic, year}} = \frac{\text{No. of papers in discipline } i \text{ for the given topic in a given year}}{\text{Total number of papers for the given topic in the given year}}$$

where, the numerator is the number of papers in discipline i, the denominator is the total number of papers, for that topic and year, respectively.

Once Equation 2.11 is calculated, the average HHI for a year is simply the mean of all HHIs for that year.

$$\text{Average HHI}_{\text{year}} = \frac{1}{T} \sum_{j=1}^{T} \text{HHI}_{\text{topic}_j,\text{year}} \tag{2.12}$$

where, T denotes the total number of topics, $\text{HHI}_{\text{topic}_j,\text{year}}$ represents the HHI calculated for each topic j in a given year.

We used Average HHI_{year} as a measure to quantify the level of interdisciplinarity within a given year by treating lower levels of concentration as topic diversity in disciplines. Therefore, a lower HHI indicates a more even distribution of documents across disciplines for a given topic and a higher level of interdisciplinarity. On the contrary, a higher HHI indicates a skewed distribution, implying that the topic is primarily associated with a particular discipline.

2.3.2.3 LDA Algorithm Parameters

Several parameters of the LDA algorithm must be set in the corpus before execution. Automatic selection of words is carried out by: 1) the Bag of Words (*BoW*) model and 2) the Term Frequency-Inverse Document Frequency (*TF-IDF*) model. The *BoW* model counts the frequency with which a word appears in a document and pairs them in the final output, while the *TF-IDF* model weights the relevance of a word in a document in the first part and assigns less weight to very frequent words (e.g. "and", "the") in the second part (i.e., Inverse Document Frequency). The *BoW* model is available as a preset parameter of the *TF-IDF* model and improves the overall quality of the resulting topics. Therefore, in this study, *TF-IDF* is used for the LDA model.

To fine-tune the LDA model, we evaluated several parameters. A key consideration was the treatment of the frequent keywords in the corpus. Unlike the study by Risius and Spohrer (2017), which had a homogeneous distribution of keywords, our study aimed to ensure a balanced representation of less frequent keywords. To achieve this, we tested various exclusion criteria, ranging from 20% to 70% for frequent terms such as "blockchain", "bitcoin", and "cryptocurrency". After the empirical evaluation, a 50% exclusion criterion was found to yield topics that were coherent and balanced in terms of keyword frequency.

In addition, we also adjusted computational parameters (i.e., "chunksize" and "iterations"). We conducted several tests by varying the chunk size between 500 and 5,000 and iterations between 500 and 2,000. These parameters can influence the computational efficiency and quality of topics; therefore, a smaller chunk size or fewer iterations can lead to incomplete loading of documents into the memory or insufficient convergence of the algorithm, respectively. After conducting multiple trials, we found that the chunk sizes of 2,000 and 1,000 iterations were optimal. These settings not only allowed for full loading of documents into memory, but also ensured that the algorithm had sufficient iterations to converge, producing the most coherent and interpretable set of topics.

To determine the relevance of a publication to a research discipline, the research areas assigned to each publication on the Web of Science and Scopus were used as reference points to create a bag of words consisting of probabilities that determine the relationship of a word to a specific research discipline. A classifier was then trained on this dictionary to determine the research disciplines of all publications for which no discipline information was available. An $F-score$ of 92% indicates the threshold of the classifier to correctly specify the discipline of the unassigned publications.

2.3.2.4 Selection of the Number of LDA Topics

The choice of the number of LDA topics, K, is also somewhat controversial. In almost all cases, the number of topics is unknown a priori. Previous researchers have determined them at random (e.g., 20, 40, and 80 are popular choices) or created multiple models with different values for K and selected K based on a visual inspection of the results. A well-known procedure based on a coherence-based method is often used to determine the optimal value of K for the number of topics. In this approach, many different LDA models are created with different values for the number of topics, and the one that provides the highest coherence value is selected. Choosing a K value that marks the end of rapid growth in topic coherence yields a meaningful and interpretable set of topics.

To determine the optimal value of K, we created many LDA models with different values of K of up to 30. The highest retraction of the coherence score occurred at $K = 8$, after which the coherence score was no longer significant; therefore, the graph shows a maximum of 30 topics. Figure 2.5 shows the coherence score on

$y - axis$ and the number of topics on the $x - axis$. The value of K in the range of 8 to 15 on $x - axis$ shows greater fluctuations in the higher coherence score than the lower values. However, at 8, the coherence score was higher than subsequent scores for each topic. Therefore, given the highest coherence score at $K = 8$, the eight-topic LDA model was selected for our analysis.

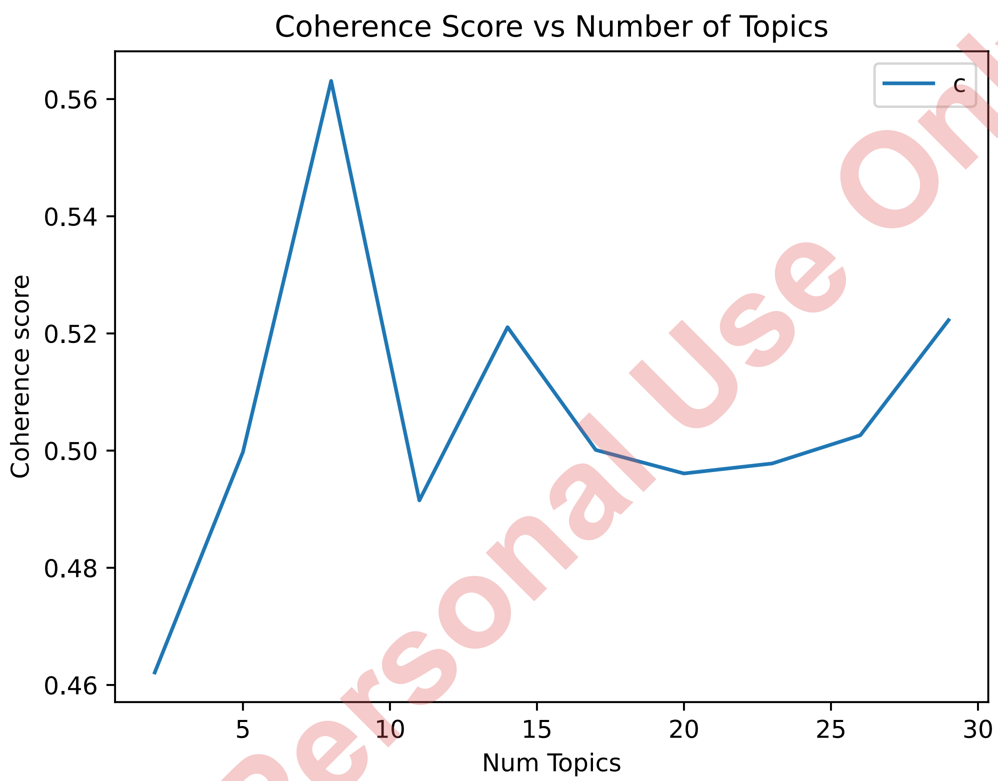

Figure 2.5: Selection of "K" topics

Notes: In this two-dimensional figure, the x-axis represents the number of topics and the y-axis represents the coherence value. The coherence value for the eight topics was optimal.

In addition to coherence-based selection of LDA topics, we also explored other approaches, such as perplexity-based selection, visual inspection, and random values K. In perplexity-based selection, the measure indicates how well the model predicts the alignment between the probability distribution and the actual distribution of words. Generally, low probability values indicate a better model as it is good at predicting the sample. However, perplexity is an evaluation metric that may not necessarily correlate with human interpretability of the topics generated by the model. Therefore, we opt for coherence scores, which are better at capturing the quality of topics for our study that revolves around thematic interdependence to assess interdisciplinarity. We visually inspect the models for topic quality; however, because this method is highly subjective, it is not suitable for our study because it is less reliable. In the random K value method, we tried 20, 40, and 80 topics; however, this randomisation did not return coherence scores that were comparable to the model with $K = 8$. This suggests that these models may be less effective in generating topics that are semantically coherent and interpretable.

2.4 Research Findings

In this section, the main variables (Table 2.1) and the topic models are analysed to answer the study's research questions. Although the analytical approach is exploratory and based on a 'trial-and-error' methodology, because interdisciplinary research lacks a specific framework for discussing results, we employ the IDEA framework given in Table 2.3 in conjunction with LDA results to discuss our findings.

To answer RQ1 (*What is the current state of BCT research in various research disciplines?*), we analysed LDA topics and keyword themes to understand the state of BCT research. These topics and keywords serve as an analytical lens through which the contours of the BCT can be mapped, indicating not only the prevailing themes but also the trends and gaps. Furthermore, using our analytical framework IDEA and LDA topics as the foundational layer, we discuss the results of RQ2 (*To what extent does the BCT research reflect interdisciplinarity?*). In particular, we evaluated the distribution of topics across research disciplines. Our analytical framework and LDA topics serve as a dual mechanism, continuously updating our understanding of both the interdisciplinary nature and the current state of BCT research.

The LDA process distributes a document onto a preset number of topics and is known as a document-topic distribution. The distribution is a percentage that shows the influence of each topic in a document. The topic with the highest percentage is called the dominant topic. Then each topic is distributed through the ten most relevant keywords. Each keyword has a topic-keyword probability (represented as β)[3], which indicates relevance to the topic, and the group of keywords with high probabilities can represent the importance of the topic, in the case of this research, the research theme. The interrelevance of topic-keyword probabilities can signal a potential title for a topic. For example, the terms "healthcare", "health", "medicine", and "patient" are interrelated and can be grouped under the topic name "healthcare". We refer to the main variables of Table 2.1 together with the LDA results to discuss further our findings in subsequent sections.

2.4.1 Understanding the Current State of Blockchain Research: Topic Models

The application of LDA to 35,604 documents resulted in eight LDA topics (as shown in Figure 2.6), based on a group of the top ten[4] keywords for each topic, respectively.

The document-topic with the highest percentage indicates its dominance. The maximum and minimum percentages for dominant topics for 35,604 documents in this study were 0.92 and 0.23, respectively, while the average value was 0.66. The black areas on the heat map in Figure 2.7 show that the order of dominance for the topics is Topic 7, 5, 4, 1, 2, 8, 3, and 6. Table 2.4 shows the distribution of the original 35,604 documents.

Each LDA topic represents a theme and topic-keywords further explain the contours

[3] β is also known as the **tolerance** value, which varies between 0 and 1. The β shows the strength of the relevance of a keyword to the topic. Therefore, a higher value of β indicates a higher relevance of a keyword to its topic.

[4] The primary goal of LDA is unsupervised text summarisation. There is no concrete method to determine the number of main keywords for each topic. Most research on topic models tends to use the top ten keywords. If the number of keywords exceeds 10, there is an increased likelihood that the purpose of LDA to succinctly summarise the text will be defeated to some extent. For example, the tolerance of $\beta > 0.01$ is far too low to show which keywords belong to each topic. This is because the main purpose of LDA is to group keywords such that topic keywords have a high probability of relevance to that topic. If a low threshold is chosen, many keywords will appear in each topic, which in turn defeats the purpose of concise text summaries with relevant keywords. To extract the most likely words, we chose a threshold of $\beta > 0.10$.

Figure 2.6: LDA word cloud

Notes: The word cloud is generated with LDA. The weight (or probability) of the words cloud represented with the size of each word shows its significance in the topic. A larger size of the word indicates a higher weight (or probability) in that topic.

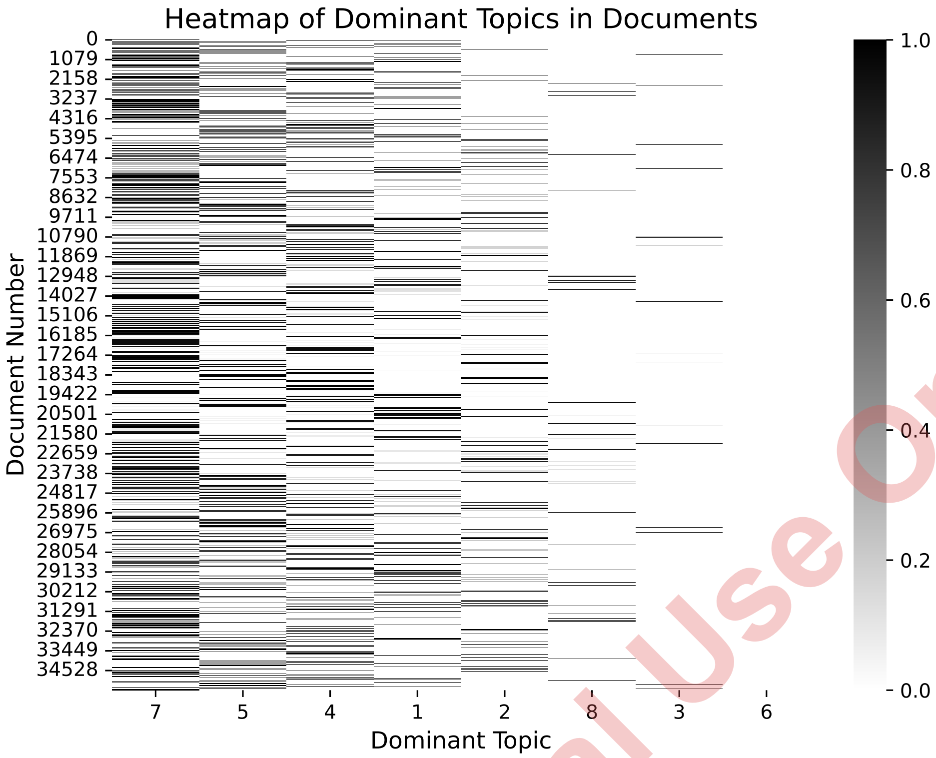

Figure 2.7: Heat maps of topics 1 to 8

Notes: In this two-dimensional diagram (a subset derived from the main diagram, displayed in the upper right corner), the y-axis represents documents, while the x-axis represents topics 1 to 8. The colour density of each box in the $x-y$ plane indicates the relevance of each document to a topic. The level of relevance ranges from 0 (low) to 1 (high). Darker colours indicate high while lighter colours indicate lower relevance to the topic.

of the theme around which BCT research is intertwined across disciplines. For example, the keyword "contract" might be related differently to "Computer Science", "Economics", and "Law". "Computer Science" may reflect the algorithmic aspect, "Economics" reflects the transactional aspect, and "Law" represents the regulatory aspect that makes the keyword "contract" to have contextual dependence due to the amalgamation of aspects from "Computer Science", "Economics", and "Law". The dynamic changes that occur in the algorithms and laws governing smart contracts over time might make the theme of "contract" even more complex to understand over time. Therefore, each LDA topic forms the building block of BCT, representing different layers that play a critical role in its interdisciplinary nature. The following sections discuss these topics and keywords in detail.

Table 2.4: LDA topics

LDA topic	Title	Top keywords	Dominance[1]
1	Blockchain and Cryptocurrency Technologies	consensus, protocol, transact, algorithm, block, nod, bitcoin, attack, mine, network	4,849
2	Healthcare, Food Supply and Agriculture Traceability	healthcar, medic, patient, health, food, suppli, chain, product, traceabl, agricultur	3,414
3	Energy Trade and Electric Power Market	energi, trade, electr, market, price, power, grid, carbon, charg, microgrid	363
4	Edge Computing and Vehicle Network Security	vehicl, network, devic, edg, comput, secur, data, internet, communic, learn	7,071
5	Data Privacy, Encryption, and Secure Access	data, privaci, scheme, vote, access, encrypt, user, secur, storag, ident	7,193
6	Multidimensional Data Fusion and BioTech Applications	fusion, cach, slice, lattic, halal, biot, recal, endors, cancer, forest	–
7	Digital Supply Chain and Industrial Research	digit, research, suppli, chain, industri, studi, develop, financi, busi, applic	12,164
8	Deep Learning for Smart Contract and Image/Video Processing	detect, contract, learn, imag, train, vulner, video, smart, neural, deep	550

[1] Number of documents in which the topic is dominant.

2.4.1.1 Topic 1: Blockchain and Cryptocurrency Technologies

There are a total of 4,849 publications on this topic, distributed over the top 10 keywords. The average contribution of Topic 1 in a given document can be 0.133 with minimum 0.00 and maximum 0.90. The probability values (β) of Topic 1 in Table 2.5 show that the first five keywords have the highest relevance to this topic than the last five keywords.

The primary focus of Topic 1 appears to be the technical aspects of blockchain technology, specifically consensus protocols, transaction processing, and network security

Table 2.5: Topic 1 keyword probabilities

Keyword	**β value**	**Keyword**	**β value**
consensus	0.012	nod	0.007
protocol	0.009	bitcoin	0.006
transact	0.008	attack	0.006
algorithm	0.007	mine	0.006
block	0.007	network	0.005

aspects within blockchain networks, such as Bitcoin. The keyword "consensus", with a probability of 0.012, implies an emphasis on consensus algorithms, a key component of blockchain technology. By ensuring the integrity of the distributed ledger, these algorithms ensure that all nodes within a blockchain network agree on the validity of transactions. The word "protocol", with a probability of 0.009, is likely related to the consensus keyword, which refers to consensus protocols that are crucial for the verification of transactions and for the functionality of blockchains. Examples include Proof of Work (PoW) and Proof of Stake (PoS). With a probability of 0.008, the term "transact" indicates that the focus will be on blockchain transactions, including how they are conducted, validated, and recorded.

With a probability of 0.007, the terms "algorithm", "block", and "nod" indicate that blockchain technology is composed of three core components. A consensus protocol refers to the processes underlying the protocol, whereas a block refers to the blocks of data that comprise the blockchain. A node, or individual computer, is a node that participates in a blockchain network. With a probability of 0.006, the keyword "bitcoin" implies a focus on the Bitcoin blockchain, the first and most famous application of blockchain technology. As with the term "attack", this refers to potential cybersecurity concerns within blockchain networks, such as 51% attack and double-spending attacks. These attacks are intended to undermine the security and integrity of these networks. The keyword "mine", with a probability of 0.006, is associated with the mining process, a crucial component of many blockchain networks, especially Bitcoin. Blocks are added to the blockchain, and transactions are validated by solving complex mathematical problems. Lastly, the term "network" is likely to refer to the entire blockchain network, underscoring the interconnected and decentralised nature of this technology.

In summary, Topic 1 appears to focus on the fundamental mechanics and security considerations of blockchain technology, highlighting the role of consensus protocols, transaction processing processes, and the structure of blockchain networks.

2.4.1.2 Topic 2: Healthcare, Food Supply and Agriculture Traceability

In this topic, there are a total of 3,414 publications distributed among the 10 most relevant keywords, whose probability values (β) are shown in Table 2.6. The average contribution of Topic 2 in a given document can be 0.11 with minimum 0.00 and maximum 0.91.

Table 2.6: Topic 2 keyword probabilities

Keyword	**β value**	**Keyword**	**β value**
healthcar	0.012	suppli	0.009
medic	0.012	chain	0.009
patient	0.011	product	0.008
health	0.011	traceabl	0.007
food	0.009	agricultur	0.007

This topic focuses primarily on the applications of blockchain technology in the healthcare and agriculture sectors, particularly in terms of supply chain management and traceability. With a probability of 0.012, both keywords “healthcar” and “medic” indicate a significant discussion related to medical and healthcare services. Blockchain technology could be applied to healthcare to ensure secure patient data management, medication traceability, and the integrity of healthcare services. The term “patient”, with a probability of 0.011, suggests a focus on patient-related information, which may indicate a possible use of blockchain to secure patient medical records. The decentralised and immutable characteristics of blockchain make it an ideal candidate for data management in the healthcare industry that is both secure and patient-centred. With a probability of 0.011, the term “health” likely refers to an overall health system that includes more than just health services, possibly including public health monitoring and preventive health measures as well. Keywords “food”, “suppli”, and “chain”, with probabilities of 0.009, indicate discussions related to the management of food supply chains. It is possible to increase transparency, efficiency and traceability in food supply chains through blockchain technology, ensuring quality and safety throughout the supply chain.

Next, a probability of 0.008 for the term “product” could refer to medical products in the healthcare industry and agricultural products in the food supply chain. Blockchains could be used to track and authenticate these products throughout their lifecycle. The term “traceabl” has a probability of 0.007 and suggests a focus on traceability aspects of these sectors. Blockchain can provide a transparent, verifiable record of product journeys, enhancing traceability and accountability. Finally, the term “agricultur”, with a probability of 0.07, could refer to the use of blockchain in agricultural supply chains, promoting sustainable farming practices and ensuring fair trade.

In general, Topic 2 appears to emphasise aspects of product traceability, supply chain efficiency, and data security in healthcare and agricultural supply chains.

2.4.1.3 Topic 3: Energy Trade and Electric Power Market

On this topic, there are a total of 363 publications distributed in the 10 most relevant keywords, whose probability values (β) are shown in Table 2.7. The average contribution of Topic 3 in a given document can be 0.05 with minimum 0.00 and maximum 0.69.

Table 2.7: Topic 3 keyword probabilities

Keyword	**β value**	**Keyword**	**β value**
energi	0.052	power	0.016
trade	0.029	grid	0.015
electr	0.023	carbon	0.012
market	0.018	charg	0.011
price	0.017	microgrid	0.010

Topic 3 appears to focus on the intersection of blockchain technology and the energy sector, especially in the context of energy trading, grid management, and renewable energy sources, such as microgrids. The keyword “energi”, with a high probability of 0.052, denotes a significant focus on energy-related issues. The application of blockchain technology in areas such as energy generation, distribution, and trading might fall under this broad theme. The word “trade” has a probability of 0.029, indicating that it is likely to refer to energy trading. Blockchain can improve the efficiency and transparency of energy trade by creating a decentralised and secure platform. The term "electr", with a probability of 0.023, indicates an interest in electricity, a key component of the energy sector, as well as issues related to electricity generation, distribution, and pricing. "Market", with a probability of 0.018, could refer to the energy market, suggesting discussions concerning the economics of energy trading, as well as the potential disruption caused by BCT in the energy sector.

With a probability of 0.017, the keyword "price" may refer to energy pricing. Blockchain could improve the transparency of energy pricing and enhance the efficiency of pricing models. "Power" and "grid", each with a probability of 0.016 and 0.015, could represent power distribution networks, otherwise known as grids. Blockchain technology can be used to optimise power distribution, manage charges, and improve the resilience and stability of the grid. With a probability of 0.012, "carbon" may refer to discussions on carbon reduction and the development of low-carbon energy sources. Blockchain is an ideal technology for carbon trading and tracking carbon credits. The term "charg", with a probability of 0.011, may be associated with the charging of electric vehicles and the management of the demand response for power grids. Blockchain may facilitate efficient and transparent energy transactions for electric vehicles. Finally, the "microgrid", with a probability of 0.010, indicates discussions on decentralised energy generation and distribution systems, such as microgrids. Blockchain technology can enable peer-to-peer energy transactions within microgrids, improving the efficiency and resilience of local energy systems.

In summary, Topic 3 discusses the applications of blockchain technology in the energy sector, including energy trading, grid management, energy pricing, and the role of blockchain in promoting low-carbon decentralised energy systems.

2.4.1.4 Topic 4: Edge Computing and Vehicle Network Security

In this topic, there are a total of 7,071 publications distributed across the 10 most relevant keywords, whose probability values (β) are shown in Table 2.8. The average contribution of Topic 4 in a given document can be 0.17 with minimum 0.00 and maximum 0.91.

Table 2.8: Topic 4 keyword probabilities

Keyword	**β value**	**Keyword**	**β value**
vehicl	0.009	secur	0.008
network	0.009	data	0.007
devic	0.009	internet	0.007
edg	0.008	communic	0.006
comput	0.008	learn	0.006

This topic appears to be primarily concerned with the application of blockchain to edge computing, such as vehicular networks, Internet of Things devices, and data security. The keyword "vehicl" with a probability of 0.009 indicates that the discussion revolves substantially around the application of blockchain in vehicular networks. Decentralised networks could be created for vehicle-to-vehicle communication, which would improve the security and efficiency of transportation systems by creating decentralised networks. Similarly, both the terms "network" and "device", which have a probability of 0.009, suggest a focus on networked systems or devices. As a result of blockchain's immutable and transparent nature, applications of blockchain in such networks can help ensure data security and privacy. This probably refers to IoT devices and their interconnected networks. "Edg" and "comput", each with a probability of 0.008, suggest that edge computing is a critical part of the discourse. By bringing data storage and computation closer to the location of need, edge computing improves system response times and saves bandwidth by lowering the distance between storage and computation. When combined with blockchain, it may enhance data handling and processing efficiency across networks while improving security.

With a probability of 0.008, "secur" reflects discussions regarding the security implications and benefits of using blockchain technology in such systems. Due to the fundamental characteristics of the immutability and decentralised control of blockchain data, edge computing systems, IoT devices, and vehicular networks can significantly improve their security. With a probability of 0.007 for the words "data" and "internet", it is likely that these keywords are related to data security and privacy issues on the Internet, a primary concern for IoT devices, vehicular networks, and edge computing systems. By implementing blockchain technology, these concerns could be mitigated, providing a secure and trustworthy platform for data exchange and storage. Furthermore, "communic" and "learn", both with a probability of 0.006, could be linked to machine learning algorithms for communication protocols in these networked systems. Incorporating machine learning into blockchain technology may result in improved communication protocols and enhanced network performance.

Generally, Topic 4 appears to focus on the integration of blockchain technology with edge computing, particularly in the context of vehicular networks and IoT devices.

2.4.1.5 Topic 5: Data Privacy, Encryption, and Secure Access

In this topic, there are a total of 7,193 publications distributed across the 10 most relevant keywords, whose probability values (β) are shown in Table 2.9. The average contribution of Topic 5 in a given document can be 0.17 with minimum 0.01 and maximum 0.91.

Table 2.9: Topic 5 keyword probabilities

Keyword	β **value**	**Keyword**	β **value**
data	0.011	encrypt	0.007
privaci	0.008	user	0.006
scheme	0.008	secur	0.006
vote	0.007	storag	0.006
access	0.007	ident	0.006

The focus of Topic 5 appears to be on the incorporation of blockchain technology into data security and privacy, emphasising encryption schemes, secure data storage, and secure access, as well as applying blockchain to voting systems and digital identity management. The keyword "data", with 0.011 probability, indicates a significant focus on data-related issues. This broad topic may encompass various topics such as data protection, data privacy, and data storage, all of which can be greatly enhanced by blockchain technology. The probability of "privaci" and "secur", each of which has a probability of 0.008 and 0.006, indicates that the importance of privacy and security in blockchain technology cannot be overstated. Clearly, these terms represent the discourse that is concerned with the security, private transactions, and data management capabilities of blockchain technology. There is a probability of 0.008 that the “scheme” refers to the different encryption and consensus schemes used in blockchain technology to protect transactions and data. With a probability of 0.007, the keyword “vote” implies a focus on voting systems. Blockchain technology offers a secure, transparent, and tamper-proof platform for conducting elections and other voting functions.

The term “access”, with a probability of 0.007, probably refers to a method of secure access. Blockchain technology can facilitate secure, permissioned access to data, improving the privacy and security of data. With a probability of 0.007, the term “encrypt” refers to the process of encoding data to prevent unauthorised access. There is a probability of 0.006 that “user” will refer to a user-centric aspect in blockchain-based applications, such as user privacy, user data, or user experience. There is a probability of 0.006 assigned to the keyword “storag”, suggesting that blockchain technology can be used to provide secure, decentralised data storage solutions, improving data integrity and resilience. The probability of “ident” being 0.006 suggests that the term refers to digital identity and its management. Blockchain technology provides a decentralised but secure means of managing digital identities, potentially reducing identity theft and protecting the identity of users.

In summary, Topic 5 covers the application of blockchain technology to improve data security, privacy, and encryption, as well as secure voting systems and digital identity management.

2.4.1.6 Topic 6: Multidimensional Data Fusion and BioTech Applications

The average contribution of Topic 6 in a given document can be 0.03 with minimum 0.00 and maximum 0.38. Table 2.4 shows that it is not a dominant topic in any of the documents in our corpus. However, we discuss the 10 most relevant keywords, whose probability values (β) are shown in Table 2.10.

Topic 6 appears to be more versatile and less straightforward compared to the previous 5 topics discussed, as it covers a variety of keywords that do not seem to fit neatly into a particular category. However, each keyword points to the potential

Table 2.10: Topic 6 keyword probabilities

Keyword	**β value**	**Keyword**	**β value**
fusion	0.018	biot	0.009
cach	0.018	recal	0.009
slice	0.018	endors	0.008
lattic	0.011	cancer	0.008
halal	0.010	forest	0.008

application of blockchain technology in various industries.

Blockchain can provide a secure and trustable platform for integrating and managing data from a variety of sources by the keyword “fusion”, which has a probability of 0.018. With a probability of 0.018, “cach” might refer to caching techniques in network or data storage systems. Blockchain technology could improve the security and reliability of these systems. There is a probability of 0.018 that the term “slice” would refer to network slicing in the context of 5G networks or beyond. Blockchain technology could serve as a tool to manage these virtual network partitions safely and securely. There is a possibility that the “lattic”, with a probability of 0.011, may refer to lattice-based cryptography, which is a method of post-quantum privacy protection. This supports the idea that blockchain technology might be able to withstand the threats posed by quantum computers.

The keyword “halal”, with a probability of 0.010, is an indication that blockchain can be explored for use in certification and supply chain management of halal products, ensuring the authenticity and compliance of the product. With a probability of 0.009, the keyword “biot” may suggest applications in bioinformatics or biotechnology. Blockchain technology could be used for the secure management of data in these fields. The keyword “recal”, with a probability of 0.009, is likely to refer to product recalls within supply chains. Blockchain can help increase transparency and efficiency in the management of product recalls. The terms “endors”, “cancer” and “forest”, each with a probability of 0.008, could represent quite diverse areas where blockchain could be applied, such as endorsement processes, health records management specifically for cancer patients, and forest management or forest supply chains.

In summary, Topic 6 appears to discuss a broad range of potential applications for blockchain technology, including supply chains and healthcare, as well as network systems and cryptography. As a result of the broad scope of these keywords, it is likely that the documents associated with this topic may cover a wide range of blockchain-related applications.

2.4.1.7 Topic 7: Digital Supply Chain and Industrial Research

In this topic, there are a total of 12,164 publications distributed in the 10 most relevant keywords, whose probability values (β) are shown in Table 2.11. The average contribution of Topic 7 in a given document can be 0.28 with minimum 0.00 and maximum 0.92.

Table 2.11: Topic 7 keyword probabilities

Keyword	**β value**	**Keyword**	**β value**
digit	0.005	studi	0.005
research	0.005	develop	0.004
suppli	0.005	financi	0.004
chain	0.005	busi	0.004
industri	0.005	applic	0.004

The focus of Topic 7 appears to be on the application of BCT in various industrial, financial, and business settings, focusing on supply chains, development, and digital applications. There is a probability of 0.005 and 0.004 for each of the keywords "digital" and "applic", indicating a focus on digital applications using blockchain technology. This could include aspects of digital finance, digital identification, and digital supply chains, among other things. Both terms "research" and "studi" have a probability of 0.005, suggesting a focus on academic research and studies on blockchain technology, its potential uses and its impact in a wide variety of industries. It is likely that the keywords "suppli" and "chain" are of particular interest to supply chain managers, as each has a probability of 0.005. Transparency, traceability, and efficiency can be improved through blockchain technology, and these terms indicate that these applications are an important component of this topic.

"Industri", with a probability of 0.005, suggests a focus on industrial applications of blockchain technology, which could encompass a variety of industries, such as manufacturing, energy and healthcare. A probability of 0.004 suggests that "develop" refers to the development of blockchain technologies and applications. This might mean both the technical development of blockchain systems and the broadening of the field of blockchain. "Financi", with a probability of 0.004, indicates a focus on financial applications of blockchain technology, including digital currencies, decentralised finance, and financial transactions. Finally, the probability of "busi", which is 0.004, suggests a focus on business applications of blockchain technology, including business transactions, business process optimisation, etc.

In summary, Topic 7 appears to dive into the various applications of blockchain technology in different sectors, with a particular emphasis on digital applications, supply chain management, and the financial sector. This may also highlight the importance of ongoing research and development in this field.

2.4.1.8 Topic 8: Deep Learning for Smart Contract and Image/Video Processing

In this topic, there are a total of 550 publications distributed in the 10 most relevant keywords, whose probability values (β) are shown in Table 2.12. The average contribution of Topic 8 in a given document can be 0.05 with minimum 0.00 and maximum 0.73.

Table 2.12: Topic 8 keyword probabilities

Keyword	**β value**	**Keyword**	**β value**
detect	0.024	vulner	0.012
contract	0.017	video	0.010
learn	0.015	smart	0.010
imag	0.014	neural	0.010
train	0.013	deep	0.010

Topic 8 focuses on the intersection of blockchain technology, smart contracts, and machine learning, particularly deep learning and neural networks, with an emphasis on detection and training processes, as well as possible vulnerabilities. The keyword "detect", with a probability of 0.024, could suggest a focus on detection mechanisms, which can be applied to various contexts, such as anomaly detection, fraud detection, or vulnerability detection in blockchain networks or smart contracts. With a probability of 0.017, the "contract" is likely to refer to smart contracts, a key element of blockchain technology. Smart contracts are self-executing contracts in which the terms of the agreement are written directly into the code. With a probability of 0.015 and a probability of 0.010 for terms "learn" and "neural" and "deep", this suggests a significant emphasis on deep learning or neural networks. This topic is likely to involve the application of deep learning techniques to blockchain technology in conjunction with these terms. "Imag" and "video", which have probabilities of 0.014 and 0.010, may indicate a preference for visual data, hinting at the possibility of applying machine learning to image or video analysis in a blockchain context.

In the context of artificial neural networks and deep learning systems, "train", with a probability of 0.013, is likely to refer to the training of machine learning models. With a probability of 0.012, the "vulner" suggests that the focus is on vulnerabilities, perhaps within smart contracts or within blockchain networks themselves. The context suggests that these vulnerabilities may be detected or mitigated using machine learning techniques. "Smart", with a probability of 0.010, is likely to refer to smart contracts, but could also refer to smart devices or IoT (Internet of Things) in a broader sense.

Topic 8 appears to focus on the application of machine learning, specifically deep learning, in the context of blockchain technology and smart contracts. This includes detecting and addressing vulnerabilities, training neural networks, and processing visual data.

2.4.2 Understanding the Extent to Which BCT Research is Interdisciplinary: Interpreting the IDEA Framework

2.4.2.1 Research Focus Evolution

BCT research has shown significant progress and covers a wide range of topics. Figure 2.8 illustrates the changes in the main topics within this field over time. An upward trend can be seen in all identified topics, with Topics 4 (edge computing and vehicle network security) and 7 (digital supply chain and industrial research) showing a particularly significant increase. This increase can be attributed to the growing requirement for computing resources in cryptocurrency mining and applications in the digital supply chain, which might attract significant attention of researchers. Topics 3 (energy trade and electric power market) and 8 (deep learning for smart contract and image/video processing), although still showing an upward trend, have done so at a slower pace. Topic 4 underwent a significant shift in focus in 2020, while Topic 5 (data privacy, encryption, and secure access) remained consistently prominent until that year.

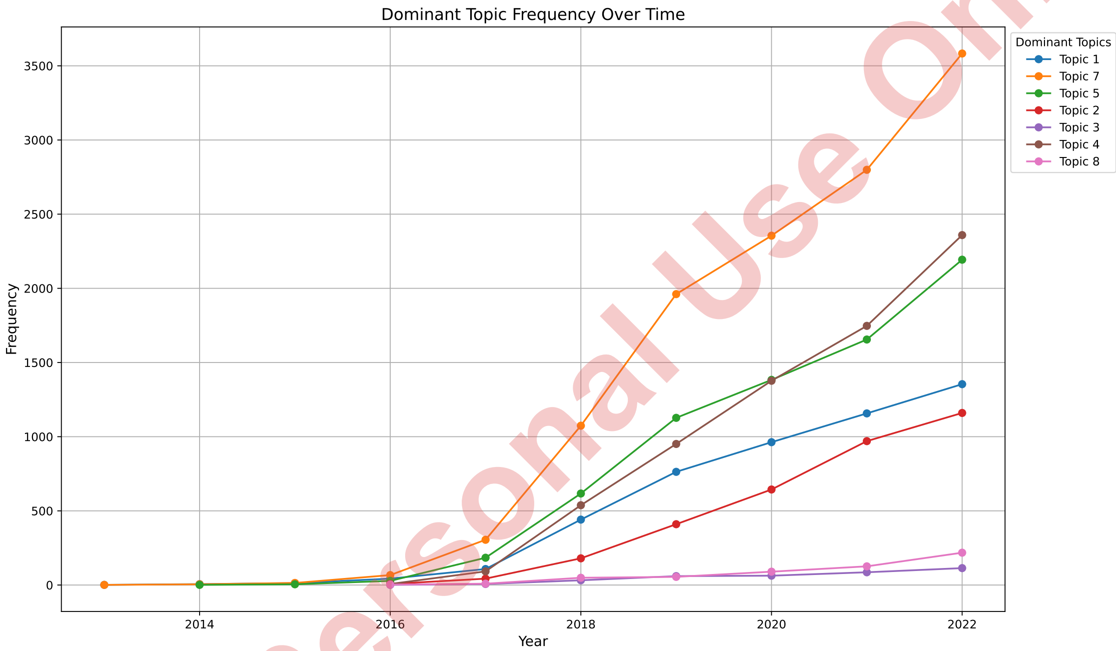

Figure 2.8: Dominant topics over time

Notes: The x-axis of the figure shows the year and y-axis shows the count of documents, categorised according to their dominant topics as identified by LDA analysis. The different colours of the trend lines correspond to various dominant topics.

2.4.2.2 Dynamic Research Interests

To understand how topic contributions emerged in each publication over time, we calculated topic contributions over time. Figure 2.9 shows that most topics have stable contributions. For example, vehicle network technologies and edge computing (Topic 4) maintained a strong presence throughout the period, reflecting their growing importance in smart city development. Similarly, research on supply chain optimisation and blockchain integration (Topic 7) remained steady, suggesting ongoing efforts to improve transparency and efficiency in various industries.

Interestingly, some topics displayed dynamic trends. The focus on blockchain consensus protocols and transaction security (Topic 1) peaked in 2013 but gradually declined as initial concerns about these aspects stabilised. This could be attributed to the maturing of blockchain technology and the emergence of new research areas within the field.

On the other hand, research on blockchain-based healthcare applications (Topic 2) and smart grid energy trading platforms (Topic 3) had lower overall contributions, perhaps due to regulatory hurdles or slower adoption rates in these complex sectors. On the contrary, data privacy and identity management (Topic 5) gained significant traction in 2014, likely spurred by increased public awareness and stricter data protection regulations. This trend continued to peak in 2017, highlighting the ongoing concerns and active research efforts in this critical domain.

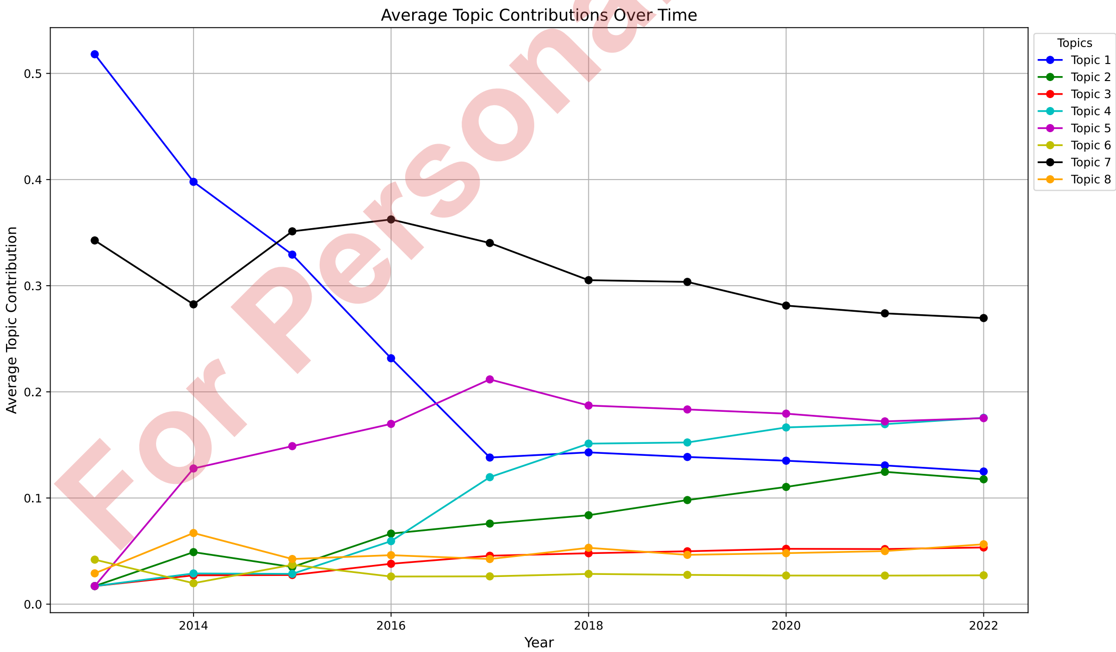

Figure 2.9: Average topic contribution in a publication over time

Notes: The x-axis shows the year, and the y-axis shows the average topic contribution in a publication. The colours of the trend lines indicate the topic.

2.4.2.3 Synergistic Knowledge Areas

To understand the synergistic relationships and knowledge areas, we performed co-occurrence mapping of the topic-keywords. The results showed three major trends: high, medium, and low keyword affinity. Figure 2.10 shows high keyword affinities by a strong bright-coloured box (e.g., data and transact or data and network), a medium-coloured box (e.g., consensus and network or consensus and data) for medium keyword affinity, and a light-coloured box (e.g., privaci and cach or privaci and imag) for low keyword affinity.

High keyword affinities indicate that the two keywords often appear together and could be a synergistic knowledge area. For example, the frequent pairing of "data" and "transact" or "data" and "network" in Topic 1 (blockchain and cryptocurrency technologies) suggests a close relationship between these concepts. On the other hand, medium or low keyword affinities indicate that the two keywords might appear together, but their relationship or association might be weak, indicating weak synergy between them. For instance, the co-occurrence of "consensus" and "network" or "consensus" and "data" in Topic 1 points to some overlap in these areas, potentially related to research on distributed consensus mechanisms for blockchain networks. Furthermore, the low affinity between "privacy" and "cache" or "privacy" and "image" in Topic 5 (data privacy, encryption, and secure access) indicates that these keywords may not be directly related to each other in the context of data privacy and security research.

Further examination of keyword affinities between topics reveal interesting insights. For example, we observe high keyword affinities between topics that might be expected to have close connections. For example, Topic 1 and Topic 4 (edge computing and vehicle network security) highlight the potential of blockchain to be applied in these emerging domains. Similarly, the high affinity between Topic 1 and Topic 7 (digital supply chain and industrial research) indicates an ongoing interest in leveraging blockchain for improved transparency and efficiency in supply chains. Furthermore, the strong association between Topic 1 and Topic 5 suggests that data privacy concerns are a significant consideration in blockchain research and development.

On the contrary, medium- to low-affinities between topics might indicate fewer direct overlaps. For example, low affinity between Topic 2 (healthcare, food supply and agriculture traceability) and Topic 4 suggests that, while there might be some potential for intersection, the research focus in these areas is currently distinct. Similarly, Topic 3 (energy trade and electric power market) and Topic 6 (multidimensional data fusion and biotech applications) indicate that these areas have largely separate research agendas within the broader field.

Overall, we can observe high affinities between keywords of Topic 1 and 4, Topic 1 and 7, Topic 1 and 5, medium to low affinities between Topic 2 and 4, Topic 2 and 7, Topic 3 and 6, Topic 4 and 8, respectively.

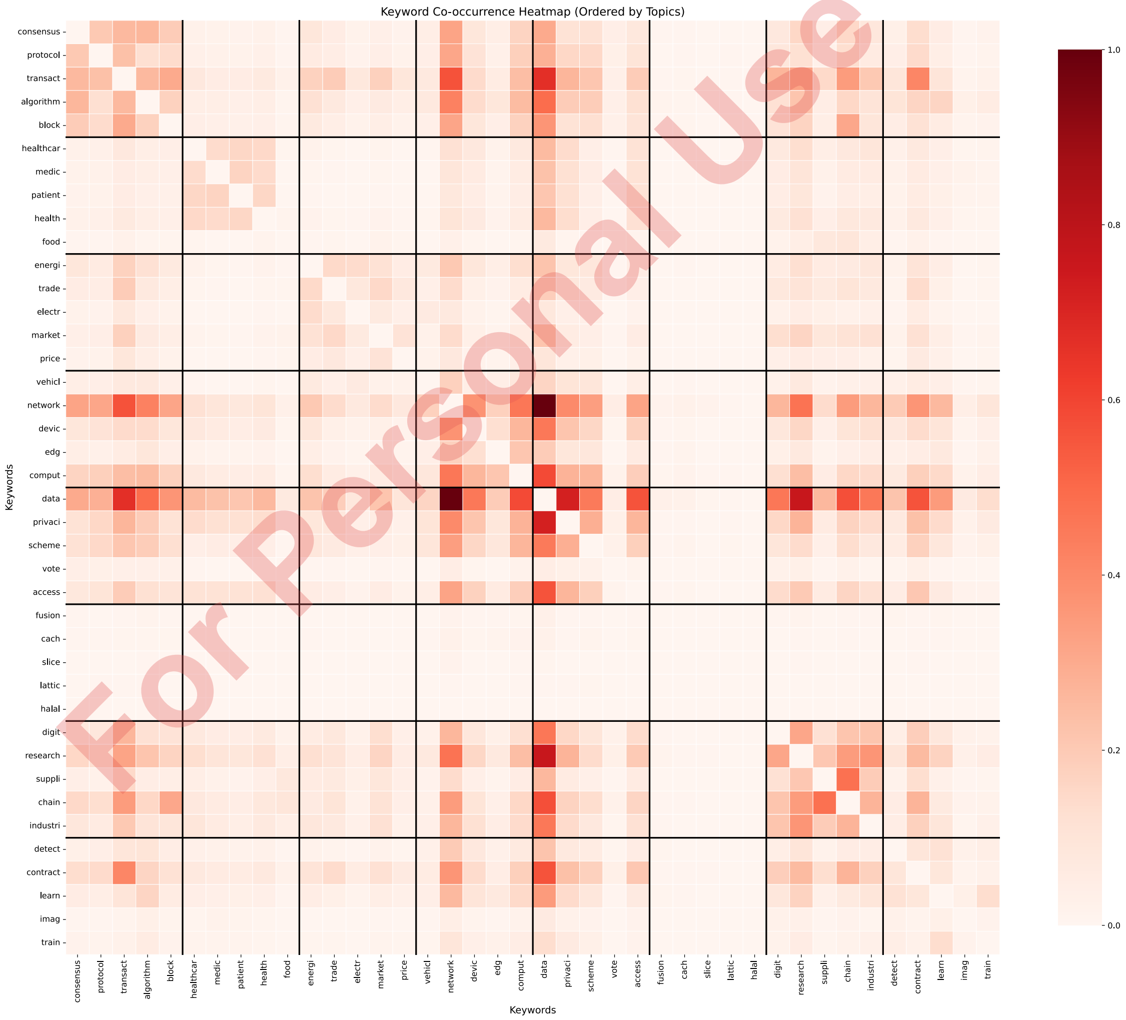

Figure 2.10: Co-occurrence matrix of top 5 keywords across all topics

2.4.2.4 Academic Convergence or Divergence

To understand the Academic Convergence or Divergence in blockchain research, Figure 2.11 shows the richness (left-hand side) and the evenness (right-hand side). A topic with high richness indicates academic convergence across multiple disciplines, potentially leading to interdisciplinary research, whereas a topic with low richness may be confined to a specific discipline, suggesting an academic divergence.

On the other hand, the evenness plot complements the richness plot by showing how evenly research contributions are distributed across the disciplines. Therefore, a high evenness index implies a balanced contribution from various disciplines, indicating academic convergence, while a low evenness index implies academic divergence.

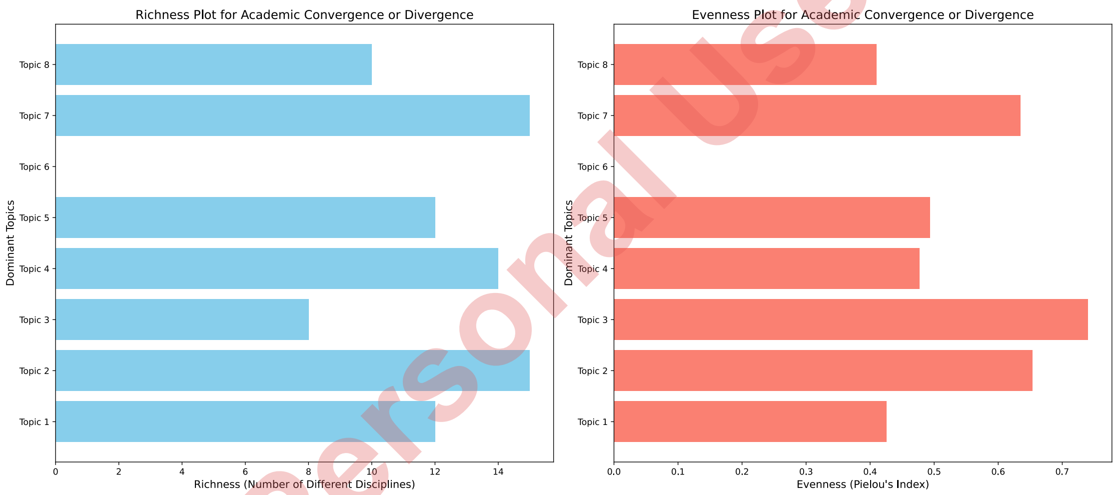

Figure 2.11: Richness and evenness

Topic 2 (healthcare, food supply and agriculture traceability) and Topic 7 (digital supply chain and industrial research) exhibit the highest richness, spanning contributions from 15 distinct disciplines. This broad disciplinary engagement highlights the widespread interest and relevance of these topics across various fields. In contrast, Topic 3 (energy trade and electric power market) demonstrates the lowest richness, suggesting a more focused research interest confined to specific disciplines. Topics 1 (blockchain and cryptocurrency technologies) and 5 (data privacy, encryption, and secure access) occupy a middle ground, indicating substantial interdisciplinary involvement but not as extensive as Topics 2 and 7.

While Topic 3 has the lowest richness, it has a high Evenness Index (0.74). This indicates that although fewer disciplines engage with Topic 3, their contributions are remarkably balanced, suggesting harmonious integration and a collaborative interdisciplinary environment. Similarly, Topic 2, despite its high richness, also maintains a high Evenness Index (0.65), implying a relatively balanced contribution from each of the 15 disciplines involved. Topic 1, on the other hand, shows a lower Evenness Index, suggesting that contributions are more skewed towards certain disciplines, indicating less balanced interdisciplinary collaboration.

Topic 7 (digital supply chain and industrial research) emerges as a prime example of high academic convergence, demonstrating both broad interdisciplinary interest (high richness) and balanced contributions from various disciplines (high evenness). Topic 3 (energy trade and electric power market) represents a niche area with focused disciplinary involvement but a balanced distribution of contributions within those disciplines. Topics 1, 4 (edge computing and vehicle network security), 5, and 8 (deep learning for smart contract and image/video processing) exhibit varying levels of richness and evenness, reflecting a spectrum of interdisciplinary engagement and collaboration within the blockchain research landscape. Topic 8 with lower values in both richness and evenness, suggests a potential area for further interdisciplinary exploration and collaboration to fully realise its potential.

2.4.2.5 Thematic Overlaps

The Figure 2.12 visually represents the interdisciplinary landscape of blockchain research, showcasing Thematic Overlaps between topics and disciplines. Nodes represent these entities, connected by flow lines whose thickness indicates the strength of research focus at each intersection, highlighting potential areas for collaboration.

For example, the flow width connecting Topic 7 (digital supply chain and industrial research) with “Computer Science,” “Engineering & Technology,” and “Business, Economics & Decision Science” is thicker, indicating significant overlap. This highlights the convergence of computational, engineering, and economic approaches in blockchain research.

Topics 4 (edge computing and vehicle network security), 5 (data privacy, encryption, and secure access), and 1 (blockchain and cryptocurrency technologies) also exhibit significant overlaps with multiple disciplines, indicating their broad relevance and potential for interdisciplinary collaboration.

In general, Topics 7, 4, 5, and 1, respectively, have thick lines connecting to all other disciplines, suggesting substantial overlaps and a volume of research. However, Topics 2 (healthcare, food supply and agriculture traceability), 8 (deep learning for smart contract and image/video processing), and 3 (energy trade and electric power market) have thinner lines, suggesting minimal overlap. Computer Science, as evident from the numerous thick connections, emerges as a major contributor to most topics, underscoring its central role in blockchain research.

2.4.2.6 Evolution of Research Complexity

Topic Contributions (HHI): The Figure 2.13 shows the HHI for each topic over the years, providing insights into the Evolution of Research Complexity within blockchain research. The x-axis of the figure represents the temporal dimension, whereas the y-axis represents the concentration of topic contributions.

We observed that all topics showed a downward trend in HHI values over time, indicating that BCT research evolved to greater complexity in the last decade, as evident by the decreasing HHI of topic contributions.

Topics 2 (healthcare, food supply and agriculture traceability) and 8 (deep learning for smart contracts and image/video processing) initially focused within their respective disciplines, similar to the early stages of BCT. Although Topic 2 gained traction in early 2016, its contributions remained largely confined to its own field, as evidenced by the increasing HHI. In contrast, Topics 4 (edge computing and vehicle network security) and 8 started as discipline-specific, but later transcended their boundaries, attracting interest from other fields. In particular, their HHIs began to increase significantly during 2017 and 2018, before entering a steady decline. This drop in

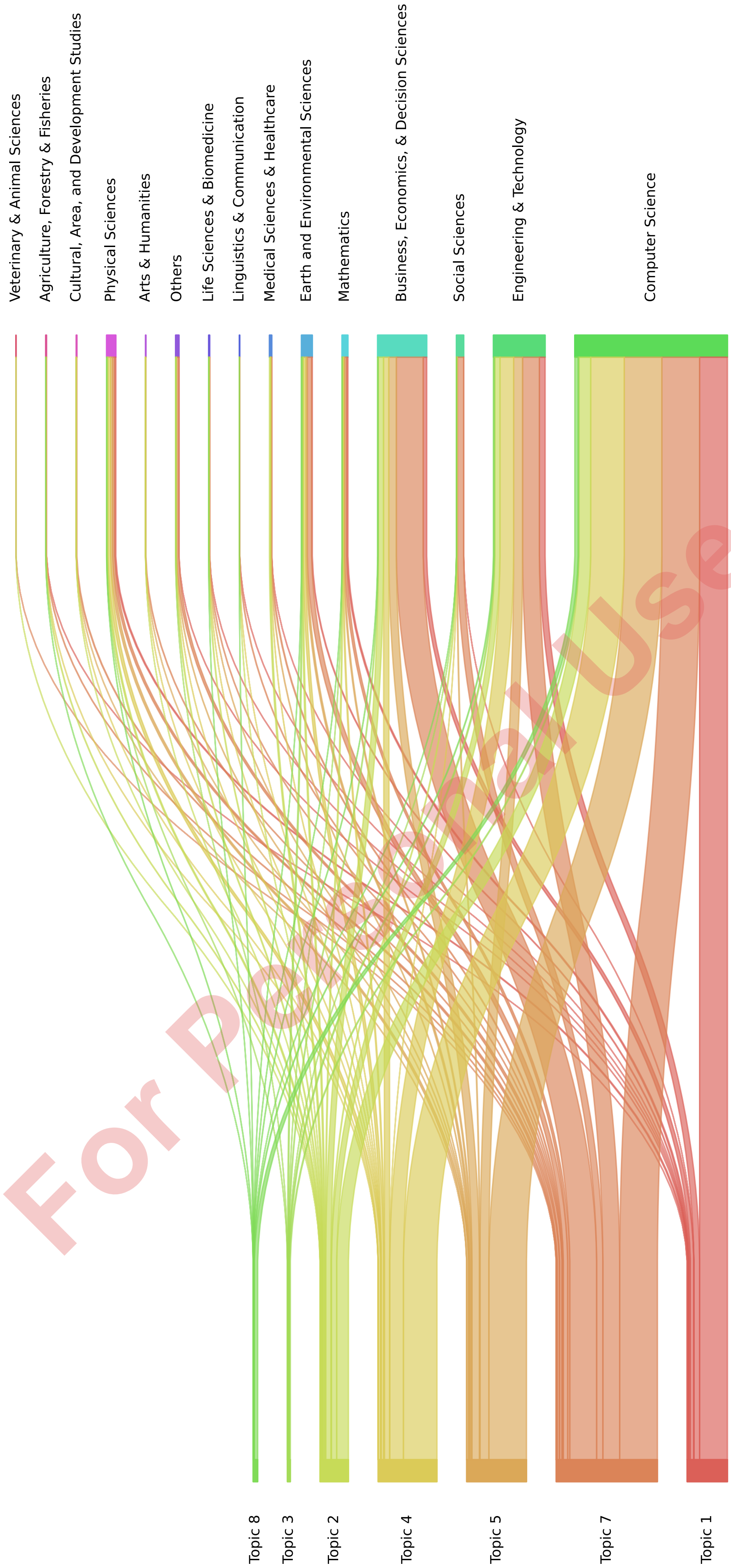

Figure 2.12: Sankey diagram showing Thematic Overlaps between topics and disciplines

Notes: This figure visually represents the interdisciplinary landscape of blockchain research, showcasing Thematic Overlaps between topics and disciplines. Nodes represent these entities, connected by flow lines whose thickness indicates the strength of research focus at each intersection, highlighting potential areas for collaboration.

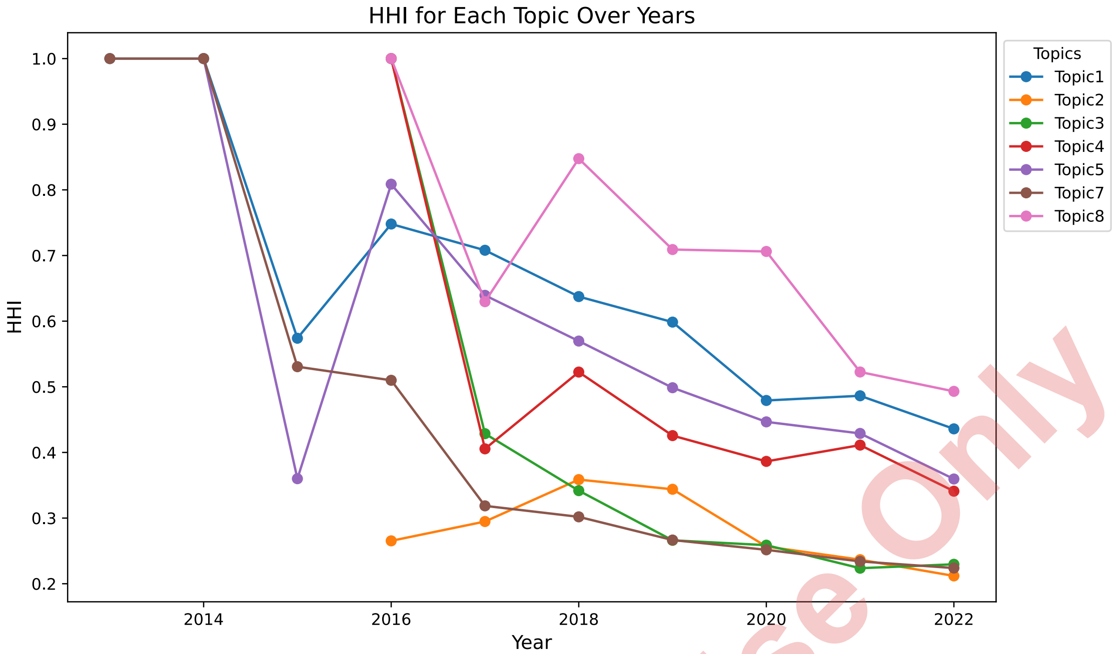

Figure 2.13: HHI values of topic contributions over time

HHI coincides with the maturation of BCT, prompting the broader application of smart contracts in various fields and fostering increased interdisciplinarity.

Topics 1 (blockchain and cryptocurrency technologies) and 5 (data privacy, encryption, and secure access) started as highly specialised within their respective disciplines. However, after 2016, both saw a consistent decline in their HHI. This can be attributed to their role as fundamental building blocks of blockchain technology. As diverse disciplines recognised their potential, these areas attracted increasing research participation, transforming them into vibrant hubs of interdisciplinary collaboration.

In contrast, Topics 3 (energy trade and electric power market) and 7 (digital supply chain and industrial research) exhibited a significantly higher change in HHI values between 2018 and 2022 compared to Topics 1, 4 (edge computing and vehicle network security), 5, and 8. This suggests that these topics underwent a more dramatic evolution in terms of cross-disciplinary engagement during this period, possibly due to emerging applications or disruptive advancements within their fields.

Unique Disciplines: The Figure 2.14 shows the temporal Evolution of Research

Complexity from the standpoint of disciplines. The increase in the number of unique disciplines every year reflects the growing complexity of BCT research. Notably, since 2014, the influx of unique disciplines suggests BCT's expanding applications beyond its foundational aspects. This trend indicates a growing recognition of BCT's potential in addressing complex challenges across various domains.

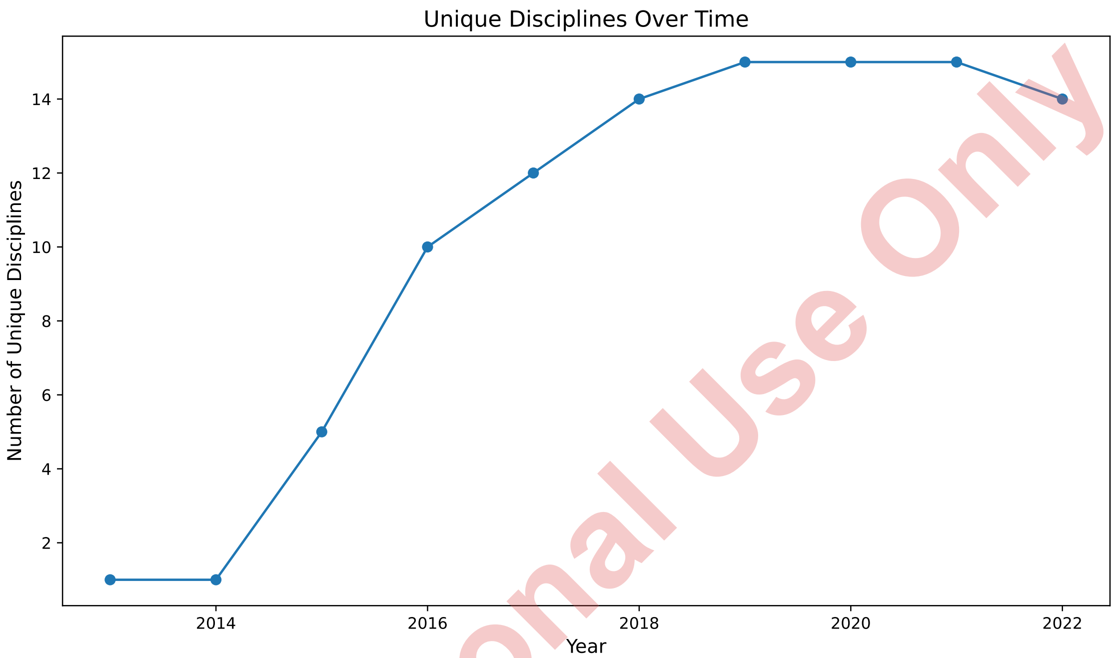

Figure 2.14: New disciplines over time

The graph shows a rapid increase in unique disciplines engaged in BCT research, with a notable jump between 2017 and 2018, followed by a recent plateau. This pattern suggests a phase of rapid diversification and subsequent maturation, with broader participation from established disciplines.

Overall, the increased engagement of unique disciplines in blockchain research over time highlights BCT's growing interdisciplinary nature and its integration into mainstream research, emphasising the need for collaborative approaches to fully realise its potential.

2.4.2.7 Research Landscape Breadth

Research Landscape Breadth is a measure of how widely a topic is studied in different research disciplines. We use Figure 2.15 for our interpretation and analysis, which serves as a quantitative anchor to understand Research Landscape Breadth.

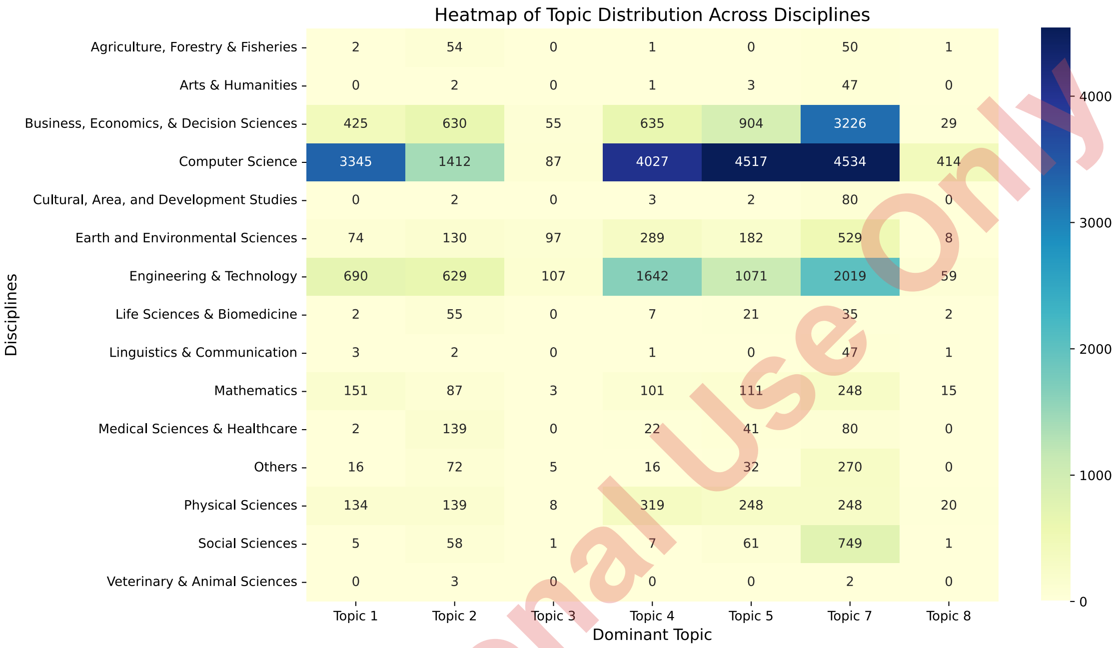

Figure 2.15: Count of dominant topics across different disciplines

Notes: This heatmap illustrates the distribution of documents across dominant topics and disciplines. The x-axis displays the topics, while the y-axis lists the various disciplines. The colour intensity of each box corresponds to the number of documents associated with a specific topic-discipline combination, with darker shades representing higher counts. This visualisation provides insights into the concentration of research efforts across different topics and disciplines, highlighting areas of greater focus and potential interdisciplinary connections.

Disciplines such as “Computer Science”, “Business, Economics, & Decision Sciences”, and “Engineering and Technology” have high counts for Topics 4 (edge computing and vehicle network security), 5 (data privacy, encryption, and secure access), and 7 (digital supply chain and industrial research), as evidenced by the corresponding cells for the dominant topics. This suggests a broader research landscape in which these topics are widely studied in a variety of disciplines and not confined to a single one. In contrast, disciplines with low occurrence values in cells suggest a more focused or narrower research landscape, potentially limiting interdisciplinary collaboration and knowledge exchange in those areas.

2.4.2.8 Scholarly Versatility

Scholarly Versatility is the engagement of an author with diverse topics within and across disciplines, reflects adaptability, and exemplifies the polymath phenomenon.

Versatility can be observed through the evolving pattern of authorship over time. In this regard, Figure 2.16 shows an exponential increase in unique authors contributing to BCT's research landscape between 2013 and 2022, particularly after 2016. This trend shows that between 2019 and 2020, unique first authors grew the least as BCT was gaining traction, while between 2021 and 2022, a substantial number of unique authors contributing to BCT research can be attributed to the maturation of BCT that provided better research opportunities.

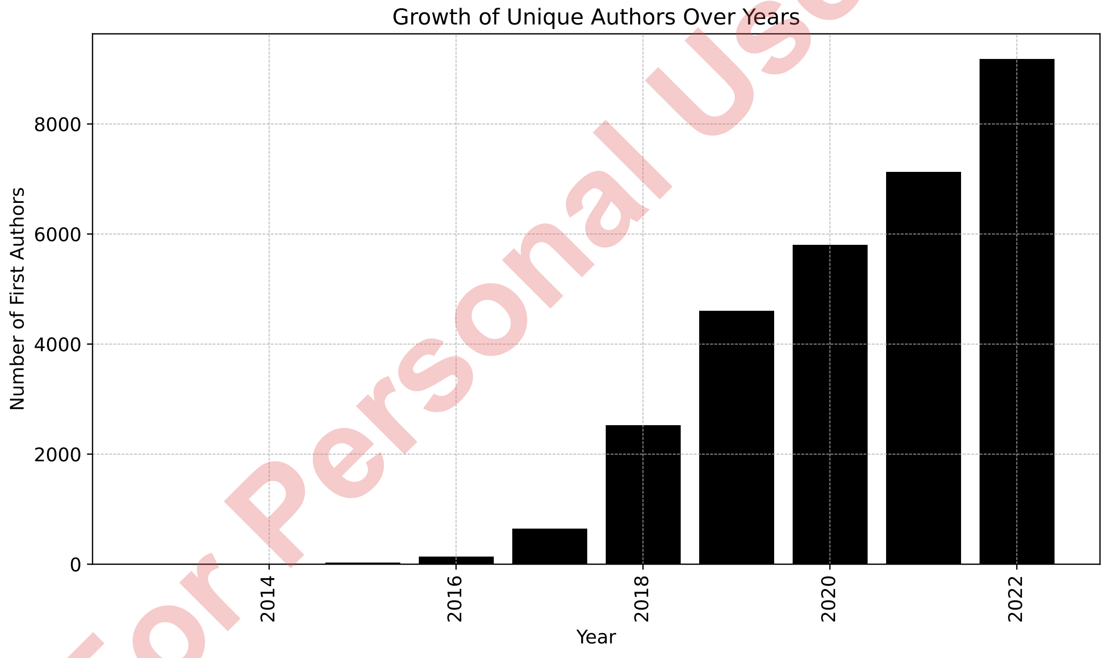

Figure 2.16: Unique authors contributing as "First Authors" over time

Author Versatility: To further complement our analysis, we use Figure 2.17 to discuss Author Versatility. The x-axis represents disciplines, and the y-axis represents the versatility score. The versatility score (VS) was calculated using Shannon's entropy, where higher entropy or VS implies that an author has contributed to a wide array of topics, implying greater versatility in their research focus. In contrast, a low VS indicates a more specialised or focused contribution.

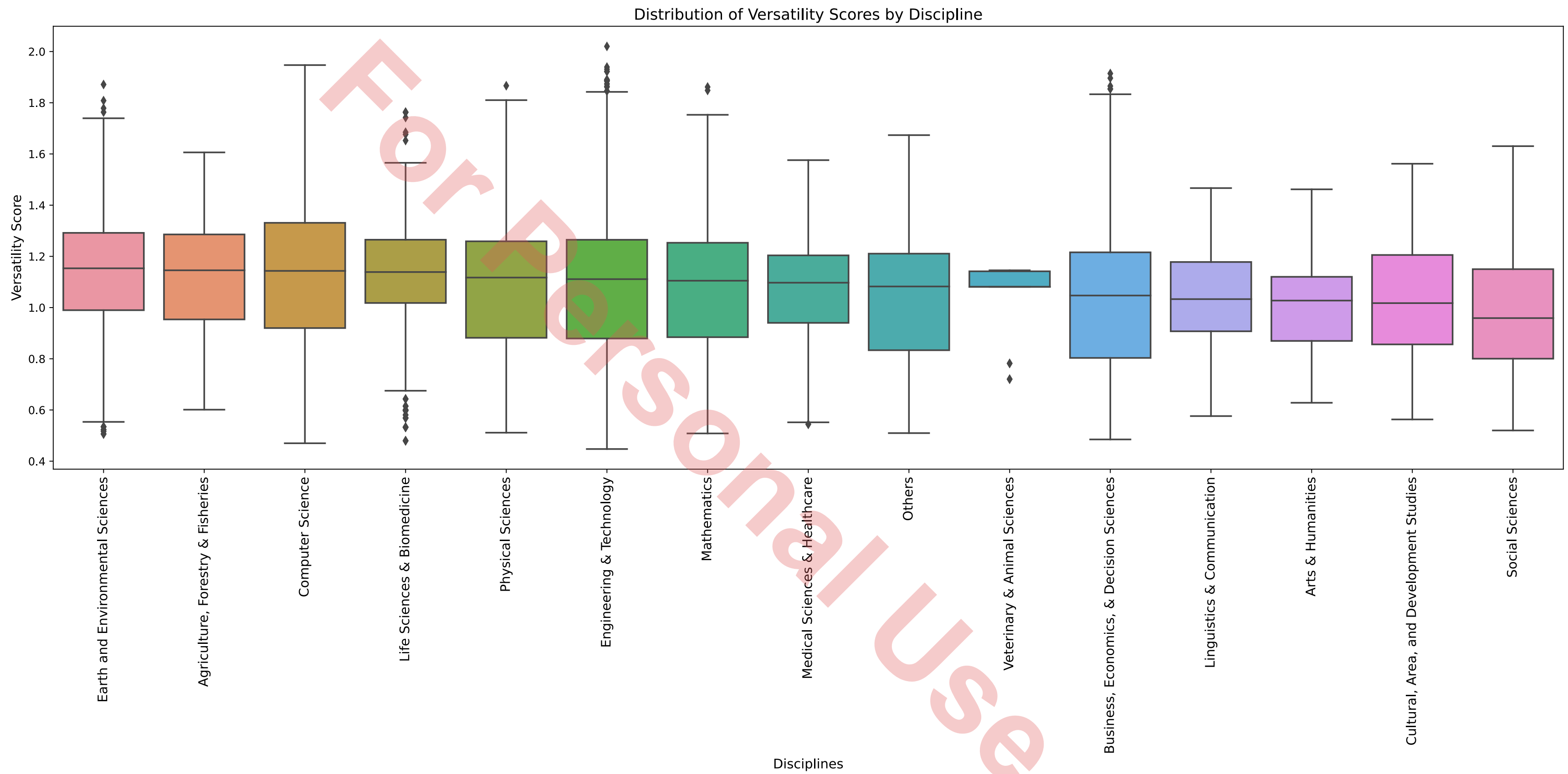

Figure 2.17: Box plot depicting the distribution of “Author Versatility Scores” across academic disciplines

This boxplot illustrates the distribution of “Versatility Scores (VS)”, calculated using entropy, among authors in various disciplines. Each box in the plot represents a discipline, and the height of the box indicates the range of VS within that discipline. The median VS for each discipline is marked by a line within the box. Disciplines are ordered from left to right based on their median VS, descending from highest to lowest. The whiskers of the boxes indicate the spread of the VS scores.

Disciplines with high VS imply the presence of both specialists and generalists. For example, disciplines at the left end of the boxplot have a higher median VS, suggesting that authors of these areas are more likely to contribute to interdisciplinary research. For example, “Computer Science”, “Engineering & Technology”, and “Business, Economics, & Decision Sciences” have a higher median VS, and the boxes are relatively spread out. This suggests that authors in these disciplines engage in a wider array of topics.

However, disciplines on the right hand side of Figure 2.17 tend to have a lower median VS, implying a more specialised or focused area of research. For example, “Arts & Humanities” and “Cultural, Area, and Development Studies” have a lower median VS and a more concentrated box, suggesting less topical diversity. Authors in these fields may be more specialised, focusing on a narrower set of issues related to blockchain, such as legal frameworks or financial applications.

Some disciplines, such as “Computer Science”, have a wider box and whiskers, suggesting a wider range of VS. This could mean that the discipline has specialist researchers focusing on niche topics and generalists exploring a range of topics. However, disciplines with narrow boxes and whiskers, such as “Arts & Humanities”, imply a more uniform versatility score for authors. This could indicate a shared focus or a common set of challenges that researchers in this discipline address.

2.4.2.9 Collaborative Trends

To complement our understanding of Collaborative Trends in blockchain research, we refer to Figure 2.18. The chart represents the diversity of the first authors who contributed to the different dominant topics. Higher values in the radar spokes indicate a higher level of diversity of the first authors contributing to that particular topic. For example, Topics 2 (healthcare, food supply and agriculture traceability), 3 (energy trade and electric power market), and 7 (digital supply chain and industrial research) show the highest diversity scores, suggesting that the first authors are versatile and contribute to that topic in a variety of disciplines. High author diversity within these topics aligns with the multifaceted nature of BCT, potentially attracting the collaboration of experts from various disciplines due to its decentralised structure, the diverse application potential, and the need for cross-domain expertise.

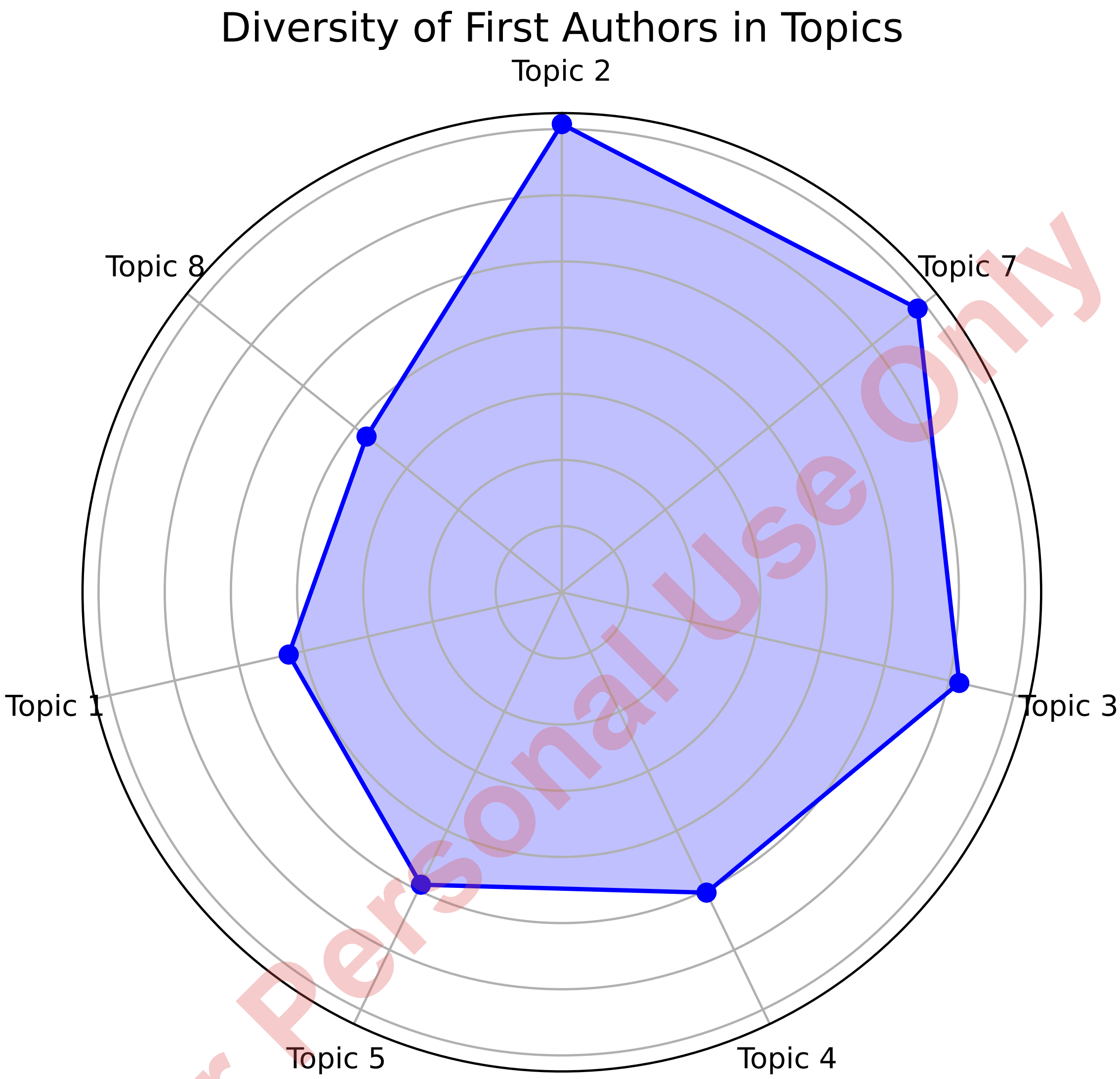

Figure 2.18: Radar chart of "First Author" diversity across topics

Notes: This radar chart visualises the "Shannon Diversity Index" for each dominant topic, highlighting the diversity of disciplines among the first authors within these topics. Higher values along the radar spokes indicate greater diversity in the disciplinary backgrounds of the first authors. This suggests a more interdisciplinary approach within the topic, reflecting the multidisciplinary nature of the authors' contributions.

On the contrary, Topics 1 (blockchain and cryptocurrency technologies), 4 (edge computing and vehicle network security), 5 (data privacy, encryption, and secure access), and 8 (deep learning for smart contract and image/video processing) have a lower author diversity, which can be attributed to inherent specialisation. These topics require in-depth disciplinary expertise, which may limit the participation of researchers outside these specific fields.

2.4.2.10 Research Focus Concentration

Topics Diversity: We used the HHI for the eight topics to quantify the level of concentration. Figure 2.19 shows that the HHI values range from 0.00 to 0.58, indicating different levels of concentration between disciplines for the different dominant topics identified. A lower HHI score indicates diversification of research across a wide range of disciplines (high complexity and interdisciplinarity), while a high HHI score indicates that research is more concentrated in a small number of disciplines (low complexity and disciplinarity)(refer to Section B.1 to see an illustrative example for Research Focus Concentration).

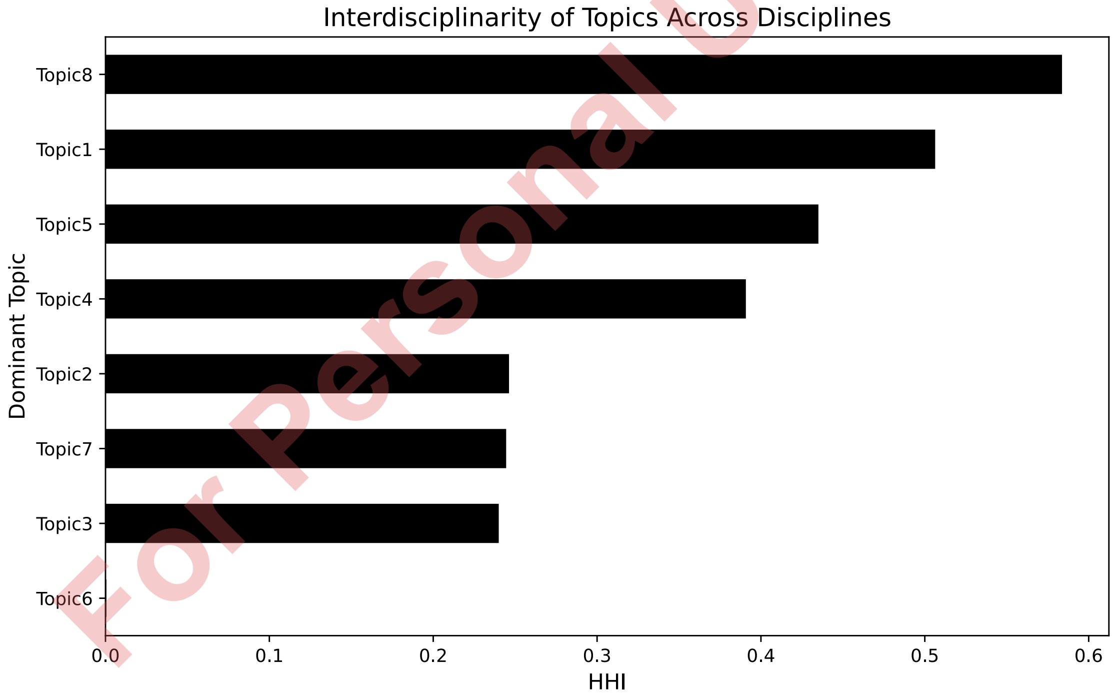

Figure 2.19: HHI values for topics

Notes: This figure illustrates the diversity of blockchain research across eight topics, represented by HHI values. Low HHI values indicate high interdisciplinarity (e.g., Topic 3), whereas high values indicate intradisciplinarity (e.g., Topics 1 and 8). The figure offers a visual representation of the extent of interdisciplinarity within blockchain research based on HHI values.

Figure 2.19 is arranged in descending order with Topics 1 (blockchain and cryptocurrency technologies) and 8 (deep learning for smart contract and image/video processing) having HHI of 0.50 and 0.58 respectively. These two topics have the highest HHI, indicating low interdisciplinarity and high disciplinary concentration. Both topics revolve around the foundational aspects and intricate applications of BCT such as smart contracts and data, suggesting niche areas of specialised disciplines such as "Computer Science".

Topics 2 (healthcare, food supply and agriculture traceability), 3 (energy trade and electric power market), and 7 (digital supply chain and industrial research) have the lowest HHI value (i.e., high interdisciplinarity) of 0.24, 0.23, and 0.24, respectively. This suggests that research on these topics is highly interdisciplinary and diverse. These topics represent application areas of blockchain where knowledge intersects with many different domains of practice, suggesting interdisciplinary research input.

Topics 4 (edge computing and vehicle network security) and 5 (data privacy, encryption, and secure access) have HHI of 0.39 and 0.43 respectively. These topics have a medium interdisciplinarity and revolve around specialised application areas of blockchain where input from other research disciplines might be infrequent.

Diversity Over Time: To understand how average topic concentration changed over time, we analysed average topic concentrations (HHI) of dominant topics across years. To achieve this, we examine the HHI values between the period 2013-2022 to understand the decades-long behaviour of research and the evolution of interdisciplinarity within blockchain technology.

Observing the annual averages of the HHI scores in Figure 2.20 revealed an interesting trend. Both years 2013 and 2014 had an HHI score of 1.0, after which there was a steep decline in 2015. There was a slight bump during 2016 after which there is a consistent decline in average HHI over time.

2.4.3 Additional Analysis

A comprehensive Jaccard similarity analysis (refer to Section A.4) was performed to identify the overlap of dominant themes between disciplines. The Jaccard similarity

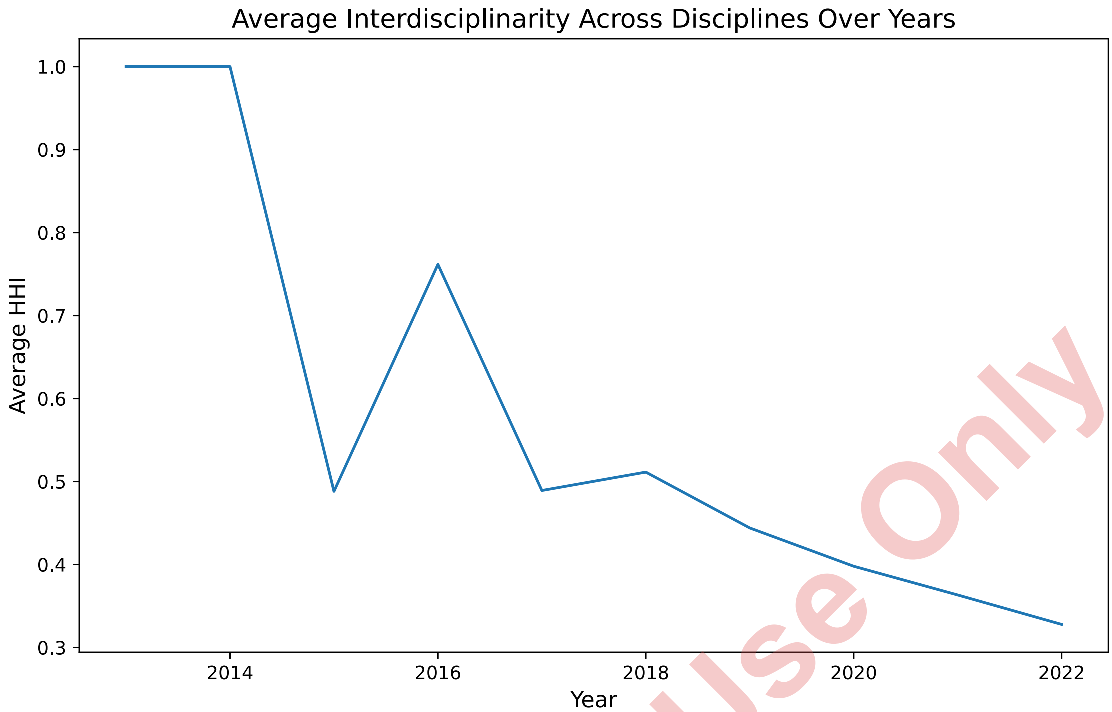

Figure 2.20: Average HHI for disciplines across 2013 to 2022

Notes: This figure traces the evolution of diversity in blockchain research from 2013 to 2022, using the HHI. A decline in HHI values from 1.0 in 2013 to 0.30 in 2022 signals an increasing dispersion of blockchain research across various disciplines, reflecting a trend towards growing interdisciplinarity in this field.

index, a statistic used to compare the similarity and diversity of sample sets, was used to measure the degree of overlap.

The results suggest that disciplines such as “Computer Science & Engineering” and “Finance & Economics” might show a higher Jaccard index, indicating a significant overlap in the topics they cover on the blockchain. This suggests an interdisciplinary synergy between these fields, perhaps in areas such as financial technology and blockchain infrastructure. On the other hand, discipline like “Law & Policy” and “Computer Science & Engineering” display a lower Jaccard index, which means fewer overlap between topics. This points to specialised niches within blockchain research that are primarily explored by experts in the respective fields, such as legal frameworks around blockchain in the case of Law, or technical aspects in the case of Computer Science.

2.5 Discussion and Analysis of Findings

This study addresses two overarching research questions: *the current state of blockchain research* and *the extent to which blockchain research is interdisciplinary*. To address the two main research questions, we developed an IDEA framework consisting of 10 different aspects to explore and four different types of analysis. We use subquestions to divide the main research questions into specific areas of inquiry. We used the IDEA framework to analyse the entire field of blockchain, regardless of the disciplinary affiliation of the publications. In this way, we aim to develop a broader understanding of the entire blockchain landscape.

Our analysis performed LDA on the entire BCT literature. The LDA model divided the entire research landscape into eight topics (or themes). We analysed these topics first, and then, we shed light on the ten aspects of our proposed IDEA framework employing its four major categories: Temporal and Evolutionary Analysis, Interdisciplinary Dynamics, Research Complexity and Behaviour, and Research Focus and Concentration. This section is structured according to these four categories in order to discuss the overall findings.

2.5.1 LDA Topics

Our analysis divides blockchain research across disciplines into eight topics. We found a large number of research papers distributed on Topics 4 (edge computing and vehicle network security), 5 (data privacy, encryption, and secure access), and 7 (digital supply chain and industrial research). These topics, which primarily revolve around the foundational and technical aspects of BCT and its applications, align with the findings of existing literature reviews (e.g., Conoscenti et al., 2016; Yli-Huumo et al., 2016). These reviews highlight the prevalence of research focused on the technical foundations of blockchain (e.g., consensus mechanisms and cryptography) and its applications in areas such as supply chain management and security. The sustained interest in both the core technology and its practical implementations is evident.

Topics 1 (blockchain and cryptocurrency technologies) and 2 (healthcare, food supply and agriculture traceability) had a moderate number of papers, while Topics 3 (energy trade and electric power market) and 8 (deep learning for smart contract and image/video processing) had the least number of papers in which these topics were dominant. The identification of Topics 3 and 8 which are heavily engineering focused, adds a nuanced layer to existing reviews. While some studies touch on the technical aspects of blockchain in these areas (e.g., Andoni et al., 2019; Kuperberg, 2020; Hewa et al., 2021), our analysis explicitly highlights their growing importance as complements to foundational research and mainstream applications.

Interestingly, Topic 6 (multidimensional data fusion and biotech applications) was not dominant in any of the research papers. This lack of dominance could be attributed to several factors. For example, multidimensional data fusion and biotech applications are highly specialised areas of inquiry and may not form a cohesive thematic area that researchers are actively exploring. Moreover, when there is a lack of cohesiveness in the keywords of a topic, there might be a lack of interest in understanding the underlying phenomenon. Additionally, the dominance of this topic could be diffused across other topics due to a high degree of overlap.

Nevertheless, this lack of dominance may represent a novel finding and suggests a potential research gap in exploring the intersection of blockchain with complex data integration and biotechnology (Engelhardt, 2017). Therefore, Topic 6 might be an emerging area for future investigation, offering opportunities to bridge the gap between blockchain technology and these specialised fields.

2.5.2 Temporal and Evolutionary Analysis

Blockchain research has undergone a paradigm shift. Beyond its initial focus on cryptocurrencies and fundamental technology, topics like edge computing, data privacy, supply chain, and industrial research have witnessed exponential growth and transformed into the field's central pillars. This prolific research output signifies their pivotal role in shaping blockchain's future, having transcended from early application spaces to foundational pillars dominating the landscape since 2013-2014.

This evolution within core blockchain research areas further manifests itself in specific application trends, reflecting both societal needs and academic interests. For example, since 2020, topics like edge computing and vehicle network security have surged in research interest, surpassing previously prominent areas like data privacy and encryption, despite their earlier dominance and substantial research output. This shift exemplifies the broader trend of blockchain research expanding beyond its financial origins, as highlighted by Risius and Spohrer (2017). Two primary factors seem to drive this shift: societal/industrial needs and academic/technological trends.

Socially, the COVID-19 pandemic served as a catalyst for the acceleration of digital transformation, as widely acknowledged by Kudyba (2020). This increased our reliance on digital technologies, amplifying societal sensitivity to secure networks and systems. Edge computing and secure vehicle networks, which address these concerns directly, probably benefited from this increased awareness. The remote work culture further fuelled the need for a robust network infrastructure, which could contribute to the interest in these areas.

Academically, “hot topics” often attract significant research focus. Edge computing and vehicle network security may have emerged as such trends, drawing increased attention from the research community. Understanding these dynamic shifts within core blockchain research areas is crucial for predicting future directions and maximising the technology’s societal impact.

On the other hand, the distribution of contributions of topics in papers is another important anchor that indicates the interdisciplinarity “within-paper”, and the evolution of topics together with other topics. We found that the average contributions of the topics remained stable from 2019 onwards. However, some fluctuations were observed before 2019. For example, the contribution of topics centred on blockchain and cryptocurrency technologies decreased significantly between 2013 and 2017. Although the decline was less consistent after 2018, it remained low in 2018. The fluctuations observed prior to 2019 may highlight the dynamic nature of emerging technologies, which undergo phases of hype and subsequent critical evaluation (Dedehayir and Steinert, 2016).

The post-2018 decline in research contributions centred on blockchain and cryptocurrency technologies can be attributed to a multifaceted phenomenon beyond mere waning interest. This phenomenon, likely driven by the maturation of foundational concepts, witnessed a paradigm shift in research focus. The initial hype surrounding cryptocurrencies led to a more nuanced understanding, prompting investigators to dive deeper into the broader potential for blockchain application beyond its financial origins.

This shift is evident in the surge of research interest in areas such as Topic 4 (edge computing and vehicle network security), Topic 5 (data privacy, encryption, and secure access), and Topic 7 (digital supply chain and industrial research). These emerging areas can be seen as testing grounds for validating the practical utility of blockchain technology in diverse contexts. This transition from theoretical exploration to real-world application validation marks a crucial stage in the evolution of blockchain research.

Finally, while topics such as edge computing, data privacy, and digital supply chain flourished, others such as energy, multidimensional data fusion, and biotech applications saw comparatively lower research investment. The energy industry and deep learning experienced an initial rollercoaster ride due to changing regulatory landscapes, stabilising thereafter. In particular, multidimensional data fusion and biotech applications showed a downward trend within publications, suggesting waning interest. This divergence highlights the varying landscape of blockchain research. Topics 4, 5, and 7 received growing attention for their practical applications, while Topics 3, 6, and 8 faced potential roadblocks, prompting researchers to explore more fertile ground. Understanding these diverse trajectories is crucial for predicting future research directions and maximising blockchain's transformative potential across different sectors.

After analysing the temporal and evolutionary development of various blockchain topics, it is equally important to understand how blockchain topics interact and overlap across disciplinary boundaries. While evolutionary analysis offered insights into the changing fundamental aspects and application areas, we need to examine the "interdisciplinarity" of these topics. The rapidly evolving blockchain landscape is not limited to individual academic disciplines; it thrives at the intersection of different disciplines and fields of study. The aim of the subsequent section is to unravel the complex network of interdisciplinary relationships that characterise blockchain research. Therefore, we examine how different areas of knowledge synergise, how academic interests converge or diverge, and how thematic overlap occurs among disciplines. Our goal is to provide a nuanced understanding of the interdisciplinary nature of blockchain research and its implications for future research and applications through a multifaceted lens.

2.5.3 Interdisciplinary Dynamics

Blockchain technology is a complex system with inherent interdisciplinarity. Our results support the assumption that blockchain research reflects the interdisciplinary nature of blockchain at the topic level. Topic keywords can not only offer a nuanced understanding of research themes but also underscore the topic-keyword relationships among them. In our analysis, we found that keywords within the same topic had a stronger co-occurrence than those across different topics. The stronger synergies of within-topic (intra-topic) keywords as compared to across-topic (inter-topic) keywords can be attributed to a specialised research focus. By a specialised research focus, we mean that the topics are rooted in groups of specific disciplines or subdisciplines. The language used in such topics is less likely to crossover into other topics, reducing the co-occurrence of keywords outside a specific discipline group.

Furthermore, our analysis of the HHI of topic-discipline co-occurrence revealed a high concentration of “Computer Science”, “Linguistics & Communication”, and many “Other” specialised research disciplines in the majority of topics. On the other hand, disciplines such as “Arts & Humanities”, “Business, Economics, & Decision Sciences”, and “Engineering & Technology”, and “Life Sciences & Biomedicine” revealed moderate to low concentration or high diversity. Because a high HHI value represents specialisation and a low HHI value represents diversity, disciplines with high HHI values in a topic represent specialised topics for such disciplines or sub-disciplines. In contrast, a low HHI value represents low concentration and high diversity, implying a more interdisciplinary discipline or subdiscipline in that topic.

The concepts of modularity and balkanisation of science (isolation) (Mundale, 2002; Ackland and Halpin, 2019) offer a framework to understand the complex interdisciplinary landscape of blockchain research. Our analysis reveals both tendencies: high HHI values indicate specialised research within specific topics, while low HHI values suggest fragmentation into isolated groups. Recognising this dynamic is crucial for promoting a balanced research ecosystem that encourages both deep specialisation and interdisciplinary collaboration to maximise the potential of blockchain technology.

To better understand interdisciplinary dynamics, consider convergence and divergence in detail. Our analysis of richness and evenness revealed a promising trend for richness of Topics 2 (healthcare, food supply and agriculture traceability), 4 (edge computing and vehicle network security), and 7 (digital supply chain and industrial research), while Topics 2, 3 (energy trade and electric power market), and 7 appeared to be highly even. The richness of Topics 2, 4, and 7 indicates a high influx of disciplines contributing to them, whereas the evenness of Topics 2, 3, and 7 indicates balanced contributions from different disciplines. Interestingly, Topics 2 and 7 are rich and even at the same time. This can be attributed to their broader applicability, social impact, and multifaceted challenges that revolve around the keyword themes represented in these topics.

Although richness and evenness offer insight into the diversity and distribution of disciplines that contribute to individual topics, it is equally important to discuss thematic overlaps. These overlaps can shed light on the intersections between different areas of research, providing nuances about how the blockchain's interdisciplinary landscape is evolving.

We observed that all topics have a high overlap in "Computer Science", "Business, Economics, & Decision Sciences", and "Engineering & Technology". For example, Topic 7 has a large volume of research done in "Computer Science", Business, Economics, & Decision Sciences", and "Engineering & Technology". On the other hand, the overlap of Topic 3 and Topic 8 (deep learning for smart contract and image/video processing) is the least of all other topics with "Computer Science" as its dominant discipline.

Having examined the interdisciplinary nature and thematic overlaps in blockchain research, we now turn to the complexities and behavioural patterns that characterise this evolving field.

2.5.4 Research Complexity and Behaviour

Blockchain technology is multifaceted and complex. Therefore, the analysis of Research Complexity and Behaviour deserves investigation. By analysing the evolution of the academic landscape from a longitudinal perspective, we shed light on how the diversity of research topics has evolved over time, how the topics covered have evolved, and how authors have managed this complex and changing terrain. Our main goal was to assess the maturity and readiness of this field to address increasingly complex research questions.

We found that the HHI of topic contributions for most topics within a document showed a consistent downward trend over time, especially in 2019. This loss of concentration and increase in diversity within topic contributions indicate that new nuances are likely to be added, indicating that publications are incorporating a mix of methods or insights from multiple topics, a finding consistent with Yli-Huumo et al. (2016) who stated that blockchain research continues to expand into diverse areas, incorporating a mix of methods and insights from multiple disciplines.

Topics 2 (healthcare, food supply and agriculture traceability), 3 (energy trade and electric power market), and 7 (digital supply chain and industrial research) showed a low magnitude loss of HHI values between 2019 and 2022 and remained within a narrow HHI threshold (between 0.1 and 0.4). The magnitude of this downward trend remained relatively low. However, Topics 1 (blockchain and cryptocurrency technologies), 4 (edge computing and vehicle network security), 5 (data privacy, encryption, and secure access), and 8 (deep learning for smart contract and image/video processing) remained in a wider HHI threshold (between 0.3 and 0.8) in the year from 2019 to 2022. The magnitude of this downward trend remained relatively high. Therefore, we can conclude that the magnitude of diversity in Topics 2, 3, and 7 is higher than that in Topics 1, 4, 5, and 8. Consequently, for new research, Topics 2, 3, and 7 might be more suitable for contributions given a disciplinary background than Topics 1, 4, 5, and 8.

In addition to the diversity of topic contributions, the addition of unique disciplines also indicates the diversity of a field. We found a consistent and above-exponential increase in the number of unique disciplines contributing to the blockchain's research landscape. Between 2018 and 2020, we observed a massive positive trend. The increase in unique disciplines can be attributed to the maturity of the formative era of blockchain around 2019, when many new applications surfaced and, being the underlying technology of cryptocurrencies, gained significant attention from the research community. Typically, a steady increase in the number of unique disciplines, added to the overall landscape of blockchain research disciplines, suggests that blockchain research is likely to draw the attention of diverse experts and perspectives (Yli-Huumo et al., 2016), enriching academic discussion. The accelerating trend also indicates that the field is not only adding disciplines, but at an increasing rate.

Quantitatively, the disciplines of "Computer Science", "Business, Economics, & Decision Sciences", and "Engineering & Technology" has the highest number of research publications on different topics. This suggests that complexity is stagnant across a specific set of disciplines. Moreover, there is an exponential increase in polymaths which suggests a high number of versatile authors who might specialise in more than one discipline, contributing to a different set of topics. The high number of first authors or polymaths may be due to the ease of publishing in open-access online communities and research repositories.

The author versatility scores complement our rationale that the disciplines of "Computer Science", "Business, Economics, & Decision Sciences", and "Engineering & Technology" has the highest number of versatile authors who are highly likely to contribute to interdisciplinary research. Our author versatility scores can be regarded as topic-specific adaptation of Porter et al. (2007) concept of researcher interdisciplinarity. This adaptation is relevant in the BCT context, where interdisciplinarity can vary significantly across different topics. Topics 2 (healthcare, food supply and agriculture traceability), 3 (energy trade and electric power market), and 7 (digital supply chain and industrial research) have the highest diversity of first authors, where authors are involved in a variety of disciplines within those topics, suggesting a more interdisciplinary area for authors. In contrast, Topics 1 (blockchain and cryptocurrency technologies), 4 (edge computing and vehicle network security), 5 (data privacy, encryption, and secure access), and 8 (deep learning for smart contract and image/video processing) have comparatively low author diversity, suggesting that the first authors who contribute to these topics are often from similar disciplines, suggesting a more specialised or focused research area.

In addition to author versatility, the collaboration of multiple authors in a research project represents diversity, as the subject of the project could be diverse and requiring the expertise of different disciplinary authors. Consequently, our results suggest that the increase in the mean number of authors points to mushroom growth in collaborative projects. This trend of collaborativeness saw a tremendous increase in 2017 and beyond. This can be attributed to an increase in the number of cryptocurrency development-related projects which led to an increasing number of people working together to set the agenda for research and practice within the blockchain's entrepreneurial and research landscape.

As we explore the intricate relationship between topic diversity, disciplinary contributions, and collaborative trends in blockchain research, it is important to narrow our focus and examine specific areas that have received significant attention in this evolving field.

2.5.5 Research Focus and Concentration

Building on the extensive analysis of the blockchain research landscape in the previous sections, this section delves into the emergence of thematic clusters within the research. Moving beyond mere breadth and depth, we aimed to identify areas where topics showed significant concentration or diversity across disciplines. Our focus was on understanding how research energy and resources were distributed, determining whether we observed thematic dispersal or significant topic concentration. By examining these patterns of topical focus and disciplinary engagement, we gained deeper insight into the current research priorities and potential future directions of blockchain scholarship.

In our analysis, we observed that Topic 8 (deep learning for smart contract and image/video processing) had the highest HHI value and Topic 3 (energy trade and electric power market) had the lowest HHI value across disciplines. HHI is a metric used to measure market concentration and disciplinarity in the context of this study. Therefore, a high HHI value of Topic 8 signifies concentration or disciplinarity, while a low HHI of Topic 3 indicates low concentration (diversity) or interdisciplinarity. Therefore, Topics 8, 1 (blockchain and cryptocurrency technologies), 5 (data privacy, encryption, and secure access), 4 (edge computing and vehicle network security), 2 (healthcare, food supply and agriculture traceability), 7 (digital supply chain and industrial research), and 3 have decreasing concentration or increasing interdisciplinarity, respectively. Consequently, Topic 3 has a high number of disciplines that contribute to this topic, making it a fertile ground or focal area of research for interdisciplinary researchers.

The disciplines of "Computer Science", "Business, Economics, & Decision Sciences", and "Engineering & Technology" dominate specialised research on topics with high HHI values. Interestingly, Topic 3 is also dominated by the same disciplines in addition to "Earth and Environmental Sciences". This implies that Topic 3 is closely aligned with the core competencies of these disciplines. This could also mean that research on Topic 3 is deeply rooted in the technological, economic, and practical imperatives that drive blockchain innovation. Topic 2 is spread over a variety of disciplines that include "Social Sciences", "Physical Sciences", "Medical Sciences & Healthcare", "Mathematics", and in addition, it consists of the same disciplines as all other topics. Therefore, Topic 2 is an interdisciplinary topic that has gained traction in the majority disciplines. In the following, Topic 7 is another highly interdisciplinary topic spread across all disciplines in our study.

Overall, our analysis indicates that on average blockchain research topics have lost concentration across disciplines, as is evident from our final analysis of average HHI values from 2013 to 2022. It also indicates that, despite some bumps in concentration across a few years, the overall research landscape is growing, with different topics pointing to interdisciplinarity. There has been a slow but steady decline in concentration over the years, which suggests the democratisation of blockchain research and illustrates its evolution from a niche, discipline-specific area to a broad, interdisciplinary area of interest. The decreasing HHI values indicate a growing dialogue around blockchain, which brings in different perspectives, methods, and thematic focuses.

In summary, the temporal analysis of HHI values offers a dynamic view of the interdisciplinarity within blockchain research, revealing a clear shift from research initially concentrated within specific disciplines towards a broader, more collaborative, and interdisciplinary focus. This trend not only underlines the universal relevance of blockchain technology, but also clarifies the development of academic discourse in this area.

2.6 Conclusion

Blockchain, an emerging field characterised by its complexity and interdisciplinarity, offers numerous research opportunities. This study aimed to examine the current state of blockchain research, assess its level of interdisciplinarity, and provide an overview of all blockchain research, its implications, and avenues for future research.

Analysis of 35,604 publications revealed a continuously evolving research landscape in blockchain technology (BCT), broadly categorised into three interconnected areas: core technological foundations, interdisciplinary applications, and specialised and frontier domains. First, the "core technological foundations" category includes fundamental aspects like consensus algorithms and encryption, highlighted by a significant number of publications. Second, the "interdisciplinary applications" category includes topics that revolve around the practical implementation of BCT in various sectors, including healthcare, energy, and supply chain management. Topics in this specific category, such as "Healthcare, Food Supply, and Agriculture Traceability" and "Digital Supply Chain and Industrial Research," serve as prime examples of the diverse application of blockchain technology in various sectors.

Third, the "specialised and frontier domains" category includes emerging and highly specialised areas like edge computing for vehicle networks, deep learning, and smart contracts representing promising avenues for future advancements.

The study highlighted the prevalence of interdisciplinary approaches in blockchain research across various topics and disciplines. "Computer Science", "Business, Economics, & Decision Sciences", "Engineering & Technology", showed high interdisciplinary contributions, while "Social Sciences" and related fields showed moderate to low levels.

Our analysis has revealed a dynamic and evolving focus in blockchain research, with interconnected topics and strong keyword affinities. This diversity highlights the crucial role of interdisciplinary collaboration. Notably, many researchers serve as first authors in multiple disciplines, showcasing their ability to integrate diverse knowledge streams.

Furthermore, we identified at least three topics exhibiting significant richness and evenness, indicating a trend toward convergence. An increase in the average number of authors per article has been observed since 2013. This trend suggests that understanding the complexities and practical applications of blockchain is increasingly a collaborative effort rather than an individual endeavour.

Our research results offer insights into potential applications and diffusion of blockchain innovations, serving as a strategic guide for science, industry, and policymakers. Recognising blockchain's complex nature, this study creates a foundation for future investigations, promoting interdisciplinary collaboration crucial for the field's advancement.

This study contributes significantly to the current literature and acts as a catalyst for reevaluating research methodologies for multifaceted technologies like blockchain. Our objective was to unravel BCT's evolutionary, dynamic, and multifaceted properties using a novel analytical framework, IDEA. The future of blockchain technology hinges on interdisciplinary collaboration, innovation, and continuous exploration to ensure its successful and exponential development.

2.7 Contributions

This study stands out in the blockchain literature by producing a comprehensive knowledge map that transcends disciplinary boundaries and contrasts with previous discipline-specific research highlighted by scholars such as Risius and Spohrer (2017), Hawlitschek et al. (2018), and Wang et al. (2019). Our integrative approach uses the Latent Dirichlet Allocation (LDA) method in conjunction with our innovative IDEA framework and four analysis metrics. This unique combination expanded the existing understanding of blockchain and delved into its interdisciplinary aspects. We gained valuable insights into the current state of blockchain technology and quantified its interdisciplinary nature, thereby illuminating the complex and evolving dynamics of blockchain research. The IDEA framework, a cornerstone of our methodology, is crucial in assessing interdisciplinarity. It enabled a comprehensive presentation and analysis of the research landscape, covering topics, authors, keyword patterns, and the extent of interdisciplinary integration. Consequently, our study provides a novel quantitative approach and a more comprehensive and nuanced understanding of the interdisciplinary dimensions of blockchain.

Second, our research approach was improved by integrating four different analytical metrics into a unified framework. This strategy significantly expanded the scope of our methodology and extended its applicability beyond traditional disciplinary boundaries. Furthermore, a main objective of this study was to establish a novel connection between the HHI values and the concept of interdisciplinarity. In this way, we aimed to extend the use of HHI, traditionally used to assess market concentration in economics, to new applications for measuring interdisciplinary research dynamics (Rogers, 2010). This expansion not only diversifies the utility of the HHI, but also provides a new perspective for evaluating the interconnectedness of academic disciplines.

Third, our study offers important insights for funding agencies that support interdisciplinary research on blockchain technology. Research proposals are typically evaluated based on criteria such as significance, novelty, and subject matter relevance. Our results illustrate the interdisciplinary potential of various blockchain research topics and thus provide a benchmarking tool for evaluating proposals. These insights can help agencies identify projects with significant potential in subject matter and topic selection. This can significantly improve the promotion and further development of innovative, interdisciplinary blockchain research. Furthermore, organisations can benefit from integrating our findings into their decision-making processes, particularly when allocating research funding and developing entrepreneurial initiatives.

Fourth, our use of LDA and our focus on topics and keywords went beyond traditional limitations by examining the underlying interactions within the system and recognising blockchain as a convergence of complexity and determinism. In this way, we expanded the scope of research in this dynamic area by identifying specific themes and keywords that encompass different perspectives. This comprehensive understanding lays a foundation for future interdisciplinary research, enriching our knowledge of blockchain technology and complex systems.

Finally, our study identifies interdisciplinary research as a potential catalyst for innovation diffusion, as highlighted by (Rogers, 2010) in his work on diffusion of innovations. The HHI serves as a quantitative measure for assessing the concentration of interdisciplinarity in blockchain research topics over time. This finding highlights the importance of interdisciplinary research in facilitating the widespread adoption and integration of blockchain technology across various domains. This reflection encourages blockchain scholars to prioritise interdisciplinary knowledge creation and build upon existing research. It complements disciplinary works of (Risius and Spohrer, 2017) and (Yli-Huumo et al., 2016), providing an interdisciplinary foundation for advancing blockchain technology and fostering knowledge contributions across various fields.

In summary, this study significantly contributes to the blockchain literature by introducing a novel approach to analysing its interdisciplinary dynamics. Our findings underscore the critical role of interdisciplinarity in driving the growth and advancement of blockchain technology. We advance the understanding of blockchain as a complex, multi-layered system and demonstrate the effectiveness of innovative methods to analyse interdisciplinary knowledge. The unique methodology and results presented here offer a valuable framework for researchers, funders, and practitioners, providing insights into the opportunities and challenges associated with blockchain research and development. Ultimately, this work supports the field's ongoing growth and bring up its practical application across diverse domains.

2.8 Limitations

Although this study makes significant advances in understanding blockchain research, it is not without limitations. It should be noted that the main focus of this literature review was to understand the overall state of blockchain research by mapping knowledge across disciplines. As a result, certain prevalent keywords in the blockchain corpus, such as "bitcoin" and "blockchain," may be under-represented due to our exclusion criteria. The under-representation of these terms might introduce noise into our findings and potentially skew the identification of dominant themes. Future studies could focus on literature that utilises these very common keywords or draws on specific types of literature, as suggested by Glaser et al. (2014), Morisse (2015), and Tschorsch (2017).

This study is innovative because in its use of the HHI to measure concentration in blockchain research. However, it is important to note that the HHI has limitations. For example, it may not fully capture the depth and quality of interdisciplinary integration. Although we have included additional aspects in our IDEA framework to assess the depth and breadth of BCT research, future studies could explore the inclusion of additional metrics of interdisciplinarity that better capture the complexity of interdisciplinary engagement.

While LDA can help identify hidden thematic structures, it is primarily a data-driven model and may not always align with theoretical expectations or real-world complexities. Its effectiveness also significantly depends on the quality of the input data. Moreover, the dynamic nature of complex systems like blockchains means that our understanding may continue to evolve, and further research is needed to explore how the complexity of such systems change over time.

Despite these challenges, we believe that IDEA framework provides a more comprehensive view of the interdisciplinary nature of blockchain, overcoming the fragmentation and oversimplification typical of traditional methods. The ability of LDA to identify interconnected themes and relationships within the data aligns with the evolving understanding of complex systems as a dynamic phenomenon, offering a more nuanced and holistic perspective.

Finally, although the methodology of this study is adequate to present our research, it is not exhaustive. Future research could assess the suitability of our methods for interdisciplinary research and suggest improved strategies for conducting interdisciplinary literature reviews. Another major limitation is the unavailability of a comprehensive framework in existing research for assessing interdisciplinary outcomes. This suggests that future work could leverage our IDEA framework to further refine it for assessing interdisciplinary research, thereby increasing the value and accuracy of studies such as ours. Despite these limitations, the primary objective of this study – to identify micro and macro indicators of interdisciplinarity in blockchain research – has been achieved.

2.9 Future Work

Blockchain technology is an interdisciplinary subject that requires a comprehensive understanding of the technology within its respective fields of study. Understanding blockchain by drawing upon reference disciplines has the potential to generate novel insights and address the theoretical obstacles presented by disciplinary boundaries. In this regard, the research conducted by Culnan (1986) can serve as a valuable addition to our existing framework for identifying reference disciplines that are relevant to the holistic advancement of blockchain technology. This study serves as an initial exploration in this area of research by synthesising the existing literature and pinpointing pertinent research areas. The dataset used in this study can be extended to incorporate data from supplementary publications in subsequent investigations.

This study presents preliminary indicators of the interdisciplinary nature of knowledge related to blockchain technology, both at the micro- and macro-level. The methods and techniques employed to extract knowledge from the literature are exploratory in nature and have potential for further enhancement. The absence of a well-defined framework or established methodology for conducting interdisciplinary research prompts an inquiry into the feasibility and analysis of such research from various disciplinary perspectives. Therefore, future research endeavours should focus on the advancement of techniques and methodologies to facilitate interdisciplinary research collaboration.

Finally, blockchain technology is undergoing continuous development, and the body of literature on this subject is expanding at an exponential rate. Currently, the academic landscape is witnessing the rapid development of novel concepts and techniques. This progress is likely to inspire future researchers to replicate our study, providing insights into evolving research patterns and contributing to the advancement of blockchain technology through these novel concepts and techniques (Wang et al., 2019; Hawlitschek et al., 2018).

Chapter 3

Standing Out to Be Seen: Legitimacy through Distinctiveness in Initial Coin Offerings

3.1 Introduction

The attainment of organisational legitimacy is a critical task for entrepreneurial ventures, as it directly affects their ability to attract vital resources (Lounsbury and Glynn, 2001; Zimmerman and Zeitz, 2002; Taeuscher et al., 2020) from resource-providing audiences,[1] such as investors and stakeholders. Legitimacy, the broad acknowledgement that an organisation's actions align with socially constructed norms and expectations (Suchman, 1995, p.574),[2] significantly shapes its ability to succeed. This acknowledgement may be based on the organisation's established track record or on the active endorsement of external parties. However, gaining this

[1]The term *resource-providing audiences* refers to investors and stakeholders with financial, non-financial, or intellectual stakes in a company.

[2]The literature presents two key perspectives on legitimacy, *legitimacy as an inherent quality* and *legitimacy as an external endorsement. Legitimacy as an inherent quality* sees legitimacy as an internal characteristic of the organisation, much like an "organisational property" (Suchman, 1995). It emphasises factors such as a long track record, established practices, and adherence to widely accepted norms as a source of legitimacy. *Legitimacy as an external endorsement* frames legitimacy as a dynamic process, contingent on the approval of external audiences. Here, legitimacy arises from active validation (Deephouse, 1996) or positive judgements made by individual stakeholders (Tost, 2011). This highlights the importance of perception and how audiences shape an organisation's legitimacy.

legitimacy poses a significant hurdle for early-stage entrepreneurial ventures (ESEVs), which often lack a proven history to establish their credibility. This absence of traditional sources of legitimacy underscores the need for ESEVs to actively cultivate their legitimacy and secure support for continued growth and development.

The challenge of establishing legitimacy is particularly acute for ESEVs due to their newness to the market (Navis and Glynn, 2011) and the rapidly evolving online media landscape. Having no proven track record or established product history (Fisher et al., 2016), they may be perceived as mere ideas rather than viable businesses (Gartner et al., 2016). This absence of traditional legitimacy markers makes it crucial for ESEVs to actively cultivate their legitimacy and attract support, as legitimacy directly affects their ability to secure funding, build strategic partnerships, and ultimately achieve long-term success.

Cultural entrepreneurship theory emphasises the transformative significance of entrepreneurial storytelling (EST)[3] in shaping how ESEVs gain legitimacy. A core principle of this theory is the concept of optimal distinctiveness[4] – the idea that a successful EST must strike a balance between being distinctive enough to capture attention and familiar enough to be understood (Lounsbury and Glynn, 2001, 2019; Martens et al., 2007; Navis and Glynn, 2010, 2011).[5] However, institutional theory has persistently underscored the potential drawbacks of distinctiveness in the search for legitimacy (Powell and DiMaggio, 2012). The underlying argument lies in the notion that an overly distinct entrepreneurial narrative can complicate rather than facilitate comprehension among audiences. By straying too far from established categorical prototypes, a venture's story may fail to resonate with familiar cognitive frameworks, thus hindering the audience's ability to readily associate with it. This deficit in understandability might erode a venture's legitimacy, as resource-providing audiences are less likely to invest in ventures that they cannot easily understand.

The complex dynamics of online media introduces a new layer to the relationship

[3]The term *entrepreneurial storytelling* or *EST* is sometimes also referred to as *entrepreneurial narratives* or *entrepreneurial stories (ESTs)*.

[4]The term *optimal distinctiveness* is sometimes also referred to as *distinctiveness*.

[5]Here, *understandability* refers to how well can the entrepreneurial venture be understood, rather than how well the entrepreneurial venture can understand.

between distinctiveness, legitimacy, and resource acquisition. While some research has presented a link between EST and legitimacy (e.g., Martens et al., 2007; Snihur, 2016; Taeuscher et al., 2020) in various funding contexts, this relationship is likely more complex in the context of dynamic online funding environments. We propose a nuanced perspective on the prevailing assumption that distinctiveness undermines legitimacy, particularly when considering the moderating role of online media. Traditional institutional theory views legitimacy as multifaceted (Aldrich and Fiol, 1994; Scott, 2014; Suchman, 1995). However, research on optimal distinctiveness has often focused on cognitive legitimacy or how easily an organisation is understood (Suchman, 1995). The distinctiveness also influences the normative legitimacy, reflecting the alignment with the expectations of the audience (Suchman, 1995). Given the variation in normative expectations among ESEV audiences (Fisher et al., 2017), we argue that distinctiveness can improve the legitimacy of ESEVs, with online media as a moderator, in the eyes of novelty-seeking audiences. In other words, emerging ESEVs might gain legitimacy specifically through their distinctiveness, attracting support from those who value innovation and uniqueness. This nuanced perspective significantly changes the distinctiveness-resource acquisition relationship for these audiences.

Furthermore, the dynamics of online media, where diverse audience segments have influence, amplify this nuanced view of distinctiveness. It offers a two-pronged mechanism that functions as either a tool for "legitimisation" or "delegitimisation", facilitating the acquisition or shift of legitimacy. Media narratives can endorse ventures by shaping stakeholder perceptions about the desirability and appropriateness of organisational actions and attributes (Elsbach, 1993; Zuckerman, 1999), and conversely. While organisation theorists have long investigated the role of "traditional media" from a strategic point of view as a "costly" legitimation device for established organisations, we propose that "online media" has gained popularity among entrepreneurs as a cost-effective and two-way communication medium that can serve to legitimise and delegitimise ventures. Discursive evaluations of users of online media, particularly the silent majority (Xie et al., 2020), can influence existing perceptions of the legitimacy of a venture by activating a legitimising or delegitimising frame (Humphreys and Latour, 2013), shifting cognitive legitimacy through implicit associations with a legitimate or illegitimate area of activity. Consequently, we believe that online media can positively or negatively affect normative legitimacy by shifting cognitive legitimacy. In that respect, we propose the dichotomous role of online media, which serves as a strategic device to manage legitimacy through venture-managed online activity, as well as a feedback system for the venture (or its offerings) residing outside the strategic control of the venture.

We employ these theoretical propositions to examine how ESEVs secure resources in the unique context of Initial Coin Offerings (ICOs), a crucial resource acquisition mechanism for these ventures (Fisch, 2019; Short et al., 2017). ICOs, which rely heavily on online discourse and community building, hold significant appeal for ESEVs, often surpassing traditional venture capital or angel investor funding routes.[6] Furthermore, ICOs often bypass traditional gatekeepers such as regulatory bodies, increasing the significance of the role of online media in moderating the relationship between distinctiveness and legitimacy, particularly given the novelty-seeking nature of ICO investors. We propose that distinctiveness offers competitive and normative advantages that outweigh potential cognitive disadvantages. Consequently, we diverge from established optimal distinctiveness models and hypothesise a strong positive relationship between distinctiveness and resource acquisition in the ICO context. Furthermore, we contend that distinctiveness' impact on legitimacy depends on the presence of alternative avenues of normative legitimacy. The absence of such alternatives amplifies the benefits of distinctiveness for ESEVs.

This study has two objectives. First, we offer a new theoretical perspective on the assumption that the acquisition of legitimacy is counteracted by distinctiveness, as reported by institutional theorists (Aldrich and Fiol, 1994; Scott, 2014; Suchman, 1995). Second, we consider the online media as a dichotomous instrument that acts both as an agenda-setter and a legitimacy-shifter.

To test our hypothesis, we analyse data from 306 ICOs in 29 market categories. We use Latent Dirichlet Allocation (LDA) (Blei et al., 2003; Hannigan et al., 2019), a topic modelling approach, to identify patterns in the content of ESEVs' ESTs in

[6] Underscoring this, CB Insights reports that in 2018, ICOs raised $12.62 billion for new companies, exceeding the $4.152 billion obtained through venture capital (CB Insights, 2019).

ICOs, subsequently quantifying the extent to which any particular EST diverges from the prototypical story in its market category. Our findings indicate that distinctive ESTs have a significant positive effect on the acquisition of resources from ICO audiences. Furthermore, we discovered that distinctive ESTs have an effect on ICO underpricing,[7] corroborating our proposal that distinctiveness can function as a basis for normative legitimacy and therefore helps – instead of impeding – the acquisition of legitimacy from novelty-seeking audiences. Finally, we confirm the role of online media in influencing the legitimacy as a mediator and a moderator.

This study advances discourse in several domains: optimal distinctiveness within organisational theory and strategic management (Barlow et al., 2019; Deephouse, 1996; Haans, 2019; Zhao et al., 2017, 2018; Zuckerman, 1999), the role of online media (Albrecht et al., 2019; Chanson et al., 2018; Etter et al., 2018), and the theory of cultural entrepreneurship (Lounsbury and Glynn, 2001, 2019; Navis and Glynn, 2011). We make four distinct contributions. First, by delineating the cognitive and normative effects of ESTs, we offer a nuanced perspective on how distinctiveness shapes different dimensions of legitimacy within the ICO context. Second, we demonstrate that the presence of novelty-seeking audiences and the dynamics of online media fundamentally reshape the traditional optimal distinctiveness trade-off. Our results demonstrate that the online media influence demand side performance by intensifying or reducing the effects of distinctiveness (Barlow et al., 2019; Haans, 2019). Third, we identify two key contingencies that determine the optimal degree of distinctiveness: 1) the strength and alignment of a venture's claims to environmental and social impact, and 2) the market category's perceived viability. Our results suggest that ventures with strong legitimating claims achieve greater success with greater distinctiveness, even in less established markets. On the contrary, ventures in high-potential markets, but with weaker legitimating claims, may benefit from a more moderate level of distinctiveness. Finally, we clarify the multifaceted role of the online media as both a mediator and a moderator in the distinctiveness-resource acquisition relationship. These theoretical and empirical insights significantly improve our understanding of the conditions under which organisations should strategically

[7] The entrepreneurial finance literature has employed IPO (Initial Public Offering) underpricing as a proxy for legitimacy. Likewise, we use ICO underpricing as a proxy for entrepreneurial venture's legitimacy.

leverage different levels of distinctiveness.

The structure of this study is as follows. Section 3.2 focuses on the theoretical background of entrepreneurial storytelling, its relationship with distinctiveness, and the interaction of the proposed theoretical constructs – online media and market category viability – with the relationship between entrepreneurial storytelling and distinctiveness. In addition, we discuss the mechanics of ICOs and demonstrate how underpricing can be considered as a sign of legitimacy. Furthermore, Section 3.3 explains the context and presents testable hypotheses. Section 3.4 presents the methodology and data. We discuss the preliminary results in Section 3.5 and important limitations of this study in Section 3.6.

3.2 Background Literature

3.2.1 Legitimacy

In his seminal article, Suchman (1995, p.574) offers a comprehensive definition of legitimacy, describing it as "a generalised perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions." Importantly, the attribution of legitimacy is a function of observers' perceptions; thus, legitimacy can vary between the observers (Lamin and Zaheer, 2012) and exists as both an objective state and a subjective social construction (Suchman, 1995, p.574).

Legitimacy is a multifaceted concept in organisational studies, and scholars have developed various perspectives to understand its nature and implications. The three major perspectives of legitimacy include legitimacy-as-property, legitimacy-as-process, and legitimacy-as-perception. Each perspective offers a different lens through which to examine how organisations achieve and maintain legitimacy.

3.2.1.1 Legitimacy Perspectives

To further understand the concept of legitimacy, it is important to examine the three major perspectives through which it has been explored: legitimacy-as-property, legitimacy-as-process, and legitimacy-as-perception.

Legitimacy-as-property: This perspective views legitimacy as an inherent trait or asset of an organisation, treating it as something that an organisation possesses (Suddaby et al., 2017, p.458). Research in this domain often examines the specific attributes or signals that make an entity appear legitimate to its audience.

Legitimacy-as-process: Grounded in a social constructivist approach (Light et al., 1967), this perspective understands legitimacy as a dynamic, ongoing process. Here, the focus is on the actions and strategies through which organisations seek affiliation with and acceptance within an established social order or category (Suddaby et al., 2017, p.462).

Legitimacy-as-perception: This perspective integrates elements of both the legitimacy-as-property and legitimacy-as-process while focusing on the audience's role in conferring legitimacy. It highlights legitimacy as a multilevel social process that begins with the perceptions of evaluators, progresses to their judgements, and culminates in actions based on those judgements (Suddaby et al., 2017), which have macro-level repercussions on the organisation. This approach underscores the importance of audience evaluations and the broader social context in understanding legitimacy.

3.2.1.2 Types of Legitimacy

Legitimacy is not a monolithic concept, rather it encompasses different forms that operate at various levels and within distinct social contexts.

Socio-political Legitimacy: This form of legitimacy refers to the extent to which an entity's actions are deemed acceptable and appropriate within the broader socio-political environment. It is closely tied to regulatory acceptance and alignment with societal norms. In the context of Initial Coin Offerings (ICOs), socio-political legitimacy might involve adherence to legal frameworks (where they exist) and demonstration of social responsibility.

Cognitive Legitimacy: This refers to the degree of public awareness and understanding surrounding a new venture or its offerings (Aldrich and Fiol, 1994). For Early Stage Entrepreneurial Ventures (ESEVs) and ICOs, cognitive legitimacy is crucial as it addresses the novelty and often complex nature of their products or services (Shepherd and Zacharakis, 2003). A lack of cognitive legitimacy can create barriers to adoption and investment due to uncertainty and unfamiliarity.

Normative Legitimacy: This type of legitimacy is based on an entity's adherence to moral and ethical standards. It reflects whether an organisation is perceived as "doing the right thing." In the ICO context, normative legitimacy might encompass ethical business practices, transparent communication, and a commitment to broader societal goals beyond profit.

Scholars have identified several potential sources of legitimacy, which Deephouse and Suchman (2012) categorises as society-at-large, relationships within the organisational field, and the media. Research in journalism and mass communication indicates that the media serves not only as a barometer of existing legitimacy, but also as an independent force capable of conferring it (Deephouse, 1996). More recently, Lundmark et al. (2017) explored the influence of social media on the development and management of legitimacy. This work demonstrated that organisational microblogging, specifically through Twitter, can confer legitimacy even prior to the execution of Initial Public Offerings (IPOs).

The acquisition and management of legitimacy have increased significance under conditions of uncertainty and complexity (Kostova and Zaheer, 1999). Investing in ESEVs that leverage nascent technologies, such as blockchain, and novel fundraising mechanisms, such as ICOs, carries substantial risk and unpredictability. Consequently, it is plausible that investors' perceptions of organisational legitimacy significantly impact ICO pricing. Although research specifically addressing the legitimacy of ICOs remains scarce, studies on legitimacy within traditional capital markets provide insight. Scholars have identified external authorities such as security analysts (Zuckerman, 1999), prestigious underwriters, and prominent media as legitimacy influencers (Pollock and Rindova, 2003). Although likely relevant to ICOs, the existing landscape of such external authorities is notably limited compared to established capital markets. The absence of a robust industry made up of professional analysts, underwriters, and media outlets diminishes the readily available sources of legitimacy signals for observers. However, within the ICO context, stakeholders (project initiators, investors, influencers, developers, etc.) maintain a highly active online presence, a domain recognised by Lundmark et al. (2017) as a potential legitimacy source. This research shifts the focus from the well-established literature on organisational legitimacy in IPOs by established entities to investigate legitimacy effects surrounding new ventures utilising ICOs.

3.2.2 Market Category Viability

Entrepreneurial ventures often find promising opportunities in emerging business environments (Santos and Eisenhardt, 2009). However, these new environments, despite their potential, are inherently characterised by uncertainty. Technological advancements, products, and processes remain unproven and subject to change (Tushman and Anderson, 1986); product definitions may lack clarity or precision (Hargadon and Douglas, 2001); and nascent market categories often present considerable ambiguity or structural fluidity (Santos and Eisenhardt, 2005). This environmental uncertainty is further amplified for ESEVs within new market spaces, as they typically possess limited resources and a still-developing organisational identity, thus facing additional challenges in establishing legitimacy. Market categories, defined as "structures of economic exchange between producers and consumers imbued with meaning by interacting participants (Kennedy et al., 2010)," play a crucial role. Due to the potential for divergent interpretations among stakeholders, the configuration of market categories takes on heightened importance. These categories continue to evolve, shaped by the contributions of entrepreneurs, prospective resource providers, media, and other key stakeholders (Tripsas, 2009). Following Lo et al. (2020), this research adopts a prescriptive approach to define the value configuration of the market category, recognising the resource acquisition process as fundamentally composed of resource seekers (entrepreneurs) and resource providers (audiences).

The existing literature presents two contrasting perspectives on the construction of the meaning of the market category: the socio-cognitive perspective (Mervis and Rosch, 1981; Porac and Rosa, 1996) and the relational perspective (Peirce, 1993; Scott et al., 2015). The socio-cognitive approach emphasises category boundaries defined by the homogeneity or internal consistency of members. In contrast, the relational perspective focuses on the position of a category within a broader system of meaning, stressing its relative "place" compared to other categories (Somers, 1995). Synthesising these views, categories can be redefined as cognitive constructs that facilitate comprehension by grouping similar entities (Glynn and Navis, 2013; Rosa et al., 1999). Categorisation based on similarities (category coherence) and differences (category distinctiveness) becomes a tool to navigate and organise complex environments (Durand et al., 2017). This dynamic is called "market category viability" (Lo et al., 2020), which highlights that a viable market category balances membership complementarity (intra-category relationships) with its role within the larger classification structure (inter-category relationships). Intra- and inter-category relationships form two spatial dimensions that define a "zone of viability", marked by an equilibrium between coherence and distinctiveness. Therefore, this research conceptualises category viability based on the two central constructs of category coherence and category distinctiveness.

Category coherence addresses the internal relationships between the entities that make up the membership of a category. These entities encompass individuals, organisations, products, and practices, etc. Their inclusion within the category is governed by rules, logic, or similarity to prototypes defined by shared characteristics (Rosch, 1983). Cognitively, strong category resemblance and the presence of distinct prototypes simplify the identification and maintenance of category boundaries. This ease reduces the cognitive stress on the audience, which promotes cognitive legitimacy (Hannan et al., 2012; Negro et al., 2011). For example, Kuilman and Wezel (2013) demonstrated that during their emergent phase, greater category coherence lowered mortality rates among UK airlines. Moreover, in line with the principle "do the right thing" central to normative legitimacy, a category's focus on financial or non-financial motives (whether economic, social, or ecological) might exist beyond its core constituents. Thus, a viable category strikes a balance of coherence among its members. Excessive homogeneity or diversity within a category diminishes its viability and, consequently, legitimacy in the eyes of audiences receptive to novelty.

Category distinctiveness refers to the relative position of a category within a larger classification system. A highly distinct category exhibits minimal overlap with other categories within that system. Previous research underscores the context-dependent nature of category meaning; relational sociology, for example, posits that the meaning of an entity arises from its position within a broader structure (Ruef, 2000; Scott et al., 2015; Somers, 1995). Applying this relational approach to categorisation, scholars have studied how connections between categories (established through co-mentioning (Kennedy, 2008) or analogical reasoning (Bingham and Kahl, 2013)) shape boundaries and meanings. Distinctiveness can also be considered from the audience's perspective, measured by the degree to which users perceive a focal category as unique or interchangeable with others based on its function. For a category to achieve viability, it must strike a balance between distinctiveness and alignment with existing category systems. Categories exhibiting insufficient or excessive distinctiveness tend towards lower viability. This balance between coherence and distinctiveness defines the "zone of viability", as illustrated in Figure 3.1; deviations from this zone lead to diminished viability.

Figure 3.1 presents a visual model with category coherence plotted on the horizontal axis and distinctiveness on the vertical axis. The central circle denotes the "zone of viability", where these dimensions are balanced. Deviations from this zone, marked by excessive or insufficient coherence and/or distinctiveness, jeopardise market category viability. Categories within the zone are more likely to provide value, remain prominent, and experience active use. This "zone" model conceptualises viability as a continuum rather than a binary state, acknowledging varying degrees of (non)viability.

As illustrated, each corner represents a unique threat to category viability, highlighting the interdependence of coherence and distinctiveness. These threats – attenuation, fragmentation, absorption, and isolation – emerge when categories fall outside the balanced "zone". The category within the "zone of viability" can be referred to as a "viable market category".

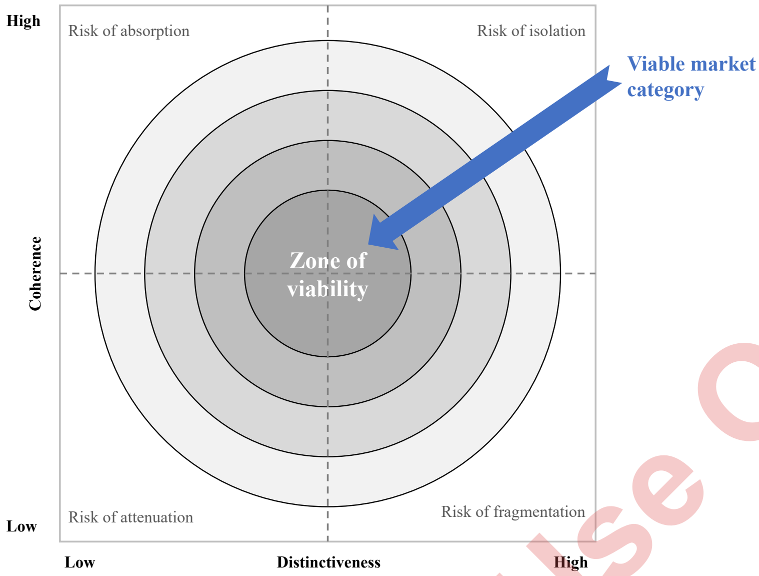

Figure 3.1: Zone of viability (Adapted from Lo et al. (2020))

3.2.2.0.1 Risk of Attenuation: Categories positioned low in coherence and distinctiveness – exhibiting excessive ambiguity and a lack of unique features – fall into the risk of attenuation. This arises from two primary factors. First, the inclusion of various members lacking shared characteristics creates perceptual confusion, undermining the category's inherent value (Rosa et al., 1999). Second, when a category fails to differentiate itself adequately from others, its appeal and utility are diminished. Audiences struggle to distinguish the unique offerings or practices of the category within the broader meaning system (e.g., Hannan, 2010).

3.2.2.0.2 Risk of Fragmentation: This threat arises when a category achieves strong distinctiveness within the broader meaning system, but suffers from internal heterogeneity. Such a combination can create instability and fragmentation. Distinctiveness, while reducing redundancy, requires greater interpretive effort from audiences attempting to reconcile the category with existing classification systems. This task becomes even more complex when the category exhibits internal incoherence, fostering debate and contestation among members and users. This dynamic may lead them to develop competing interpretations of meaning or practice, resulting in subdivision – a vertical shift in which the focal category fragments or spawns variants (Kennedy et al., 2010). If these alternatives gain prominence or utility over the original, the viability of the focal category will likely decline.

3.2.2.0.3 Risk of Absorption: This threat describes the potential for a focal category to be subsumed within a higher-level or better-positioned (i.e., more balanced) alternative. Specifically, a category exhibiting both high internal coherence and low distinctiveness may simply be perceived as a minor variant of an existing category. Its lack of unique features renders it redundant, increasing the likelihood of absorption. Under these conditions, the category may be deemed overly specific and, thus, lacking in utility compared to the broader, better-positioned category. Consequently, its viability may decline.

3.2.2.0.4 Risk of Isolation: This threat applies to categories that demonstrate both excessive coherence and distinctiveness. This imbalance can lead to what Lo et al. (2020) terms "isolation," or the risk of being perceived as marginal, irrelevant, or suspect. Although previous research generally highlights the positive aspects of coherence and clear boundaries (e.g., Hannan et al., 2012; Negro et al., 2011), category viability requires a delicate equilibrium. Extreme coherence and distinctiveness can hinder a category's ability to function as a "boundary object," a concept that values interpretive flexibility to facilitate interactions between diverse groups (Star and Griesemer, 1989). A category that is too rigid in its definition and is isolated from broader frameworks will likely struggle to achieve or maintain widespread awareness and utility.

Recalling the highly uncertain environment in which ICOs function and their limited sources of legitimacy, it follows that categories placed within the "zone of viability" will be perceived as having enhanced cognitive and normative legitimacy. On the contrary, ICOs belonging to categories with a higher potential for failure due to imbalances in coherence and distinctiveness will likely be seen as riskier and less legitimate. Given that category viability is ultimately shaped by audience perceptions, this study posits that novelty-seeking audiences will actively avoid ICO categories associated with elevated risk.

3.2.2.1 Legitimacy and Role of (Online) Media

The literature on organisational studies has long recognised the crucial role of the media as an information intermediary (or "infomediary") with the power to influence market behaviour (e.g., Pollock and Rindova, 2003; Ross et al., 2021). Media functions as a repository of public knowledge and sentiment, not only directing attention to specific issues, but actively shaping the agenda of public discourse (Rogers et al., 1993). Furthermore, through "persistent patterns of cognition, interpretation and presentation, as well as selection, emphasis, and exclusion (Levitt and Gitlin, 1981, p.7)," media framing provides a context for evaluating actions based on institutional and cultural norms (Elsbach, 1993). Therefore, the media's actions – informing, highlighting, and framing – provide market participants with cues that directly impact impression formation and, consequently, the legitimation of firms.

Although organisational theorists recognise the central role of the media in shaping legitimacy, precise mechanisms remain contested (Pollock and Rindova, 2003). Two research perspectives dominate: one frames the media as an expert monitor that facilitates market transactions (Biglaiser, 1993), while the other views it as a strategic tool used by organisations to manipulate public perception (Elsbach, 1993; Zuckerman, 1999). A key limitation of the traditional media's unidirectional communication model is its vulnerability to manipulated information, allowing organisations to construct an artificial facade of legitimacy. However, the ascendance of online media has fundamentally altered this dynamic. Its inherent bidirectionality enables audiences to critically engage with and evaluate organisational claims. This introduces complex theoretical and methodological challenges to understanding the full impact of online media on organisational legitimacy.

The online media plays a dichotomous role in shaping legitimacy, functioning as both a mediator and a moderator. As mediator, firms exert influence through strategic online activity aimed at increasing public exposure and framing narratives. This extends the influence over how stakeholders interpret and evaluate organisational information. However, the online media also acts as a moderator, where the role of the audience emerges independently of the organisation's control. Here, legitimacy is impacted by factors such as discursive evaluations, emotional expressions, and taken-for-granted beliefs.

Online media empower ordinary citizens to actively evaluate organisations, express opinions, and shape a larger consensus (Castelló et al., 2013). Platforms like Twitter facilitate these discussions, with evaluative discourse categorised by linguistic markers (for example, #ICOs) (Albu and Etter, 2016). This discourse represents a social process in which individuals express subjective opinions to forge an "intersubjective consensus of values" (Hart, 2014, p.43). This dynamic is crucial to understanding legitimacy, a "generalised" perception, not merely individual viewpoints. Moreover, online judgements significantly impact legitimacy construction, as information from sources such as Twitter is considered credible by ordinary citizens seeking to gauge the suitability of organisational actions (Castelló et al., 2013; Westerman et al., 2014; Sotiriadis and van Zyl, 2013). The influence, perceived credibility, and sheer volume of these online judgements make social media a key driving force in the co-construction of organisational legitimacy.

Individuals actively contribute to legitimation or delegitimation through communication (for example, text and talk), influencing others, and forging shared understandings that culminate in aggregate legitimacy judgements (Bitektine and Haack, 2015; Johnson et al., 2006; Tost, 2011). Evaluation can be defined as "substantive perceptions and beliefs that underlie judgements of an entity's legitimacy, often classified as instrumental (pragmatic), relational, and moral (Tost, 2011, p.687)." Unlike cognitive legitimacy, often viewed as static and taken-for-granted (Suchman, 1995; Suddaby et al., 2017), normative legitimacy arises from discursive evaluations (Golant and Sillince, 2007; Suchman, 1995). As these evaluations can change audience beliefs through rational argumentation, we argue that cognitive legitimacy can, in fact, be influenced by changes in normative legitimacy. This proposition challenges the widespread assumption of cognitive legitimacy as an inflexible construct.

Unlike the rational and measured evaluations typical of institutional evaluators (e.g., experts commenting on television), online communities express opinions with greater emotional valence (Deephouse and Carter, 2005). This allows semi-private expressions of strong feelings, such as outrage or praise, often expressed in highly individualistic styles (Papacharissi, 2009; Perelló-Sobrepere, 2017). Although institutional evaluators adhere to professional norms (Shoemaker and Reese, 2013) and may offer limited space for the full spectrum of public sentiment, online communities rely on affect-based responses (Haack et al., 2014). Joy (positive affect) or disappointment (negative affect), triggered by perceptual input (e.g., claims of an organisation), form the basis for individual judgement (Haidt, 2001). These judgements, expressed as sentiments about the organisation, may even diverge from rational assessments or reflect herd behaviour. Accordingly, we argue that positive sentiment functions as a legitimising force, while negative sentiment contributes to delegitimisation. Thus, online expressions of judgement actively shape organisational legitimacy.

3.2.3 The Theory of Cultural Entrepreneurship

3.2.3.1 Entrepreneurial Storytelling and Optimal Distinctiveness

Cultural entrepreneurship theory positions entrepreneurial storytelling (EST) as a transformative force that enables new ventures to establish identity, acquire legitimacy within their markets, and attract the essential resources to fuel their growth (Lounsbury and Glynn, 2001; Snihur, 2016). ESTs commonly refer to deliberately constructed narratives about the venture and usually encompass elements such as founder values ("We value users who trust us to keep their money safe."), business models ("Our model combines commission fees with a free marketplace."), market opportunities ("Demand for blockchain-based business transitions is growing rapidly."), underlying technology ("Our product uses Ethereum's Plasma Cash Protocol."), and sustainability commitments ("Our product mitigates the environmental effects of typical crypto mining.") (refer to Section D.1 to see excerpts of ICO stories). Given that new ventures lack access to traditional sources of legitimacy, these entrepreneurial stories become paramount for external evaluations (Lounsbury and

Glynn, 2001).

The concepts of optimal distinctiveness (Lounsbury and Glynn, 2001; Martens et al., 2007) and legitimate distinctiveness (Navis and Glynn, 2011) form a cornerstone of the theory of cultural entrepreneurship, highlighting the dynamic interaction between a venture's unique identity and its perceived legitimacy within the market. This proposition, which addresses the balance between differentiation and legitimacy (Deephouse, 1999; Zhao et al., 2017), emphasises how distinctiveness within ESTs influences new ventures. Since these ventures often lack access to traditional sources of differentiation, a distinct EST becomes crucial to establish a market presence in a market category (Navis and Glynn, 2011). However, excessive distinctiveness hinders cognitive legitimacy, as audiences struggle to comprehend the essence of an ESEV (Aldrich and Fiol, 1994).

Resource providing audiences face significant uncertainty when evaluating ESEVs: they lack familiarity with these new ventures and their uncertain environment (Davis et al., 2009, p.420). Previous research shows that this uncertainty hinders cognitive legitimacy for such ventures among various audience groups (Martens et al., 2007; Pontikes, 2012). New ESEVs gain popularity when aligned with category prototypes, allowing audiences to grasp the unfamiliar concept more easily (Suchman, 1995). For example, the research of Elsbach and Kramer (2003) revealed a key factor in the success of Hollywood film pitches: a clear connection to established category prototypes. As Elsbach and Kramer (2003) found, matching an EST with a known category prototype increases favourability as it becomes immediately identifiable (Bitektine, 2011). This cognitive association reduces ambiguity (Lo et al., 2020; Lounsbury and Glynn, 2001; Navis and Glynn, 2011), whereas an EST significantly deviating from the prototypes raises concerns about future viability. Therefore, prototype similarity strengthens cognitive legitimacy, while divergence undermines it (Glynn and Navis, 2013; Navis and Glynn, 2011, 2010; Wry et al., 2014). According to theoretical frameworks, ESTs are poised to secure optimal resource allocation when they achieve an equilibrium between the competitive advantages derived from differentiation and the cognitive disadvantages associated with reduced legitimacy due to excessive distinctiveness.

3.2.3.2 Entrepreneurial Storytelling as a Pathway to Normative Legitimacy

Entrepreneurial storytelling can serve as a mechanism for establishing normative legitimacy. Normative legitimacy, as defined within the scholarly discourse, is related to an organisation's alignment with the normative expectations that are prevalent within its institutional environment. This legitimacy is based on the perception that new companies engage in activities deemed as the "appropriate" or "do the right thing" in their social context. This concept is elaborated on by Suchman (1995) as moral legitimacy and by Aldrich and Fiol (1994) as socio-political legitimacy, highlighting its multifaceted nature. The perception of normative legitimacy is promoted when stakeholders recognise that the actions of a venture are in accordance with societal norms and values.

The theoretical framework of cultural entrepreneurship, as articulated by Lounsbury and Glynn (2001), posits normative legitimacy as a crucial aspect of EST. For these narratives to be effective, they must resonate with the interests and norms of the target audience, facilitating a favourable interpretation of the venture. ESTs play a critical role in enhancing the perceived normative legitimacy of a venture by demonstrating its alignment with the audience's normative expectations.

The effectiveness of an entrepreneurial narrative in achieving normative legitimacy is contingent on its ability to meet the audience's expectations regarding what constitutes "doing the right thing". Within the domain of organisational finance, the notion of "doing the right thing" is contextualised within a social framework where the congruence between the actions of a firm and the expectations of society is assessed based on the contribution of the action to societal well-being.

This study reinterprets the concept of "doing the right thing" within the context of ICOs as encompassing suitability across economic, social, and ecological dimensions. It posits that in the realm of ICOs, which is characterised by boundaryless markets[8] and uncertainty, audiences may adjust their legitimacy expectations by considering the cumulative benefits across these three vectors. Ventures that demonstrate their

[8] A boundaryless market can be characterised by lack of geographic limitations and regulatory ambiguity.

commitment to societal well-being across multiple dimensions are posited to signal a stronger reduction in information asymmetry compared to those focusing on a singular aspect.

The economic vector is concerned with the financial benefits accruing to investors; the social vector addresses contributions to societal well-being; and the ecological vector focuses on environmental benefits. This multidimensional approach to evaluating "doing the right thing" underscores the importance of aligning entrepreneurial actions with economic, social, and ecological objectives to foster normative legitimacy in uncertain and globally orientated markets.

Audiences possess diverse normative expectations regarding appropriate behaviour. These expectations are influenced by economic, social, and ecological preferences, which ultimately shape how audiences assess legitimacy (Fisher et al., 2017). While audience expectations vary, we focus specifically on novelty expectations. Consequently, novelty expectations can be defined as "the degree to which an audience anticipates the offering and business model of a venture to demonstrate originality or uniqueness." Audiences with strong novelty expectations (that is, novelty-seeking audiences) are more inclined to perceive a venture as legitimate when they consider it to be innovative.

We posit that ventures accrue greater normative legitimacy among novelty-seeking audiences when their EST differs from the market category's prototypical story. As new ventures often lack demonstrable indicators of uniqueness (Rindova et al., 2007), a distinctive EST becomes a vital yardstick by which audiences gauge novelty. Consequently, for new ventures targeting novelty-seeking audiences, distinctiveness becomes a powerful source of legitimacy. While distinctiveness may present challenges to cognitive legitimacy (i.e., how easily an audience can understand the venture), it simultaneously enhances normative legitimacy by signalling alignment with the values of audiences who embrace innovation and difference. When normative gains surpass cognitive costs, distinctiveness strengthens a venture's overall legitimacy. This insight directly challenges the notion that distinctiveness invariably undermines legitimacy; instead, we assert that distinctiveness is a contributing factor to legitimacy. We challenge the prevalent belief that new ventures must make a binary choice between differentiation and legitimisation, instead proposing that distinctiveness acts as a crucial catalyst for resource acquisition through normative, cognitive, and differentiation mechanisms. The nature of the legitimacy-resource relationship depends on the specific context and the relative influence of these three mechanisms.

3.2.4 Startup Finance

3.2.4.1 Initial Public Offerings (IPOs)

Initial public offerings (IPOs) provide a common pathway for companies to raise the financial capital crucial for growth and competitiveness. The IPO process commences with the company, typically along with its underwriter(s), filing a registration statement with the regulatory body overseeing IPOs (e.g., the Securities and Exchange Commission (SEC) in the United States). Once the IPO is formally registered, the issuing company embarks on a roadshow, a series of presentations aimed at potential institutional investors. These presentations serve to share information about the company and evaluate investor interest. Based on this interest, institutional investors submit limit orders outlining both the number of shares they are interested in and their maximum purchase price. These data, combined with the prevailing market conditions, guide the issuing company and its underwriter(s) in setting the offer price and share allocation. The offer price represents what institutional investors will pay for the shares immediately before the IPO (e.g., Bigelow et al., 2014; Ljungqvist et al., 2006). On the day of the IPO launch, institutional investors, now owning shares in the issuing company, release a portion on the stock exchange. The opening price marks the debut trading price of the shares, whereas the closing price reflects the value upon the first day's market close. Figure 3.2 shows an outline of the IPO process.

3.2.4.1.1 Legitimacy in IPOs: IPOs are inherently marked by information asymmetry, with potential investors operating with significantly less knowledge than the issuing company (Carter and Manaster, 1990; Leland and Pyle, 1977; Ross, 1977;

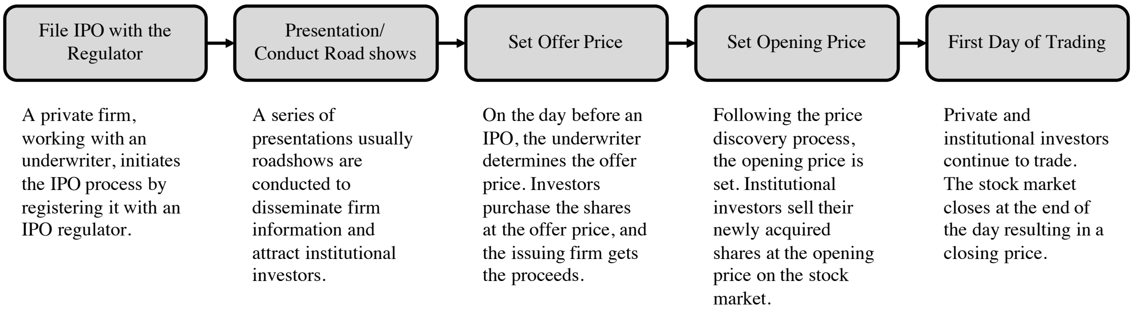

Figure 3.2: IPO process

Stuart et al., 1999). Issuers have deep insight into the company's operations, financial prospects, and the capabilities of its management and employees (Leland and Pyle, 1977). Meanwhile, public-market investors generally have limited information on IPOs (Carter and Manaster, 1990; Ross, 1977; Stuart et al., 1999). This disparity is further amplified by potential conflicts of interest that could lead to less than optimal information disclosures (Downes and Heinkel, 1982; Leland and Pyle, 1977; Ross, 1977).

To address information asymmetry and establish legitimacy in the eyes of investors, companies resort to costly and observable signals (Connelly et al., 2011). Examples include substantial media advertising investments, comprehensive CSR (Corporate Social Responsibility) reporting in environmental (Cho and Patten, 2007; O'Donovan, 2002) and social domains (Patten, 1992), affiliations with reputable venture capital firms (Gulati and Higgins, 2003), or strategic alliances with established multinational corporations (Ozmel et al., 2013). Investors perceive these actions as costly visible signals of legitimacy (Díez-Martín et al., 2020).

3.2.4.1.2 IPOs Underpricing: Underpricing measures the first-day return on a newly issued stock, the percentage gain (or loss) on its public market debut. This initial return is typically calculated as the percentage change relative to the offer price (for example, the closing price minus the offer price). Within the management literature, a higher return is generally viewed as undesirable (Arthurs et al., 2008, 2009; Pollock and Rindova, 2003). This signifies lost potential or suboptimal capital raised during the IPO. This perspective often hinges on the agency theory assumption of misaligned incentives between the issuing firm and underwriters who initially price the offering. Underwriters might favour a lower offer price to benefit institutional investors, creating underpricing and shifting potential gains away from the issuer. As Loughran and Ritter (2002) demonstrate, companies appear to deliberately increase underpricing as a strategic move.

Finance scholars offer several potential explanations for this trend. Cliff and Denis (2004) link increased underpricing to unofficial compensation for post-IPO analyst coverage. Welch (1989) suggests that reputable companies can deliberately underprice their stock to signal market quality and capitalise on a subsequent offer. On the contrary, Aggarwal et al. (2002) posit that strategic underpricing generates informational momentum, favourably shifting the demand curve for the firm's stock. Regardless of the specific rationale, there exists a connection between perceived organisational legitimacy and IPO underpricing. Firms deemed more legitimate (i.e., appropriate, proper, desirable) will likely experience higher underpricing levels, reflected in their first-day closing price.

3.2.4.2 Initial Coin Offerings (ICOs)

Initial Coin Offerings (ICOs), or token sales, represent a decentralised fundraising mechanism where blockchain-based tokens are issued to the public. These tokens function as entries on a digital ledger (the blockchain), which ensures a secure and transparent record of all cryptocurrency transactions. There are three distinct token categories:

Utility Tokens: The most prevalent type, utility tokens grant holders the right to access a future product or service developed by the issuing entity. They generally do not convey ownership rights and, due to a lack of widespread regulation, have become the most common instrument of ICOs. Their hybrid nature, blending payment and investment aspects, creates a compelling focus for research.

Security Tokens: These tokens often incorporate voting rights and, based on compliance with the Howey test,[9] fall under securities laws (Momtaz, 2020). Security tokens represented a small share (approximately 3%) of ICOs until the end of 2020.

Cryptocurrency Tokens: Acting as a value store or a medium of exchange, cryptocurrency tokens are exemplified by widely recognised assets such as Bitcoin and Ether.

Figure 3.3 outlines the core dynamics of an ICO. Before launch, the investment-seeking venture establishes two key smart contracts on its ICO blockchain. These contracts define the ICO's crucial parameters and govern the tokens to be issued. Parameters include the maximum funding target (hard cap), ICO duration, project-specific coin price, and the maximum allowed token creation. Once deployed, these smart contracts facilitate investor participation. Investors transfer capital to the ICO smart contract, a transaction distinct from direct payment to the project itself. Upon the investor's payment, the remaining process adheres strictly to the predetermined rules within the smart contracts. The project accesses its collected funds from the ICO smart contract, while the token smart contract automatically distributes tokens to investors. Thus, the central exchange of capital for tokens, the heart of the ICO process, operates as a fully automated, blockchain-driven system, solidifying its classification as an Information Technology (IT) artefact.

The following sections discuss the concept of legitimacy within the context of ICOs. We examine underpricing as a proxy for legitimacy, draw comparisons between ICOs and traditional financing mechanisms such as IPOs, and investigate the theoretical significance of underpricing dynamics in both IPOs and ICOs.

3.2.4.2.1 Legitimacy in ICOs: ICOs present an environment of increased information asymmetry compared to IPOs, where costly and observable signals serve to mitigate risk and uncertainty. IPOs generally operate under strict registration requirements imposed by regulatory bodies (Ofir and Sadeh, 2019). In contrast, ICOs often lack standardised registration processes, a dynamic that can vary based

[9]The Howey Test is a legal framework established by the United States of America. It's used to determine whether a transaction qualifies as an "investment contract" and, therefore, falls under the regulations of securities law.

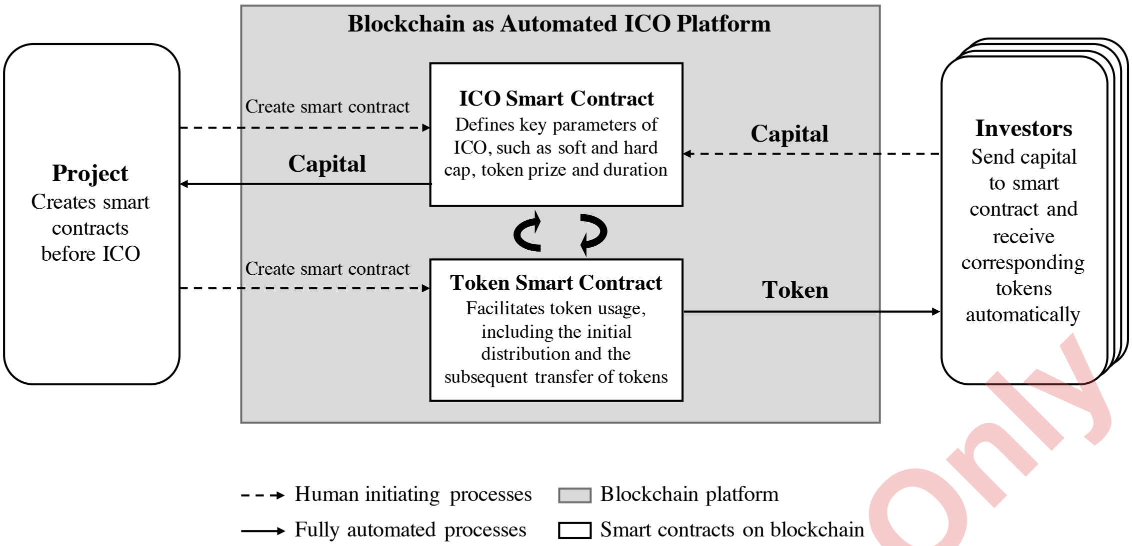

Figure 3.3: ICO process (Adapted from Chanson et al. (2018))

on their function and jurisdiction. This regulatory landscape makes it particularly attractive for ESEVs that offer nascent technologies, unfinished products, and face an untested market to pursue ICO funding; such ventures often lack tangible evidence to demonstrate their quality (Murray and Marriott, 1998; Nagy et al., 2012). Access to the type of costly, observable signals that establish track records is severely limited or non-existent. In this context, absent traditional credibility markers (Connelly et al., 2011), shaping audience perceptions – a vital component of legitimacy – becomes crucial for ESEVs using ICOs. Their focus shifts to reducing information asymmetry and projecting the venture's appropriateness to potential resource-providing audiences.

This dynamic, where legitimacy concerns become paramount in environments with high uncertainty, aligns with current scholarly understandings of *legitimacy-as-perception*. This viewpoint draws on social psychology and microsociology. It integrates elements of both *legitimacy-as-property* and *legitimacy-as-process* while expanding the scope to incorporate multiple levels and institutional audiences. Following this perspective, legitimacy unfolds as a "multilevel social process that begins with the perceptions of the evaluators of a legitimacy object, progresses to their judgements about it, and finally to their actions based on that judgement, which has macro-level repercussions on the object (Suddaby et al., 2017, p.468)."

While aligning with the view of legitimacy as a dynamic process and organisational goal, the focus changes subtly. Emphasis is placed on the audiences from whom legitimacy emerges, with legitimacy itself framed as a favourable assessment rather than an inherent property.

The existing literature examines the legitimacy of IPOs through the lenses of "legitimacy-as-property" and "legitimacy-as-process." However, the unique dynamics of ICOs suggests that these perspectives could have less explanatory power in this context. ICOs often lack the traditional hallmarks that signal legitimacy in IPOs. For example, while IPOs can provide tangible indicators of product standards (costly signals mirroring *legitimacy-as-property*) and verifiable performance (observable signals aligning with *legitimacy-as-process*), ICOs might offer limited evidence of either product quality or performance. These circumstances, coupled with the increasing reliance on online communities in the startup funding space, underscore the growing importance of the *legitimacy-as-perception* perspective in ICOs. Therefore, the distinctions between the IPO and ICO landscapes suggest a potentially weaker relevance of traditional legitimacy frameworks for understanding ICOs.

3.2.4.2.2 ICOs Underpricing: Underpricing, a well-documented phenomenon in IPOs, is also prevalent in ICOs. Following the methodology of Chanson et al. (2018), we calculate underpricing as the difference between the coin issuance price and its closing price after a workweek on at least one public exchange.[10] While the mechanics of price setting differ between IPOs and ICOs, the underlying rationale for underpricing shares some commonalities. In both cases, underpricing can act as a signal of quality, a mechanism to attract investors, and a means to secure long-term benefits.

The rationale behind IPO underpricing is complex. One viewpoint holds that issuing companies view underpricing unfavourably, as it implies lost potential revenue. This is often attributed to misaligned incentives arising from the underwriter's role in price setting (Ritter and Welch, 2002). On the contrary, several theories posit that

[10] While synonymous with ICO investor return, the term "underpricing" aligns with established terminology in management and finance literature.

underpricing serves long-term issuer interests. These include its use as a quality signal to secure beneficial terms in future financing rounds (Allen and Faulhaber, 1989), the pursuit of widespread ownership (Booth and Chua, 1996), generating post-listing demand (Aggarwal et al., 2002), and incentivising future analyst coverage (Sackley, 2005; Cliff and Denis, 2004).

However, the dynamics of ICO pricing differ significantly from those of IPOs. A small founding team determines the issuance price based largely on projected expectations, enjoying considerable valuation flexibility due to the project's early stage and lack of a historical track record. Furthermore, since ICOs leverage the blockchain for direct access to the capital market, there is no intermediary underwriter influencing the process. Additionally, the founding team typically retains a substantial portion of project-specific coins (often 10-30%) and controls a project endowment for future funding needs. Taken together, these factors lead us to posit that underpricing in ICOs is likely to prove beneficial for the issuing project in the long term. Therefore, we argue that ventures deemed more legitimate will experience higher levels of underpricing. We expect that this outcome will reflect the perceptions of the audience during the initial five-day trading period, mirroring observations made by scholars studying IPO underpricing (e.g., Lundmark et al., 2017).

While opportunistic behaviour may occasionally confound this relationship, the overall pattern of underpricing in legitimate projects should outweigh these isolated instances. The initial five-day trading period following the ICO launch is particularly crucial, as it reflects the market's early assessment of the project's legitimacy, mirroring observations made in IPO underpricing literature (Lundmark et al., 2017). Higher levels of underpricing during this period suggest a stronger belief in the project's long-term prospects, thereby serving as a proxy for investor judgement.

3.2.4.3 Comparisons Between ICOs and IPOs

ICOs diverge significantly from traditional financing methods, such as IPOs, resulting in different legitimacy dynamics. This section compares IPOs and ICOs across four key dimensions: 1) startup or firm characteristics, 2) investor characteristics, 3) deal setup characteristics, and 4) post-deal characteristics. Table 3.1 provides a summary of these distinctions.[11]

Table 3.1: Difference between IPOs and ICOs

Characteristic	IPO	ICO	Explanation
		Startup or firm characteristics	
Funding stage	After funding stage	Theoretically all stages	IPOs align with specific funding stages. On the contrary, ICOs offer theoretical applicability across all funding stages, though entrepreneurial firms primarily use them for capital acquisition.
Investment amounts	Greater than \$10m	Greater than \$100k	IPOs traditionally cover funding amounts up to \$10M, while successful ICOs demonstrate a far wider range, from \$100K to a remarkable \$4.2B (as of July 2018).
Issuance	Equity shares	Utility tokens, security tokens, or cryptocurrencies	IPO investors receive equity shares in the company. In contrast, ICOs can issue a variety of assets: equity shares (security tokens), rights to future products or services (utility tokens), or mediums of exchange (cryptocurrency tokens).
Organisation type; Track record	Established; High	Early-stage; Low	IPOs are typically associated with established organisations with long track records, while ICOs are characteristic of early-stage companies with limited operational history.
Information disbursement channel	Traditional media/Online media	Online media	IPOs generally use traditional media (e.g., newspapers, television) for information disbursement. ICOs use online media to circulate and disseminate information.

Continued on next page

[11] For a more detailed comparison of ICOs and conventional funding, see Benedetti and Kostovetsky (2021) and Blaseg (2018).

Table 3.1 – continued from previous page

Characteristic	IPO	ICO	Explanation
Product/service degree of maturity	High	Low	IPOs usually have a working product as compare to ICOs that usually do not have a working product.
Investor characteristics			
Aimed at	Institutional and private investors	Primarily crypto community; All types	While cryptocurrencies may represent a growing asset class for mainstream investors, ICOs currently remain primarily focused on attracting participants within the existing crypto community.
Motivation	Financial	Financial and non-financial	While financial motives primarily drive IPO investors, ICO investors often prioritise a blend of financial and non-financial incentives (e.g., altruism, product interest, social/ecological concerns, and the desire to provide feedback).
Deal characteristics			
Transaction costs	High	Low	Transaction costs of IPOs are high while ICOs have close-to-zero transaction costs.
Information basis	IPO prospectus	Whitepaper	IPOs prospectus is a legal document and follows stringent standards; provides significant information about the business, company, plans, and track record. ICOs whitepaper is not a legal document with no prescribed standard; provides information about business, team, token distribution, and proposed technical solution.

Continued on next page

Table 3.1 – continued from previous page

Characteristic	IPO	ICO	Explanation
Degree of regulation	Strong	Weak	IPOs operate within a comprehensive regulatory framework overseen by national agencies like the Securities and Exchange Commission. In contrast, ICOs operate largely within a self-regulated environment.
Legal backing	Yes	Generally not	ICO-related contracts often lack robust regulatory oversight and may not be legally enforceable by traditional courts of law. In some cases, the platform that issues the ICO may serve as the primary arbitrator.
Reporting	Strong	Weak	Weak reporting requirements contribute to substantial variation in transparency levels among ICOs.
Duration to setup	4-5 months (Approx.)	1-month (Approx.)	The rigorous regulatory processes involved in IPOs make them significantly more time-consuming to initiate compared to ICOs.
Underwriter	Yes	Rare, but pre-ICOs	Underwriting is a standard practice for IPOs, providing price stability and risk mitigation. In contrast, ICOs lack this intermediary process, with the issuing firm directly determining the initial token price.
Strategic aim	Exit and company development	Entry of a specific project	IPOs offer established companies both exit options for owners and development capital. In contrast, ICOs are frequently used to initiate new projects.

Continued on next page

Table 3.1 – continued from previous page

Characteristic	**IPO**	**ICO**	**Explanation**
Currency	Fiat	Crypto coins	IPOs involve purchasing stock with traditional fiat currency. In contrast, ICOs utilise cryptocurrencies or tokens as the exchange medium, typically acquired using existing crypto assets.
Fraud	Rare	Rather often	The requirement for companies to demonstrate a positive track record for IPOs acts as a safeguard against potential fraud. In contrast, the current lack of robust ICO regulation increases the risk of fraudulent projects.
Risk	Relatively low	High	ICOs carry significant risks due to factors such as a lack of regulation, non-existent track records, weak reporting requirements, the potential for fraud, and pronounced herding behaviour among investors.
		Post-deal characteristics	
Exchange	Listed on one (or some) exchange(s)	Easily listed on (many) trading platforms	Dual listings for IPOs are determined by the issuing company at or following the IPO. In contrast, trading platforms ultimately decide whether to list an ICO project. Additionally, IPOs adhere to the operating hours of their respective stock exchanges, while ICO trading occurs around the clock.
Liquidity	High	High (if listed)	IPOs have stricter regulations for soft or hard caps for fund accrual.
Trading	During stock exchange opening times	Continuously	IPOs are confined to traditional stock exchange operating hours. However, ICOs enjoy continuous trading without restrictions such as weekends or bank holidays.

Continued on next page

Table 3.1 – continued from previous page

Characteristic	IPO	ICO	Explanation
Voting rights	Well defined rights, like dividends and voting	Security tokens: Yes; Utility tokens and Cryptocurrencies: No; Exchange medium, access to services or products, dividends and/or other rights	While IPO rights broadly align with a company's overall performance across projects, ICO contracts offer diverse rights tied to the success of specific initiatives.
Exit options	Open market	ICO, open market	IPOs may offer open market exits, but achieving these often depends on company maturity. However, ICOs provide the earliest open market exit potential. Token liquidity facilitates a timely exit at any point after listing, addressing a common constraint of IPOs.

3.2.4.3.1 Startup or Firm Characteristics: IPOs and ICOs serve fundamentally different purposes within the entrepreneurial funding landscape. Unlike IPOs, which are geared towards established businesses seeking substantial growth capital, ICOs theoretically find application at any funding stage. This flexibility is evident in the diverse range of capital successfully raised through ICOs – from $100K to $4.2 billion (as of July 2018, see Momtaz (2018)). Another crucial distinction lies in the nature of the assets offered. IPOs grant shareholders residual rights linked to differing perceptions of legitimacy (Chan and Makino, 2007). ICOs, in contrast, issue tokens representing equity (security tokens), product/service access rights (utility tokens), or mediums of exchange (cryptocurrency tokens). Early-stage ventures' typical reliance on ICOs inherently creates a gap in their track record, prompting audiences to evaluate legitimacy based on alternative signals. On the contrary, IPO-bound companies have verifiable track records (Pahwa, 2018) that foster credibility, reduce information asymmetry, and thus facilitate legitimacy conferral.

3.2.4.3.2 Investor Characteristics: ICOs attract a diverse investor pool, from early adopters and altruistic individuals to institutional investors, while IPOs traditionally appeal to more sophisticated investors. Investor motivations also differ. IPO investors primarily prioritise financial returns. In contrast, ICO investors often combine financial incentives with non-financial motivations, such as aligning with the principles "do the right thing," altruism, product interest, and the opportunity to provide feedback (see Lipusch, 2018). Strict regulatory and broker requirements in the IPO space function as signals of quality and legitimacy. However, due to the largely unregulated nature of ICOs, participation remains open to anyone with Internet access, regardless of background or experience.

3.2.4.3.3 Deal Characteristics: ICOs have exploded in popularity largely due to near-zero transaction costs, minimal documentation, and a lack of regulation similar to crowdfunding campaigns. However, they enable ventures to potentially raise capital amounts rivalling those of heavily regulated IPOs. Interestingly, during the first half of 2018, the largest ICO, ranked among the top three global IPOs, in terms of funding (see Howell et al., 2019). It has even exceeded the total amount raised on Kickstarter since its founding in 2009 (Fisch, 2019). While intermediaries such as underwriters and accountants serve to mitigate risk in IPOs, their participation remains rare in ICOs (Cerezo SSnchez, 2018). Consequently, ICOs carry significantly higher risk than IPOs. This is consistent with the observations that cryptocurrency investors generally exhibit greater risk tolerance than traditional stock market participants (Lee et al., 2018).

3.2.4.3.4 Post-deal Characteristics: The liquidity offered by ICOs is a major draw for investors. Unlike IPOs, which do not provide immediate liquidity, tokens from ICOs often gain exposure on round-the-clock trading platforms within three months of the ICO's closure. This is consistent with liquidity discount theories (e.g., Officer, 2007), which posit that a lack of liquidity negatively affects the value. ICOs also innovate by offering flexibility in conveying voting rights depending on the token type. However, perhaps the most revolutionary aspect of ICOs lies in their exit mechanisms. IPOs generally require reaching a mature stage and have longer preparation timelines, hindering fast exits. On the contrary, ICOs offer the earliest potential exit among financing methods. This is achieved by delegating future development to a decentralised network, often preprototyping. And even though projects typically retain a portion of tokens, token liquidity ensures the potential for quick exits whenever tokens are listed.

3.2.4.4 Empirical Studies of ICO and IPO Underpricing

Three central research streams approach ICO underpricing through the lens of IPO theory. First, both IPOs and ICOs exemplify the "lemons" problem (Akerlof, 1970), where information asymmetry favours the seller (startup) over the buyer (investor). This asymmetry is particularly severe in ICOs, where lack of regulation increases the potential for fraud (Felix and von Eije, 2019). Second, underpricing might be a deliberate issuer strategy to mitigate information disparities (e.g., Allen and Faulhaber, 1989). Furthermore, Welch (1992) theorises that sequential share sales allow future investors to glean insights from previous purchases. Therefore, issuers may underprice initial offerings to signal quality and potentially reap higher prices in subsequent offerings.

A third perspective focuses on how the specific characteristics of the ICO project could dictate the levels of underpricing. Early studies in this area (Adhami et al., 2018; Momtaz, 2018) sought to use these characteristics to predict ICO success, but did not directly connect their findings to the established IPO literature. Subsequently, further research was conducted on associations between ICO variables and underpricing (e.g., Felix and von Eije, 2019; Howell et al., 2019; Lyandres et al., 2019). Using a sample of underpriced token sales, Felix and von Eije (2019) established a significant negative correlation between underpricing and issue size (amount raised). Additionally, they linked underpricing to market sentiment and initial listing trading volume, both liquidity indicators. These conclusions are corroborated by Momtaz (2018) and Momtaz (2019), who employed different measures of issue size.

However, Felix and von Eije (2019)'s findings diverged from some previous work. They did not observe a significant link between underpricing and the percentage of tokens retained by the project or whether the fundraising target was reached, a connection supported by Chanson et al. (2018). Furthermore, while Benedetti and Kostovetsky (2021) found that the offer price and the pre-ICO negatively affected the underpricing, Felix and von Eije (2019) only confirmed the latter as a determinant. Lastly, the effect of ICO underpricing regulation remains largely unexplored. However, IPO research offers potential connections: Shi et al. (2013) found cross-country evidence suggesting that security law disclosure requirements may negatively impact IPO underpricing. Overall, signalling theory and project characteristics inform the current scholarly understanding of underpricing in IPOs and ICOs.

3.3 Research Context and Hypotheses

This study tests theoretical assumptions employing ICO underpricing as a proxy for the legitimacy conferred on early-stage entrepreneurial ventures (ESEVs). Our theoretical model (illustrated in Figure 3.4) provides a comprehensive overview of these relationships.

Previous research suggests that perceived uniqueness and innovation drive investor interest in ICOs (Calic and Mosakowski, 2016). This aligns with social media sentiment analysis, where users emphasise the appeal of blockchain-based projects that offer novel solutions. For example, a user says that "I see ICOs as unique blockchain products that you will not find anywhere else." Statements that highlight

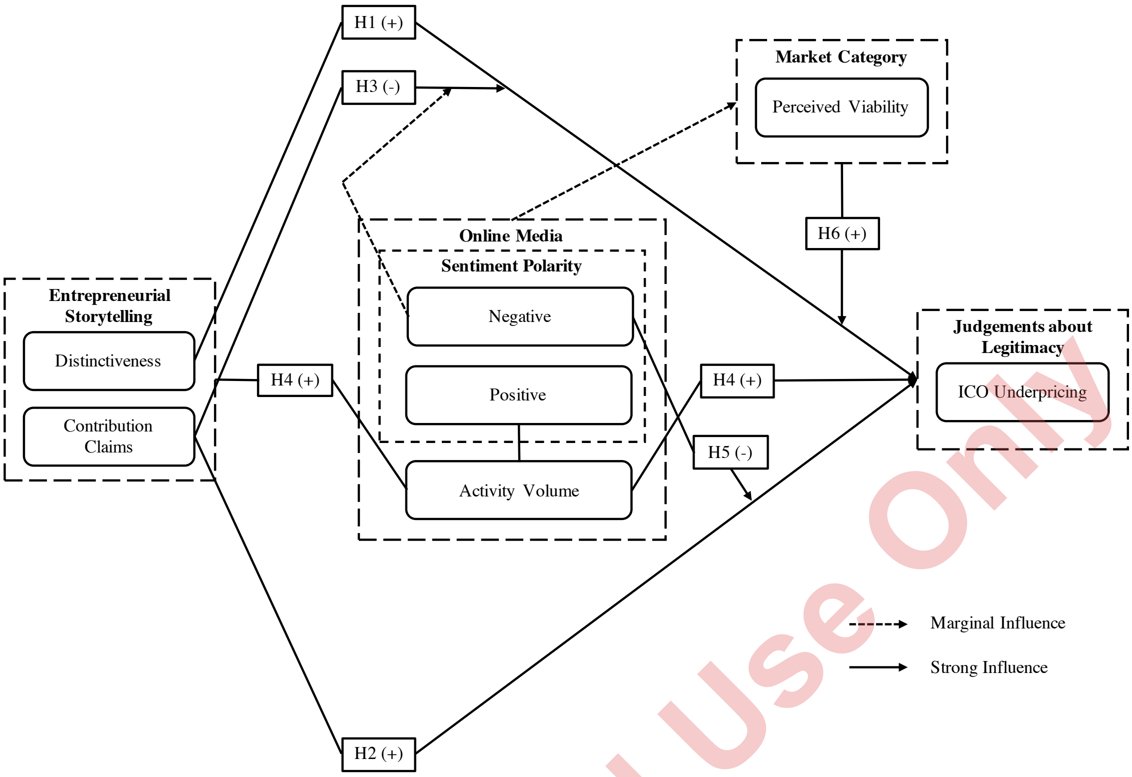

Figure 3.4: Theoretical model. H: Hypothesis; +: Positive relationship; -: Negative relationship

the desire to be among early investors in groundbreaking projects underscore this novelty-seeking behaviour. For example, a user says, "I love being able to invest quickly in new projects provided by blockchain technology and being able to claim that I was one of the first to invest in ICO projects." Further evidence comes from the practices of ICO listing websites such as ICObench (`https://icobench.com`). By prioritising ICOs with novel products or services, they reinforce investor expectations of finding distinctive offerings. These collective findings characterise the ICO investor community as fundamentally novelty-driven.

ICO ventures function within a dynamic and fluctuating landscape where stakeholders seek validation from reputable sources to gauge an ICO's legitimacy. Adapting the framework established by Salamanca et al. (2017), we delineate six primary sources of legitimacy and demonstrate how they inform our theoretical constructs for a legitimacy-focused analysis. Figure 3.5 illustrates these sources: justice, participation and deliberation, transparency, accountability, coherence, and effectiveness.

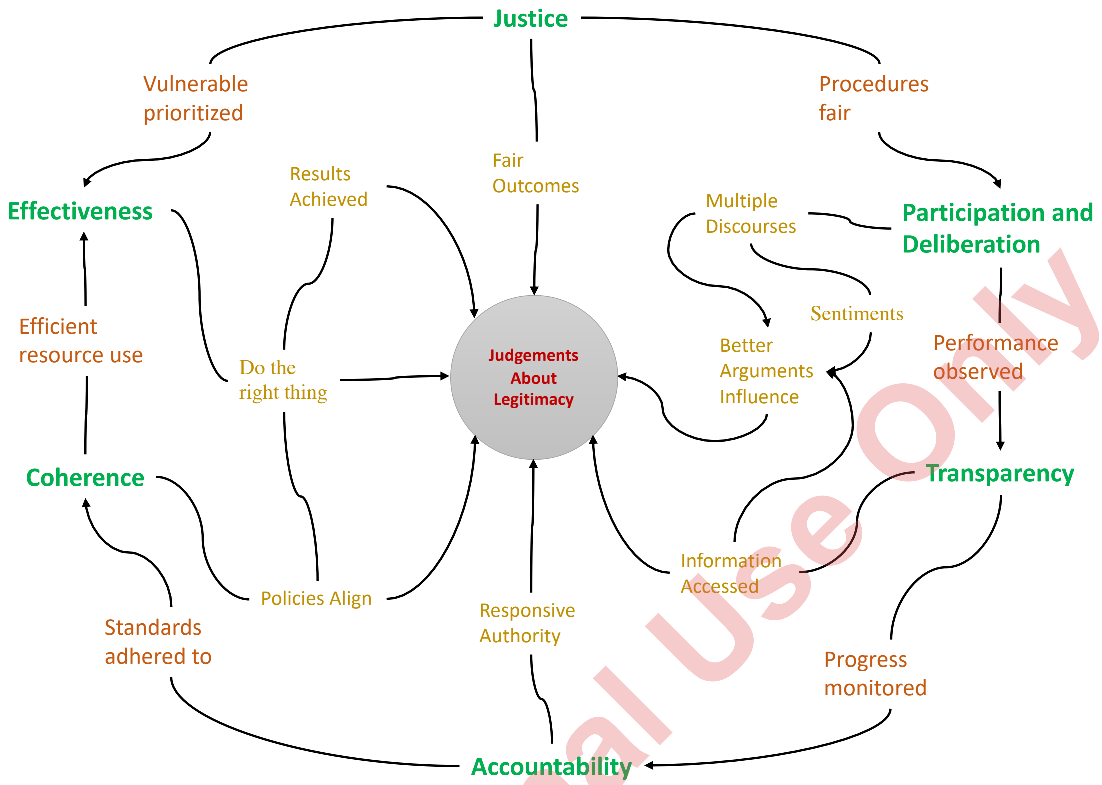

Figure 3.5: Sources of legitimacy (in green) distinguished with illustrative mechanisms and relationships

Four of these sources (participation and deliberation, transparency, accountability, and coherence) emphasise input or procedural legitimacy. In contrast, effectiveness centres on output legitimacy, while justice encompasses both the input and the output dimensions. A detailed explanation of each source is given below.

1. **Social Justice** is a non-monetary motive where attention is paid to the norms, principles and ethical arguments to legitimise the endeavour. Since they believe it is "doing the right thing," audiences are more inclined to partner and embrace a venture that promotes justice values (Grasso, 2011). However, justice on paper and in processes is insufficient unless it leads to more equal and fair results, or distributional justice (Dore and Lebel, 2010).

2. **Participation and Deliberation** refers to those who are subject to a decision that have been included, or well represented in decision-making processes, then they are more likely to accept the authority of those who decide and act (Dingwerth and Eichinger, 2010). This source is aided by the inclusive participation of the online media, which creates multiple discourses, which, given sufficient deliberative space, contributes to higher discursive quality and better arguments, which contribute more to decisions (Schultz et al., 2013).

3. **Transparency** refers to the availability of information about procedures and key decisions. In certain cases, education-based interventions that increase actors' knowledge of their own actions can also improve transparency (Mitchell, 2011). Transparency incentivises better performance (Bernstein, 2014), but transparency criteria alone cannot be sufficient if knowledge remains unavailable or difficult to comprehend (Gupta, 2010). While our theoretical constructs might not directly signal transparency, sentiment polarity in the discursive space of online media might do so indirectly.

4. **Accountability** refers to ventures that are held accountable for unfulfilled claims or acts (Biermann and Gupta, 2011). Stakeholders whose interests it seeks to regulate may sanction or keep an accountable venture liable for poor results or other reasons. Accountability often involves the willingness to assume responsibility and act in a manner that is appropriate (Becker-blease and Sohl, 2012). As a result, claims that conform to a legal system are more likely to be accepted as legitimate.

5. **Coherence** refers to the way mutual interests of key actors are addressed. Entrepreneurial stories that resonate with investors' financial and non-financial objectives, as well as normative expectations that support "do the right thing," are often critical to the acceptance of ICO projects (Woods and Fisher, 1989). Coherence implies a fit with priorities and procedures that can foster ownership and reduce contradictory and redundant policies.

6. **Effectiveness** refers to the story of the successful venture that strongly supports the link between the ICO characteristics and the perceived expectations of the audience. Adherence to broader sociopolitical norms fosters confidence,

which can help to reduce false opinion polarisation that often emerges from discursive space of online media, which might be a product of herding but often results in losses for ESEVs.

Finally, the sources of legitimacy do not operate independently from each other (Figure 3.5). Participation in evaluative judgements and providing solutions in online media's space can help to promote accountability and building of trust. Accountability often requires transparency. Transparency, however, may not empower information users in situations where sentiments are highly polarised, or reduce legitimacy where a venture's claims are perceived as arbitrary or ineffective. Even where there is a lack of policy coherence or effectiveness in aspects of "do the right thing," audiences' deliberation and social learning could legitimise entrepreneurial claims.

3.3.1 Distinctiveness and Initial Coin Offerings (ICOs)

This section explores the normative benefits, cognitive liabilities, and competitive advantages of distinctiveness within the context of ICOs. Given strong expectations for innovation, we hypothesise that distinctiveness will lead to more favourable evaluations of a venture's normative legitimacy among ICO investors.

An ESEV's perceived novelty increases when its EST diverges substantially from the market category prototype. Research on venture financing (e.g., crowdfunding) shows that ESEVs often lack patents or similar objective novelty indicators (Catalini and Gans, 2016; Mollick, 2014). Furthermore, investors typically put minimal effort into ESEV due diligence (Catalini and Gans, 2016; Short et al., 2017). Hence, it is likely that EST distinctiveness will sway investors' perceptions of an ESEV's novelty, strengthening its normative legitimacy.

Distinctiveness is likely to reduce investors' cognitive legitimacy assessments of ESEVs due to inherent ambiguity about the venture's potential. Audiences with low tolerance to ambiguity (a reluctance to embrace ambiguous situations (Stanley Budner, 1962)) tend to disfavour ventures lacking clearly defined paths (Aldrich and

Fiol, 1994). This aversion is due to a preference for certainty, categorisation, and established norms (Bochner, 1965).

On the contrary, micro-level research shows that novelty-seeking individuals exhibit a high tolerance to ambiguity, often actively seeking it (Farley and Farley, 1967; Hirschman, 1980; Kahn, 1995). Frequent exposure to novelty promotes improved cognitive abilities in these individuals to handle ambiguity (Hirschman, 1980; Stanley Budner, 1962). Consequently, we argue that investors, driven by their novelty-seeking nature, possess a higher tolerance for ambiguity. This suggests that they may not automatically dismiss ambiguous ESEVs as investment targets. Therefore, we anticipate that ICO investors will be more tolerant of the cognitive challenges posed by distinctiveness.

Beyond the cognitive aspects, distinctiveness provides a crucial competitive edge in the often-saturated ICO market. With low entry barriers (Reuber and Fischer, 2009; Taeuscher, 2019), ventures face intense pressure to stand out from competitors (Reuber and Fischer, 2009). In this competitive environment, investment-seeking ventures gain a significant advantage by differentiating themselves.

Within the context of ICOs, distinctiveness functions as a powerful differentiator, allowing ESEVs to capture investor attention more effectively. Consequently, we posit that distinctiveness yields significant competitive advantages for ESEVs seeking ICO funding.

In general, we anticipate that the significant normative and competitive advantages conferred by distinctiveness will outweigh the potential cognitive challenges within the ICO landscape. This study challenges the conventional wisdom on optimal distinctiveness, which often posits an inverted U-shaped relationship between distinctiveness and positive performance outcomes (e.g., Deephouse, 1999). Instead, we propose the following hypothesis:

H1: *Distinctiveness positively affects the judgements about legitimacy of an ICO venture.*

3.3.2 Establishing Normative Legitimacy through Contribution Claims

Ventures can strategically improve their normative legitimacy by making claims that resonate with their audience's normative expectations (Lounsbury and Glynn, 2001). Fisher et al. (2017) highlights the power of community norms, suggesting that investment-seeking ventures can enhance legitimacy by articulating how they will contribute and uphold these norms. These "contribution claims" communicate the venture's values and its intended positive impact on its audience, signalling alignment and fostering a sense of shared purpose.

Our qualitative evidence supports this proposition, suggesting that claims resonating with both financial and non-financial investor motivations become particularly salient in uncertain environments. This finding is based on previous studies focused on non-financial claims in specific contexts (e.g., Díez-Martín et al., 2020). Furthermore, unlike researchers who confine contribution claims to the social realm (e.g., Farooq et al., 2014), this study examines their total outcome across social, economic and ecological vectors.

Illustrative examples include an investor tweet expressing preference for "Who says you cannot save the world and get rich at the same time? I am on the hunt for these types of deal." Another investor emphasises how "financial" normative expectations drive investment behaviour: "Being retired from the government, you must find ways to make your money work for you. I am always looking for investments with a good monthly payout, some of them are surprisingly decent!"

We predict that contribution claims within ESTs act as a powerful source of normative legitimacy. By aligning narratives with investor expectations, these claims enhance a venture's perceived suitability for investment. Consequently, we propose the following hypothesis.

H2: *Contribution claims positively affects the judgements about legitimacy of an ICO venture.*

Resource-providing audiences assess an ESEV's normative appropriateness based on

EST distinctiveness and contribution claims. However, the legitimising effects of these factors are not just additive (Zimmerman and Zeitz, 2002). The legitimacy–resource acquisition relationship often features an "acceptability band" (Deephouse, 1999). Once an ESEV achieves sufficient legitimacy within this band, additional sources of legitimacy yield diminishing returns. When an ESEV is deemed "legitimate enough", audiences shift their focus to other evaluative criteria to differentiate among legitimate ESEVs (Deephouse and Carter, 2005).

An ESEV can draw on multiple sources to establish legitimacy. However, since the positive effects of legitimacy have limits, ESEVs benefit only from additional sources until they meet their audience's acceptability band. We posit that benefits from multiple sources of normative legitimacy will partially overlap. Thus, distinctiveness and contribution claims serve a similar legitimising function.

If a venture establishes sufficient legitimacy through numerous contribution claims, its reliance on distinctiveness as a legitimating force decreases. This suggests that the positive effect of distinctiveness on legitimacy is most pronounced in the absence of contribution claims and diminishes as number of contribution claims becomes more prevalent. Therefore, we hypothesise the following.

H3: *Contribution claims negatively affect the positive relationship between distinctiveness and judgements about the legitimacy of an ICO venture.*

Regarding H3, we lack a formal hypothesis on specific factors that might compromise the audience's "acceptability band". However, we posit that negative sentiment, which we term "polarised herd" behaviour, within the online media discourse may narrow this band. This could ultimately weaken the negative impact of contribution claims on the distinctiveness-legitimacy relationship.

As the acceptability band narrows, the need for additional legitimacy signals intensifies to offset the distinctiveness-legitimacy trade-off. This assertion challenges the original "acceptability band" theory of Deephouse (1999), which links established organisations and quantifiable or cost-orientated signals to reducing information asymmetries. The highly ambiguous context of ICOs presents a clear contrast.

Consequently, in the early stages of an ICO venture's life cycle, novelty-seeking audiences largely depend on venture-driven online activity for cues about distinctiveness and contribution claims. This activity, spread through social media channels, amplifies participation and prompts multiple discourses that lead to discursive evaluations (Hart, 2014) and sentiments. Evaluations with a high density of positive sentiment, when aligned with the volume of online activity, positively correlate with the legitimacy of a venture (Barger et al., 2016; Chanson et al., 2018; Salamanca et al., 2017).

This process strengthens intersubjective consensus among audiences, refining perceptions of the venture's contribution claims and distinctiveness. Participation and deliberation form the theoretical foundation for this consensus-building mechanism.

Therefore, ventures that engage in more online media interactions prior to their ICO will improve the reachability of their communication and, when combined with positive sentiments, will positively mediate judgements about legitimacy. This leads to our hypothesis about mechanisms that increase the appropriateness of EST and, potentially, indirectly increase the degree of underpricing found in an ICO.

H4: *Online media activity volume and positive sentiment about the distinctiveness and contributions claims prior to the ICO positively mediates the judgements about the legitimacy of the ICO venture.*

Unlike the effect of sheer volume, audiences develop effective responses, such as joy (positive sentiment) or disappointment (negative sentiment) in the online media (Haack et al., 2014). These responses emerge from aggregated individual judgements within online discourse (Bitektine and Haack, 2015; Johnson et al., 2006; Tost, 2011). Consequently, these aggregated effective responses can shape the formation of individual judgements (Haidt, 2001).

Therefore, we expect that the volume of online media activity will generate sentiment aggregation, evolving into sentiment polarity. The polarity of positive sentiment generally supports legitimacy, while negative sentiment undermines the legitimising impact of contribution claims. This leads to our final prediction regarding legitimacy-conferring mechanisms, which may indirectly influence the level of underpricing observed in ICOs.

H5: *Online media's negative sentiment negatively moderates the relationship between contribution claims and judgements about the legitimacy of an ICO venture.*

3.3.3 Market Category Viability

Investors can also infer the novelty of a proposed venture on the basis of its market category. The theory of cultural entrepreneurship emphasises the critical role of market categories in shaping the perceptions of the audience of new ventures (Navis and Glynn, 2010). In uncertain markets, ventures use the reach and diversity of online media to shape public perceptions. The use of multiple sources of legitimacy helps to manage uncertainty and instability (Goel and Nelson, 2020) by signalling an appropriate balance between differentiation and assimilation.

We emphasise the viability of the market category – the degree to which a category maintains an optimal balance between coherence and distinctiveness (Lo et al., 2020) – as a key categorical property. As a category emerges and matures, the need for viability increases. Given that investors, as novelty seekers, seek ventures that ethically "do the right thing" and generate profits, it follows that they reduce risk by favouring ventures in viable market categories (where distinctiveness and coherence are balanced). In the absence of more specific evaluative data, investors draw on multiple sources of information. Thus, membership in a market category with appropriate viability offers an additional signal, helping investors perceive member ventures as more suitable investment targets than those in unviable categories.

Building upon our arguments on market category viability, we anticipate that market category viability will mitigate the cognitive challenges associated with evaluating a new venture, reducing audience uncertainty. This facilitates a sociocognitive perspective (emphasising coherence) alongside a relational perspective (emphasising distinction), fostering a perception of viability aligned with the concept of the "zone of viability." This aligns with prospect theory (Kahneman and Tversky, 2018), which posits that investors seek to minimise losses and maximise returns (both financial and non-financial) in unpredictable situations.

Consequently, these assumptions likely create pressure on ventures to convey category viability through distinctive entrepreneurial storytelling while remaining within the "zone of viability." Therefore, the distinctiveness of EST will yield stronger cognitive and normative benefits for ventures embedded in highly viable (versus unviable) market categories. Therefore, we propose the following hypothesis.

H6: *Perceived market category viability positively moderates the relationship between distinctiveness and judgements about legitimacy of an ICO venture.*

3.4 Research Methodology

To empirically investigate our proposed hypotheses, we collected a comprehensive sample of ICO self-descriptions, prices, and related social media data of ICO.

3.4.1 Data and Sample

ICObench (`https://icobench.com`) and Coinmarketcap (`https://coinmarketcap.com`) were used to compile the ICO sample. We randomly selected an independent sample of 306 ICOs for manual data collection. Of these, 13 lacked the necessary five-day trading data for underpricing calculations, primarily due to non-listing on public exchanges. An additional 11 ICOs were excluded due to missing data necessary for analysis, such as Twitter or discussion forum information.

To maintain a quality standard, we also excluded ICOs that raised less than $1 million (that is, 3 ICOs) or did not reach 50% of their funding target (that is, 7 ICOs). This resulted in a final dataset of 272 ICOs. Coinmarketcap provided coin prices, and we manually collected 3,188 tweets from each ICO's official Twitter account. The forum activity was obtained from Bitcointalk (`https://bitcointalk.org`) and relevant Reddit subreddits (`https://reddit.com`), totalling 5,584 crawled threads.

Data for control variables were collected from ICODrops (`https://icodrops.com`), project websites, and white papers.

On ICObench, entrepreneurial ventures are assigned to multiple representative categories, signalling the degree of relatedness between the venture's offerings and existing sectors. For example, a "decentralised affiliate marketing" venture could be categorised as "marketing," "business services," "platform," and "smart contract," with "marketing" and "business services" having the highest relatedness. These multiple categories offer audiences a useful cognitive framework to understand a proposed venture and assess its novelty. Therefore, we consider the representative categories of ICObench to be analogous to the market categories.

To enhance external validity, our approach was to select a sample representative of the various market categories in the ICO landscape. Existing research suggests that there are differences in the way audiences perceive and assess ESEVs in sectors distinguished by their technological nature (e.g., Fisher et al., 2016). Consequently, our sample design encompassed all market segments, including both product-orientated and service-orientated ventures.

Entrepreneurial ventures craft narrative-style self-descriptions in their whitepaper, website, and on ICObench's profile section. We used custom python scripts to gather these narratives from all three sources, and concatenated the descriptions to build detailed and rich dataset. Initial data analysis led to the removal of extreme outliers. Exclusion criteria included: ICOs with a hard cap exceeding $10 million, narratives under 50 words, and campaigns previously launched by the same individuals under identical names. This process yielded a final sample of 272 ICOs.

3.4.2 Data Analysis Procedure

To test our hypotheses, we employed a series of OLS regressions using Python packages (Pandas, NumPy, StatsModels, and MatPlotLib). We specified regression models with robust standard errors to mitigate potential heteroskedasticity (White, 1980). Variance inflation factors (VIFs) were calculated to assess multicollinearity.

The mean VIF values ranged from 1.4 to 2.6 (with a maximum individual VIF of 6.61). Since VIF values greater than 10 generally indicate multicollinearity concerns (Stock and Watson, 2012), we concluded that our models were unaffected.

We identified influential outliers through Cook's distance analysis. For example, Filecoin with its $257 million raise exhibited the greatest influence and leverage. As such outliers were considered relevant, no further sample adjustments were made. Normality and linearity were evaluated using residual QQ plots and normality tests. Although White's test did not indicate significant heteroskedasticity, we applied homoskedasticity-consistent standard errors (HC1) as a precautionary measure to ensure robust estimates (Long and Ervin, 2000).

3.4.3 Measures

3.4.3.1 Dependent Variable

ICO Underpricing: We employ ICO underpricing as our dependent variable, serving as a proxy for legitimacy, given the challenge of measuring legitimacy directly (Lundmark et al., 2017; Zimmerman and Zeitz, 2002). Following Chanson et al. (2018), we calculate the underpricing based on the difference between the closing price of five days and the listing price. This approach accounts for the extreme volatility characteristic of the initial days following a coin's listing, a period often marked by uncertainty regarding the venture's value.

Unlike IPOs, ICOs exhibit higher volatility due to factors such as traffic surges that occasionally overwhelm exchanges, phased user access to mitigate system overload, and often unannounced public trading launch times. We use Ether, the predominant ICO currency, as the basis for price calculations, as it enjoys both widespread use and acceptance within the ICO ecosystem. Therefore, underpricing is defined as the percentage change between the closing price of the fifth day and the ICO issue price:

$$\text{Underpricing (\%)} = \left[\frac{\text{fifth day closing price} - \text{ICO issue price}}{\text{ICO issue price}}\right] \times 100 \qquad (3.1)$$

3.4.3.2 Independent Variables

Distinctiveness: We define distinctiveness as a metric that evaluates the extent to which an EST diverges from the prototypical entrepreneurial story associated with its market category (refer to Section F.1 to see a conceptual example for venture distinctiveness). To identify common topics within self-descriptions and online media, we used Latent Dirichlet Allocation (LDA) (Blei et al., 2003), a widely used topic modelling technique (Hannigan et al., 2019). The entrepreneurial stories were then modelled as probabilistic representations of these topics. This approach enables robust and context-sensitive quantification of multidimensional constructs like distinctiveness (Haans, 2019; Du et al., 2022) and novelty (Kaplan and Vakili, 2015).

Following established conventions (e.g., Haans, 2019; Shahid and Hahn, 2020), we specified 20 topics to strike an optimal balance between variation and interpretability. Figure 3.6 illustrates the most representative words for the extracted topic from "BlaBlaGame's" ICO.

The word sizes in the visualisation represent their probabilistic weight, indicating relative importance within our topic. We validated our topic models using the procedures established by DiMaggio (2015). This model measures how much an EST differs from the prototypical entrepreneurial story within its market category. For example, the distinctiveness of a "banking" venture reflects how much its entrepreneurial story varies from the average content found in the entrepreneurial stories of all "banking" ventures.

Following Haans (2019), we calculate the distinctiveness using the following equation:

$$D_i = \sum_{T=1}^{20} \left| \vartheta_{T,i} - \vartheta_{T,M}^{-} \right| \qquad (3.2)$$

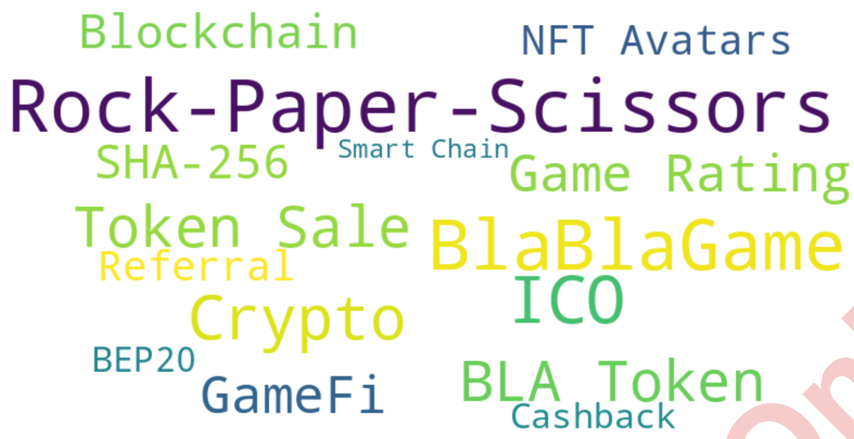

Figure 3.6: Word cloud for extracted topic

Where, D_i represents the distinctiveness of the venture i, the summation goes over $T = 1$ to $T = 20$, indicating that the calculation includes 20 topics, and $\left|\vartheta_{T,i} - \vartheta_{T,M}^{-}\right|$ calculates the absolute difference between the weight of the venture i on the topic T and the average weight of the market category M on the topic T. Therefore, the measure of distinctiveness of a venture is the aggregate of all absolute deviations between its individual topic weights and the mean topic weights for its market category, spanning the 20 topics (refer to Section F.2 to see an illustrative example for venture distinctiveness calculation). A distinctiveness of zero indicates that the entrepreneurial story mirrors the average proportions of topics found within its market category.

Contribution claims: Contribution claims are defined as the explicit or implicit statements made by an ICO venture regarding its potential positive impact on various stakeholders and the broader ecosystem. These claims can be financial (e.g., promising returns on investment), non-financial (e.g., social or environmental impact), or a combination of both. They highlight the value proposition of the project and how it aims to contribute to the community and society at large.

We employ a hybrid approach that combines computer-assisted text analysis ($CATA$), word embedding techniques, and software such as *DICTION*,[12] to systematically quantify and analyse contribution claims within ICO narratives. *CATA* provides a structured framework for extracting information from large-scale textual data (Short et al., 2010), word embeddings enrich the vocabulary, and *DICTION* allows accurate term matching and quantification.

Our process consists of four steps. First, using established CATA practices (e.g., Short et al., 2010), we begin with a deductive approach. We used tools such as Relatedwords (`https://relatedwords.org`) to identify synonyms and semantically similar words to our target concept of contribution claims (e.g., "environment," "sustainability") (Moss et al., 2011, 2018). These form the foundation of our dictionary, managed within *DICTION*.

Second, a pre-trained word embedding model (that is, *GloVe*) is used to expand our initial dictionary. We find words with high vector similarity to our CATA keywords, expanding our ability to detect diverse expressions of contribution. We focus on words with high vector similarity to our CATA keywords. Candidate words are carefully reviewed and added to the *DICTION* word list if they align with our contribution claim concept.

Third, we use *DICTION* to scan ICO narratives, highlighting sentences that contain words from our expanded dictionary. Additionally, sentence-level embeddings are compared to an embedding representing an ideal contribution claim statement. High similarity offers a signal even without exact dictionary matches.

Fourth, the hybrid approach is validated by comparing its results with manual human coding of a sample of ICO narratives. Precision, recall, F1 score, and strong Pearson correlation (that is, $r > 0.8$) between automated and manual coding support the conclusion that this method reliably and accurately measures contribution claims (refer to Section F.3 to see an example identifying contribution claims).

We operationalise **online media activity** as the volume of strategic engagement a venture undertakes with audiences across various online platforms (e.g., official Twitter account, Bitcointalk, Reddit, and Medium). To avoid reverse causality

[12]The software version 7.2.1 of *DICTION* was used in our analysis.

biases, we focus on the 30-day period immediately preceding the ICO. Measures include:

ICO Tweets: All tweets (including retweets) posted from the venture's official account within the 30-day period.

Followers: The total number of Twitter followers (in thousands) on the venture's official account the day before the ICO. Following Hoffman and Fodor (2010), this serves as a proxy for popularity, as users follow ventures to collect investment-related information.

Threads: The cumulative (logarithmic) count of online discussion threads that specifically mention the venture on selected platforms during the 30-day period. Mention suggests recognition and attention, and frequent mentions could indicate greater influence of the venture.

Sentiment Polarity: Using 134 distinct hashtags, we collected 1,433 tweets authored by 398 users to analyse sentiment within Twitter conversations. To identify positive and negative judgements about ventures (Haack et al., 2014), a supervised machine learning classification model (Crammer et al., 2006) was used to analyse the sentiment of tweets.

We manually coded a training set (10% of retweets, $n = 318$) for sentiment: negative (-1), neutral (0), or positive ($+1$) relative to the venture (reliability: $Kalpha = 0.81$, $p = 0.026$). Employing a Passive-Aggressive (PA) classifier (Crammer et al., 2006) with pairwise majority voting (Hastie and Tibshirani, 1998), we then performed sentiment analysis on the remaining tweets.

Market Category Viability: We aim to devise a measure of market category viability that captures both intracategory and intercategory optimality, focusing on the degree to which a category's online discussion aligns with the concept of "zone of viability." To achieve this, we focused on two core dimensions: category coherence and category distinctiveness.

Category coherence is the degree of similarity in themes or ideas discussed within a market category. We evaluated coherence through topic modelling of online discussions. On the contrary, category distinctiveness is the degree to which a market category's core topics differ from those of adjacent categories. We operationalised this through topic modelling, analysing thematic overlap with related categories.

Online media coverage serves as a robust indicator of the future prospects of a category. Expert commentary and critiques within these platforms contribute to audience familiarity and shape broader perceptions of category viability. Following established category research (e.g., Navis and Glynn, 2010), we analysed category-specific online discussions, including articles, comments, and opinions, to assess the perceived viability of the market category.

Our primary data source was online media discussions on platforms such as Bitcointalk and Reddit. To ensure precision in identifying market category coverage, we focused on articles explicitly referencing the relevant category (e.g., art, technology). This approach mitigated false positives, especially for ambiguous category labels.

We collected and analysed all available text within the selected forums, with a primary observation period of 2008 to 2020. This comprehensive approach accounts for the path-dependent nature of category viability and minimises the impact of short-term fluctuations.

We use LDA (Blei et al., 2003) to extract key topics from online discussions. After careful consideration of model interpretability and fit, we extracted 30 topics. To assess category coherence, we calculated pairwise cosine similarity scores between the topic distributions of the discussions within each category. We measured distinctiveness by analysing the degree of topic overlap between a category and those of adjacent categories, again using cosine similarity to quantify the similarity of topic distributions.

To ensure comparability across categories, we normalised topic-based measures using z-score standardisation. This transforms the coherence and distinctiveness scores to have a mean of zero and a standard deviation of one.

Next, we calculate the "category coherence index" and "category distinctiveness index" by taking a simple average of the normalised scores for each dimension. Categories with index scores that fall within the theoretically derived "medium" ranges for both coherence and distinctiveness (approximately one standard deviation above and below the mean) were deemed to reside within the "zone of viability."

To capture the theoretical concept of the "zone of viability", we combined the category coherence index and the category distinctiveness index into a single "market category viability index (MCVI)". We used a simple average of the normalised scores, as we did not have a theoretical reason to prioritise one dimension over the other. Categories with MCVI scores closer to zero represent those residing within the optimal "zone of viability" (refer to Section F.4 to see an illustrative example for market category viability calculation).

Our search yielded 2,612 articles across the selected platforms (Bitcointalk: 1,264; Reddit: 1,348). Throughout the observation period, an average of 7.1 articles per category-month were published, highlighting the dynamic nature of online discourse surrounding market categories.

There is significant variability in the viability patterns of categories, both between and within categories. While some categories (e.g., platform) experienced relatively consistent media coverage, others (e.g., business) exhibited notable surges over time. For further insights, descriptive statistics are provided in Table 3.2.

3.4.3.3 Control Variables

To mitigate the impact of potential confounding effects beyond those explicitly hypothesised, we incorporated twelve project-level, five creator-level, and one platform-level control variables into our models.

Venture Age: Indicates the company's age at the time of the ICO (its founding date minus the ICO date). Previous research suggests that older ventures may benefit from a greater established legitimacy, which could influence underpricing (Lundmark et al., 2017).

Raised Amount: Denotes the total sum raised during the ICO (in thousands of

Table 3.2: Descriptive statistics

Variables	Mean	Median	Min	Max	Std.Dev.
ICO underpricing	6.46	5.18	-13.41	28.00	17.36
Distinctiveness	0.91	0.42	0.34	1.89	0.23
Contribution claims	4.31	2.47	0.00	53.00	3.34
Tweets	64.76	44.31	0.00	276.00	59.90
Followers	1.63	6.34	0.30	97.00	67.23
Threads	3.41	3.35	7.00	22.00	2.33
Market cat. v. index*	0.00	7.00	-4.00	3.00	1.00
Venture age	96.21	45.01	30.00	188.00	39.25
Raised amount	12.032	7.762	0.50	21.81	4.703
Valuation	28.46	13.44	6.09	97.41	14.63
Oversubscribed	0.34	1.00	0.00	1.00	0.44
ICO duration	16.33	8.00	6.00	30.00	9.48
Min cap	2.21	3.44	3.10	32.65	7.45
Max cap	0.22	0.00	0.00	1.00	0.41
Length**	0.00	6.61	-4.00	11.00	1.62
U-Token	0.37	0.31	0.00	1.00	0.41
S-Token	0.18	0.14	0.00	1.00	0.27
Bonus	0.339	0.148	0.00	1.00	0.467
GitHub	0.70	1.0	0.00	1.00	0.50
Team	4.00	2.00	2.00	30.00	13.368
Advisor	2.65	4.18	0.00	20.00	3.51
Projects-backed	3.00	9.49	1.00	29.00	13.19
Creator-projects	1.00	1.04	0.00	22.00	8.97
Insider	9.246	5.34	3.98	15.991	6.251
Rating	3.96	2.14	0.00	5.00	1.743

* After normalisation. ** After orthogonalisation.

Ether). As ICOs that generate substantial funds often command greater attention and perceived desirability, this variable could also impact underpricing.

Valuation: The implied value of the venture, calculated based on the total funds raised divided by available coins (in thousands of Ether). A higher valuation may confer greater perceived legitimacy within the cryptocurrency community, affecting underpricing dynamics.

Oversubscribed: A dummy variable indicating whether the ICO reached its funding cap (1=oversubscribed). Excess demand in oversubscribed cases may potentially increase underpricing upon coin listing.

ICO Duration: Length of the ICO in days. Variations in duration can signal fluctuations in demand over time.

Min/Max Cap: The minimum (Min Cap) and maximum (Max Cap, a dummy variable) investment levels for participation in ICO, both in Ether. Caps influence demand dynamics, which could impact underpricing.

Length: The logged word count of the narrative of the venture. Length potentially proxies campaign quality (Calic and Mosakowski, 2016), with multicollinearity issues mitigated through the modified Gram-Schmidt method (Golub and Loan, 1991).

Token Type: Dummy variables denoting token classification: U-Token for utility, S-token for security, and C-Token for currency. Due to its invariant nature across the dataset, C-Token is used as a constant in the analysis.

Bonus: Dummy variable that indicates if an early investor bonus was offered (1 = yes).

GitHub: Dummy variable that indicates the existence of a preICO GitHub repository (1 = yes).

Team: Number of team members reported preICO; typically, larger teams suggest greater project resources.

Advisor: The number of project advisors included indicates the support of industry experts.

Projects-Backed/Creator-Projects: Control for creator experience. "Projects-Backed" represents the number of projects supported by the creator; "Creator-Projects" captures previous projects launched by the same creator (log-transformed to mitigate the skew caused by highly active creators).

Insider: The percentage of the total token supply reserved for the ICO team ("Insiders").

Rating: Overall ICO quality rating provided by the hosting platform (range 0-5, with higher scores indicating higher quality).

3.5 Research Findings

Preliminary findings, obtained using Ordinary Least Squares (OLS) regression models with ICO underpricing as a dependent variable, are presented in Table 3.3.

Our regression analysis investigates the hypothesised relationships between distinctiveness, contribution claims, online media, market category viability, and their influence on ICO underpricing – a proxy for early-stage entrepreneurial ventures (ES-EVs) legitimacy. Model 1 supports Hypotheses 1 and 2: distinctiveness significantly and positively impacts underpricing ($\beta = 0.670, p < 0.001$), suggesting its strong link to increased legitimacy perceptions. Contribution claims also consistently exert a positive effect, emphasising the importance of aligning with normative expectations for enhanced legitimacy.

Model 2 tests Hypothesis 3, revealing a negative interaction between distinctiveness and contribution claims ($\beta = -0.15, p < 0.001$). This means that while both factors can independently foster legitimacy, the marginal benefit of distinctiveness declines in the presence of numerous contribution claims.

Model 3 supports Hypothesis 4, highlighting the mediating role of online media. Both a venture's activity volume and positive sentiment surrounding its distinctiveness and contribution claims positively influence underpricing. This underscores the importance of active online participation in shaping legitimacy judgements.

Model 4 validates Hypothesis 5: negative online sentiment acts as a significant moderator, negatively impacting the relationship between contribution claims and underpricing ($\beta = -0.21, p < 0.001$). This emphasises the vulnerability of ventures to negative online discourse, which can potentially undermine their legitimising narratives.

Finally, Model 5 validates Hypothesis 6. The impact of distinctiveness on underpricing is positively influenced by the perceived viability of the market category ($\beta = 0.178, p < 0.001$). This finding highlights the increased effectiveness of distinctiveness in improving legitimacy within market categories that are considered

Table 3.3: Regression results

Variables	Model 1		Model 2		Model 3		Model 4		Model 5	
	β	ρ	β	ρ	β	ρ	β	ρ	β	ρ
Distinctiveness	0.670***	0.000	0.89***	0.000	0.39*	0.000	0.12	0.000	-0.240	0.221
Contribution claims	0.08***	0.000	0.16***	0.000	0.08**	0.001	0.22***	0.000	0.040***	0.000
Tweets	0.57*	0.000	0.35*	0.100	-0.001*	0.002	0.0161*	0.000	0.34*	0.000
Followers	0.66	0.002	0.336	0.000	0.026*	0.012	0.034**	0.000	0.045	0.010
Threads	0.056*	0.013	0.564	0.000	0.079***	0.027	0.003	0.000	0.004	0.100
Market cat. v. index	-0.11***	0.000	0.17***	0.000	0.13**	0.000	0.117	0.000	0.134***	0.000
Venture age	0.14***	0.000	0.14***	0.000	-0.007	0.0071	0.654	0.000	0.14***	0.000
Raised amount	-2.356**	0.000	0.454	0.000	0.001	0.001	0.556	0.000	0.021**	0.034
Valuation	0.665**	0.000	0.012	0.000	-0.000	0.000	0.543	0.000	0.565***	0.017
Oversubscribed	-0.256	0.000	0.665	0.000	1.000***	0.428	0.054*	0.000	-0.045	0.000
ICO duration	0.115	0.000	0.432	0.000	0.030**	0.0091	0.1121	0.000	0.001***	0.000
Min cap	0.432	0.000	0.212	0.000	-0.035	0.138	0.003	0.000	0.021	0.001
Max cap	0.212*	0.000	0.657	0.000	0.424**	0.237	0.554	0.000	0.542**	0.065
Length	0.21***	0.000	0.21***	0.000	0.06	0.000	0.055	0.000	0.21***	0.000
U-Token	0.562	0.000	0.573	0.000	0.324	0.000	0.652	0.000	0.651	0.000
S-Token	0.866	0.000	0.781	0.000	0.799	0.000	0.901	0.000	0.931	0.000
Bonus	0.076*	0.000	0.061*	0.000	0.076*	0.000	0.581**	0.000	0.551**	0.000
GitHub	0.255	0.000	0.371	0.000	0.169	0.000	0.191	0.000	0.197	0.000
Team	0.026	0.000	0.0266	0.000	0.034	0.000	0.034*	0.000	0.036*	0.000
Advisor	0.059	0.000	0.554	0.000	0.055	0.000	0.441**	0.000	0.573**	0.000
Projects-backed	0.01***	0.000	0.01***	0.000	0.01***	0.000	0.01**	0.000	0.01**	0.000
Creator-projects	-0.00	0.428	-0.00	0.543	-0.00	0.643	-0.00	0.254	-0.00	0.131
Insider	0.600	0.000	0.631	0.000	0.698	0.000	0.573	0.000	0.451	0.000
Rating	0.544	0.000	0.636	0.000	0.064**	0.000	0.671	0.000	0.356	0.000
Dist. x cont.claims			-0.15***	0.000						
Dist. x cont.claims x omav[a]x p.sent.					0.058***	0.000				
Cont.claims x n.sent.							-0.21***	0.000		
Dist. x Market cat. v. index									0.178***	0.000
R^2	0.613		0.516		0.531	0.105	0.506	0.0001	0.216	0.006
P	0.000		0.000		0.000		0.000		0.000	

N=272. * p<0.05. ** p<0.01. *** p<0.001.

[a] Overall media activity volume.

Notes: Regression models for the dependent variable *ICO underpricing*. Standard errors based on the *272* samples are presented in separate columns next to the coefficient estimates.

viable.

In general, our analysis provides compelling evidence for the complex ways in which legitimacy operates within the ICO landscape. These results illuminate how distinctiveness, contribution claims, online media interactions, and market category viability collectively shape an ICO venture's success.

3.5.1 Robustness Checks and Supplementary Analysis

To rigorously test alternative explanations and quantify the magnitude of observed relationships, we performed a series of comprehensive robustness checks.

3.5.1.1 Testing for a Curvilinear Relationship

To examine the possibility of a curvilinear relationship between distinctiveness and underpricing, we introduced a squared term for distinctiveness in Model 1, aiming to detect a U-shaped impact on underpricing. The lack of statistical significance in the negative effect ($p = 0.119$) suggests that the relationship between distinctiveness and underpricing is not curvilinear. Furthermore, the use of the utest function in Stata to test the joint hypothesis for a curvilinear link between distinctiveness and underpricing resulted in the rejection of the curvilinear hypothesis ($p = 0.219$). These results reinforce the findings supporting Hypothesis 1, indicating a linear rather than curvilinear relationship between distinctiveness and ICO underpricing.

3.5.1.2 Alternative Operationalisation of Distinctiveness

In assessing the sensitivity of our results to the operationalisation of distinctiveness, we explored various prototype specifications. By comparing ICO campaigns' distinctiveness against different benchmarks (e.g., prototypical ICO narrative), including the average topics across all market categories, within fundamental (for example, the prototypical narrative within the technology categories) and specific market categories annually (for example, the prototypical cryptocurrency narrative in 2019) and quarterly, we aimed to ensure the robustness of our findings. The analysis, detailed in Appendix C, shows consistent results at these alternative reference points (refer to Section C.1 to see alternative operationalisation of distinctiveness). Moreover, robustness tests varying the topic model's number of topics to 50 (refer to Section C.2 to see regression model with topics set to 50) or 100 (refer to Section C.3 to see regression model with topics set to 100) did not affect the findings, confirming the stability of our results regarding ICO underpricing. These tests validate that our conclusions about the influence of distinctiveness on ICO underpricing are robust and not contingent on the specificity of the distinctiveness operationalisation.

3.6 Discussion and Conclusion

3.6.1 Theoretical Contributions

The quest for optimal distinctiveness – the delicate balance between being unique and fitting within market expectations – lies at the heart of cultural entrepreneurship (Lounsbury et al., 2019; Lounsbury and Glynn, 2001, 2019; Martens et al., 2007; Navis and Glynn, 2011). This study extends the optimal distinctiveness framework to the dynamic realm of Initial Coin Offerings (ICOs), demonstrating that online media plays a crucial mediating role in how distinctiveness translates into legitimacy. Our findings highlight the nuanced relationship between distinctiveness, online discourse, and audience perceptions of market viability. Specifically, we show that in markets perceived as highly familiar and viable, distinctiveness becomes an even greater asset for entrepreneurial ventures. These results suggest that the appeal of distinctiveness is amplified when investors and consumers place greater confidence in the overall market category. This finding enriches the optimal distinctiveness discourse, emphasising the interplay between firm-level strategies, audience perceptions shaped by online media, and broader market sentiment.

This study advances the understanding of legitimacy within cultural entrepreneurship by dissecting the cognitive and normative dimensions as they relate to distinctiveness.

We introduce normative legitimacy as a central mediator between distinctiveness and underpricing. Although previous research on optimal distinctiveness has emphasised the cognitive legitimising effect of entrepreneurial stories (Aldrich and Fiol, 1994; Suchman, 1995; Martens et al., 2007; Navis and Glynn, 2011), our work underscores the equally important normative legitimising role, particularly within the amplified landscape of online media. By demonstrating how distinctiveness can, under specific conditions, bestow normative legitimacy, we challenge the assumption that high distinctiveness inherently compromises legitimacy. In an era where online discourse heavily shapes public perception, our findings suggest a potential trade-off where increasing distinctiveness might slightly lessen cognitive legitimacy while augmenting the venture's normative legitimacy. This enriches the concept of optimal distinctiveness: rather than the previously assumed inverted U-shape, a strictly positive relationship between distinctiveness and underpricing might occur when differentiation benefits, normative gains amplified by online media, and cognitive liabilities are weighed within a specific institutional context.

In addition, our study reveals that normative expectations of audiences, particularly their preference for novelty, exert substantial influence on the optimal balance between categorical similarity and distinctiveness. To achieve legitimacy from novelty-seeking audiences such as crowdfunders, ICOs must not only express their novelty, but actively resonate with audiences who crave innovation. Distinctive entrepreneurial stories become powerful tools, shaping these audiences' perceptions of a venture's novel potential and fostering normative legitimacy. This audience-level dynamic highlights a significant constraint on the theorised optimal distinctiveness trade-off: ventures seeking resources from novelty-seeking audiences should prioritise crafting narratives that emphasise their bold departure from the ordinary.

Building on existing work on entrepreneurial storytelling that examined traditional or professional investors (e.g., Villanueva, 2012; Becker-blease and Sohl, 2012; Fisch and Momtaz, 2020), our study reveals the transformative impact of audience expectations on perceptions of similarity and difference, highlighting the dynamic and context-dependent nature of legitimacy for new ventures (Fisher et al., 2017). In contrast to the centralised, firm-driven strategic use of traditional media, we emphasise the reliance of new ventures on online media as a tool for establishing legitimacy. Online media provides a cost-effective platform where both ventures and their audiences engage, collaboratively shaping the venture's legitimacy capital. By activating legitimating or delegitimating narratives, online media exert a direct influence on normative legitimacy. This dynamic, in turn, can potentially reshape cognitive legitimacy, which is often assumed to be more stable. Thus, online media have the power to either improve or undermine a venture's normative legitimacy, with cascading effects on its perceived cognitive legitimacy.

In addition, the study emphasises that the link between distinctiveness and underpricing depends on two main factors: alternative sources of legitimacy and the magnitude of negative sentiment. Previous research has indicated that ventures can gain legitimacy through various sources (Zimmerman and Zeitz, 2002), helping them to achieve the audience's "range of acceptability (Deephouse, 1999)." However, we offer two counterarguments: first, the legitimising effect of distinctiveness may diminish when alternative legitimacy sources are already present. Second, strong negative sentiment, particularly in online media, can restrict the "range of acceptability." Our study shows that contribution claims, a recognised source of legitimacy in peer-to-peer funding settings (Fisher et al., 2017), weaken the distinctiveness-underpricing relationship. Furthermore, negative online sentiment has the potential to either exacerbate the diminishing returns of contribution claims or entirely negate the benefits of alternative legitimacy sources. However, entrepreneurial storytelling remains the most effective in attracting resources when conveying high distinctiveness alongside robust contribution claims. Additionally, we posit that perceived viability of a market category can improve a venture's novelty and future performance expectations, serving as an additional legitimacy source. Our findings support this and indicate that market category viability acts as a positive moderator on the distinctiveness-underpricing relationship. These results further suggest that ventures operating within less viable categories face decreased resource acquisition potential. By highlighting the role that market categories play in shaping normative legitimacy, we advance understanding beyond their often-emphasised cognitive legitimacy function (Glynn and Navis, 2013; Lounsbury and Glynn, 2001; Navis and Glynn, 2010). Ultimately, the impact of contribution claims and market category viability underscores our core propositions: distinctiveness is a vital legitimacy source for novelty-seeking audiences, while the presence of alternative legitimacy sources or negative online sentiment can diminish its positive effects.

Our theoretical approach and empirical results significantly advance the discussion on optimal distinctiveness in "strategic management" and "organisational theory" (e.g., Deephouse, 1999; Zhao et al., 2017). We bridge a vital gap in existing research by showing that audience perceptions play a key role in mediating the link between distinctiveness and organisational performance (Zhao et al., 2017). By highlighting the influence of audience expectations and tolerance for ambiguity, we provide a compelling explanation for seemingly contradictory results regarding the similarity-distinctiveness trade-off. This audience-centric lens helps contextualise studies such as Barlow et al. (2019), where a clear negative correlation between prototype similarity (or lack of distinctiveness) and the performance of the mobile application aligns with the expectation of novelty among application users. Therefore, our work reinforces a growing body of research challenging the assumption that moderate distinctiveness invariably yields optimal results (e.g., Navis and Glynn, 2011; Fisher, 2020).

3.6.2 Generalisability and Limitations

Our empirical findings hold the most relevance for scenarios where individuals or organisations prioritise novelty in their choices among numerous competing options. For example, similar patterns may emerge in "corporate innovation contests" (e.g., Boudreau et al., 2011; Piezunka and Dahlander, 2015), where judges select promising ideas from a wide variety of submissions. The results extend to various cultural industries driven by the consumer's appetite for novelty (e.g., Lampel et al., 2000), as well as other online marketplaces characterised by novelty-seeking customers (e.g., app markets). Furthermore, insights can be translated directly to entrepreneurial competitions such as Shark Tank,[13] where judges prioritise the identification of

[13]Shark Tank is an award-winning business reality television series that debuted on ABC in 2009.The show features aspiring entrepreneurs pitching their business ideas to a panel of five seasoned venture capitalists, known as "Sharks." These Sharks evaluate presentations and decide whether or not to invest in the ventures.

groundbreaking startups. Although our study focuses on distinctiveness through entrepreneurial stories, the principles could potentially apply to other differentiation strategies, such as product catalogues (e.g., Cennamo and Santalo, 2013; Bhansing et al., 2014) or unique characteristics (Lounsbury and Glynn, 2001; Krzeminska et al., 2021), provided that adherence to the specified boundary conditions.

Although this study represents a valuable initial exploration of entrepreneurial storytelling and resource acquisition in the context of ICOs, certain limitations and opportunities for further research warrant mention.

One potential limitation of our study is its reliance on online media content and coverage as the indicator of market category viability. While this is a valuable metric, complementary data sources, such as venture funding flows within a category, could provide a more comprehensive assessment of its perceived viability among financial stakeholders.

The unique characteristics of the ICO landscape made this alternative measure less suitable for our specific context. Unlike traditional venture funding, ICOs often rely heavily on crowd-sourced early-stage investments from highly speculative retail investors. Moreover, online sentiment may further influence the speculative nature of the ICO landscape. As such, venture funding flows may not offer an accurate reflection of overall market category viability, as perceived by the broader ICO investor audience. Given our focus on understanding how these audiences assess category viability, the analysis of online media discourse offered a more direct and contextually appropriate measure of these perceptions.

Furthermore, our sample size of 272 ICOs could be expanded to strengthen the validity of our findings. Including additional variables would contribute to a more comprehensive analysis. Organisational legitimacy is an intrinsically complex construct (Zimmerman and Zeitz, 2002), underpricing serves as an established but indirect proxy. Future investigations could examine alternative legitimacy proxies within the ICO landscape to develop a deeper understanding of these dynamics.

3.6.3 Implications for Entrepreneurs and Entrepreneurial Ventures

ICOs present a cost-effective fundraising avenue for emerging projects, but navigating this landscape poses distinct challenges for entrepreneurs (Short et al., 2017). This study offers valuable insights for entrepreneurs seeking to leverage ICOs for capital acquisition. First, our findings suggest that ventures significantly benefit from cultivating a high degree of distinctiveness. This empowers entrepreneurs to clearly differentiate their entrepreneurial story within a given market category. Second, incorporating tangible contributions, such as demonstrated expertise or strategic partnerships, within the ICO narrative can enhance funding success (refer to Section E.1 for a table of keywords representing contribution claims). Finally, our results indicate that the choice of market category has a considerable impact on crowdfunding outcomes. Entrepreneurs improve their chances of securing resources by strategically aligning their venture with a market category known for its growth potential and viability.

In general, this study underscores the critical interaction between entrepreneurial storytelling, online media, and audience expectations in shaping success. Our findings offer entrepreneurs a compelling basis to invest in developing entrepreneurial stories that align with the values and preferences of their target audiences. Crucially, a venture's legitimacy is not fixed, it emerges from the expectations and norms held by its audience, as reflected and amplified in online media. For those catering to novelty-seeking audiences, cultivating distinctiveness takes on paramount importance, as it actively fosters both legitimacy and a competitive edge.

Chapter 4

Conclusion

Blockchain has emerged as one of the most promising and potentially transformative technologies in recent times. Studies have shown that it is capable of comprehensively disrupting established business processes and creating trust and integrity while offering disintermediation and immutability. These aspects have huge implications for firms considering this technology. Furthermore, studies attribute blockchain to the territory of narrow-minded and perhaps misguided visionaries who fail to look through the lens of pragmatism and realism, the need to reinvent entire industries, establish new cultural norms, and transform new business processes. The challenge for blockchain is to maintain the integrity of its potential to transform industries while overcoming the many limitations of the technology.

This study attributes the stunted growth of blockchain to the lack of knowledge that lies in the interstices between the disciplinary boundaries. Based on an analysis of the existing literature, the results of this study suggest a tremendous growth of blockchain research in inherently technical research disciplines. For example, due to their close proximity to design and business-related aspects, "Computer Science", "Engineering & Technology", and "Business, Economics, & Decision Sciences" are leading the way in blockchain research. In contrast, disciplines such as "Social Sciences", which is influenced by Blockchain as a technology-as-a-service, are discovering the potential of it in their own environments. However, the research gaps that exist between disciplines can be closed by investing in interdisciplinary research. For example, new sub-disciplines can be formed according to the concept of fusion in order to redefine the strict disciplinary boundaries. Redefining disciplinary boundaries can increase the breadth of knowledge by strengthening explanatory power and the depth of knowledge by exploring the applicability of blockchain features in different contexts.

The financing of entrepreneurial ventures is a promising area of blockchain technology, where Initial Coin Offerings (ICOs) have gained significant traction as a method of raising capital. However, the success of entrepreneurial ventures depends on the legitimacy with which they are perceived by the resource-providing audiences. Research shows that legitimate ventures are more likely to survive. However, current methods of establishing legitimacy are only suitable for established organisations. Therefore, I draw attention to distinctive storytelling as a source for conferring legitimacy. I argue that distinctive entrepreneurial stories involve quantifiable effort and can be perceived as costly signals of legitimacy by resource-providing audiences.

Established organisations typically rely on traditional media to signal legitimacy, in contrast to entrepreneurial ventures that rely on online media. While traditional media is associated with one-sided views, online media can provide audiences with two-sided views by enabling feedback. Therefore, the assumption that cognitive legitimacy is taken for granted and fixed is insignificant in the context of ICOs. This study provides empirical evidence that discourse in online media can change normative legitimacy by shifting cognitive legitimacy.

Although this study draws attention to the distinctive entrepreneurial stories as sources of legitimacy, further research is needed to identify additional methods of obtaining legitimacy in the context of ICOs. Researchers can use balanced approaches by drawing parallels between existing and new financing methods and then addressing the challenges posed by the new financing methods.

Bibliography

Ackland, R. and Halpin, D. R. 2019. Change or stability in the structure of interest group networks? Evidence from Scottish Public Policy Consultations. In *Journal of Public Policy*, volume 39, pages 267–294. Cambridge University Press.

Adhami, S., Giudici, G., and Martinazzi, S. 2018. Why do businesses go crypto? An empirical analysis of initial coin offerings. *Journal of Economics and Business*, 100.

Aggarwal, R. K., Krigman, L., and Womack, K. L. 2002. Strategic IPO underpricing, information momentum, and lockup expiration selling. *Journal of Financial Economics*.

Akerlof, G. A. 1970. The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*.

Albrecht, S., Lutz, B., and Neumann, D. 2019. How Sentiment Impacts the Success of Blockchain Startups – An Analysis of Social Media Data and Initial Coin Offerings. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Albu, O. B. and Etter, M. 2016. Hypertextuality and Social Media: A Study of the Constitutive and Paradoxical Implications of Organizational Twitter Use. *Management Communication Quarterly*.

Aldrich, H. E. and Fiol, C. M. 1994. Fools Rush in? The Institutional Context of Industry Creation. *Academy of Management Review*.

Allen, F. and Faulhaber, G. R. 1989. Signalling by underpricing in the IPO market. *Journal of Financial Economics*.

Andoni, M., Robu, V., Flynn, D., Abram, S., Geach, D., Jenkins, D., McCallum, P., and Peacock, A. 2019. Blockchain technology in the energy sector: A systematic review of challenges and opportunities.

Arthurs, J. D., Busenitz, L. W., Hoskisson, R. E., and Johnson, R. A. 2009. Signaling and initial public offerings: The use and impact of the lockup period. *Journal of Business Venturing*, 24(4):360–372.

Arthurs, J. D., Hoskisson, R. E., Busenitz, L. W., and Johnson, R. A. 2008. Managerial agents watching other agents: Multiple agency conflicts regarding underpricing in IPO firms. *Academy of Management Journal*, 51(2):277–294.

Balietti, S., Mäs, M., and Helbing, D. 2015. On disciplinary fragmentation and scientific progress. *PLoS ONE*, 10(3).

Bammer, G. 2012. Strengthening interdisciplinary Research. *The Australian Council of Learned Academies*, (February):1–29.

Barger, V., Peltier, J. W., and Schultz, D. E. 2016. Social media and consumer engagement: a review and research agenda. *Journal of Research in Interactive Marketing*, 10(4):268–287.

Barlow, M. A., Verhaal, J. C., and Angus, R. W. 2019. Optimal distinctiveness, strategic categorization, and product market entry on the Google Play app platform. *Strategic Management Journal*, 40(8):1219–1242.

Becker-blease, J. R. and Sohl, J. 2012. The sources of legitimacy for new venture angel investors. Technical report.

Benedetti, H. and Kostovetsky, L. 2021. Digital tulips? Returns to investors in initial coin offerings. *Journal of Corporate Finance*, 66:101786.

Benz, D., Hamzah, M., Ghazali, M. F., and Asli, M. F. 2022. Bringing Blockchain Technology in Innovating Industries: A Systematic Review. *Lecture Notes in Networks and Systems*, 322:391–416.

Bergstrom, C. T., West, J. D., and Wiseman, M. A. 2008. The Eigenfactor™ metrics.

Bernstein, S. 2014. Legitimacy in Global Environmental Governance. *Forestry*, 1(2003).

Bhansing, P. V., Leenders, M. A. A. M., and Wijnberg, N. M. 2014. Product Innovativeness, Organizational Legitimacy and Reputation in the High Arts: What Signals Drive Decisions by Intermediaries? Technical report.

Biermann, F. and Gupta, A. 2011. Accountability and legitimacy in earth system governance: A research framework. *Ecological Economics*, 70(11).

Bigelow, L., Lundmark, L., McLean Parks, J., and Wuebker, R. 2014. Skirting the Issues: Experimental Evidence of Gender Bias in IPO Prospectus Evaluations. *Journal of Management*, 40(6):1732–1759.

Biglaiser, G. 1993. Middlemen as Experts. *The RAND Journal of Economics*, 24(2):212.

Bingham, C. B. and Kahl, S. J. 2013. The process of schema emergence: Assimilation, deconstruction, unitization and the plurality of analogies. *Academy of Management Journal*, 56(1):14–34.

Bitektine, A. 2011. Toward a theory of social judgments of organizations: The case of legitimacy, reputation, and status. *Academy of Management Review*.

Bitektine, A. and Haack, P. 2015. The "macro" and the "micro" of legitimacy: Toward a multilevel theory of the legitimacy process. *Academy of Management Review*, 40(1):49–75.

Blaseg, D. 2018. Dynamics of voluntary disclosure in the unregulated market for initial coin offerings. *Available at SSRN 3207641*.

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Bochner, S. 1965. Defining Intolerance of Ambiguity. *The Psychological Record*.

Booth, J. R. and Chua, L. 1996. Ownership dispersion, costly information, and IPO underpricing. *Journal of Financial Economics*.

Bornmann, L. and Daniel, H. D. 2008. What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1):45–80.

Boucher, P., Nascimento, S., and Kritikos, M. 2017. How blockchain technology could change our lives - Publications Office of the EU.

Boudreau, K. J., Lacetera, N., and Lakhani, K. R. 2011. Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Science*, 57(5):843–863.

Bromham, L., Dinnage, R., and Hua, X. 2016. Interdisciplinary research has consistently lower funding success. *Nature*, 534(7609):684–687.

Calic, G. and Mosakowski, E. 2016. Kicking Off Social Entrepreneurship: How A Sustainability Orientation Influences Crowdfunding Success. *Journal of Management Studies*, 53(5):738–767.

Calzadilla, J. F. and Villa, A. 2017. Systematic Literature Review of the use of Blockchain in Supply Chain. Technical report.

Carter, R. and Manaster, S. 1990. Initial Public Offerings and Underwriter Reputation. *The Journal of Finance*, 45(4):1045–1067.

Casino, F., Dasaklis, T. K., and Patsakis, C. 2019. A systematic literature review of blockchain-based applications: Current status, classification and open issues.

Castelló, I., Morsing, M., and Schultz, F. 2013. Communicative Dynamics and the Polyphony of Corporate Social Responsibility in the Network Society.

Catalini, C. and Gans, J. S. 2016. Some Simple Economics of the Blockchain.

CB Insights 2019. Blockchain Trends In Review.

Cennamo, C. and Santalo, J. 2013. Platform competition: Strategic trade-offs in platform markets. *Strategic Management Journal*, 34(11):1331–1350.

Cerezo SSnchez, D. 2018. An Optimal ICO Mechanism. Technical report.

Chan, C. M. and Makino, S. 2007. Legitimacy and multi-level institutional environments: Implications for foreign subsidiary ownership structure. *Journal of International Business Studies*, 38(4).

Chanson, M., Risius, M., Gjoen, J., and Wortmann, F. 2018. Initial coin offerings (ICOS): The role of social media for organizational legitimacy and underpricing. In *International Conference on Information Systems 2018, ICIS 2018*.

Chaturvedi, A. and Kumaran, G. M. M. 2019. Blockchain: From invention to transformation. In *Proceedings of the 2019 6th International Conference on Computing for Sustainable Global Development, INDIACom 2019*, pages 1134–1139.

Cheng, M. Y., Hen, K. W., Tan, H. P., and Fok, K. F. 2013. Patterns of co-authorship and research collaboration in Malaysia. *Aslib Proceedings: New Information Perspectives*, 65(6):659–674.

Cho, C. H. and Patten, D. M. 2007. The role of environmental disclosures as tools of legitimacy: A research note. *Accounting, Organizations and Society*, 32(7-8):639–647.

Chukwu, E. and Garg, L. 2020. A systematic review of blockchain in healthcare: frameworks, prototypes, and implementations. *IEEE Access*, pages 1–1.

Cliff, M. T. and Denis, D. J. 2004. Do initial public offering firms purchase analyst coverage with underpricing?

Connelly, B. L., Certo, S. T., Ireland, R. D., and Reutzel, C. R. 2011. Signaling theory: A review and assessment.

Conoscenti, M., Vetro, A., and De Martin, J. C. 2016. Blockchain for the Internet of Things: A systematic literature review. Technical report.

Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., and Singer, Y. 2006. Online passive-aggressive algorithms. *Journal of Machine Learning Research*, 7.

Culnan, M. J. 1986. The Intellectual Development of Management Information Systems, 1972–1982: A Co-Citation Analysis. *Management Science*, 32(2):156–172.

Davis, J. P., Eisenhardt, K. M., and Bingham, C. B. 2009. Optimal structure, market dynamism, and the strategy of simple rules. *Administrative Science Quarterly*, 54(3):413–452.

De Domenico, M., Omodei, E., and Arenas, A. 2016. Quantifying the diaspora of knowledge in the last century. *Applied Network Science*, 1(1).

Dedehayir, O. and Steinert, M. 2016. The hype cycle model: A review and future directions. *Technological Forecasting and Social Change*, 108:28–41.

Deephouse, D. L. 1996. Does isomorphism legitimate? *Academy of Management Journal*, 39(4).

Deephouse, D. L. 1999. To be different, or to be the same? It's a question (and theory) of strategic balance. *Strategic Management Journal*.

Deephouse, D. L. and Carter, S. M. 2005. An examination of differences between organizational legitimacy and organizational reputation. Technical Report 2.

Deephouse, D. L. and Suchman, M. 2012. Legitimacy in Organizational Institutionalism. In *The SAGE Handbook of Organizational Institutionalism*, pages 49–77. SAGE Publications Ltd.

Díez-Martín, F., Blanco-González, A., and Prado-Román, C. 2020. The intellectual structure of organizational legitimacy research: a co-citation analysis in business journals. *Review of Managerial Science*.

Dignam, A. 2020. Artificial intelligence, tech corporate governance and the public interest regulatory response. *Cambridge Journal of Regions, Economy and Society*, 13(1):37–54.

DiMaggio, P. 2015. Adapting computational text analysis to social science (and vice versa).

Ding, Y. 2011. Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of Informetrics*, 5(1):187–203.

Dingwerth, K. and Eichinger, M. 2010. Tamed transparency: How information disclosure under the global reporting initiative fails to empower. *Global Environmental Politics*, 10(3):74–96.

Dore, J. and Lebel, L. 2010. Gaining public acceptance: A critical strategic priority of the World Commission on Dams. *Water Alternatives*, 3(2):124–141.

Downes, D. H. and Heinkel, R. 1982. Signaling and the Valuation of Unseasoned New Issues. *The Journal of Finance*, 37(1):1–10.

Du, K., Dudar, J., and Schöch, C. 2022. Evaluation of Measures of Distinctiveness Classification of Literary Texts on the Basis of Distinctive Words. *Journal of Computational Literary Studies*, 1(1).

Durand, R., Granqvist, N., and Tyllstrom, A. 2017. *From categories to categorization : studies in sociology, organizations and strategy at the crossroads*. Emerald Group Publishing.

Dutta, P., Choi, T. M., Somani, S., and Butala, R. 2020. Blockchain technology in supply chain operations: Applications, challenges and research opportunities. *Transportation Research Part E: Logistics and Transportation Review*, 142:102067.

Elder, G. H., Johnson, M. K., and Crosnoe, R. 2003. The Emergence and Development of Life Course Theory. pages 3–19. Springer, Boston, MA.

Elsbach, K. D. 1993. Managing Organizational Legitimacy in the California Cattle Industry: What Makes Verbal Accounts Effective? *Academy of Management Proceedings*, 1993(1):212–216.

Elsbach, K. D. and Kramer, R. M. 2003. Assessing creativity in hollywood pitch meetings: Evidence for a dual-process model of creativity judgments. *Academy of Management Journal*.

Engelhardt, M. A. 2017. Hitching Healthcare to the Chain: An Introduction to Blockchain Technology in the Healthcare Sector. *Technology Innovation Management Review*, 7(10):22–34.

Eom, S. B. 2010. *Author Cocitation Analysis*. IGI Global.

Etter, M., Colleoni, E., Illia, L., Meggiorin, K., and D'Eugenio, A. 2018. Measuring Organizational Legitimacy in Social Media: Assessing Citizens' Judgments With Sentiment Analysis. *Business and Society*, 57(1):60–97.

Farley, F. and Farley, S. V. 1967. Extroversion and Stimulus-Seeking Motivation. *Journal of Consulting Psychology*, 31(2):215–216.

Farooq, O., Payaud, M., Merunka, D., and Valette-Florence, P. 2014. The Impact of Corporate Social Responsibility on Organizational Commitment: Exploring Multiple Mediation Mechanisms. *Journal of Business Ethics*, 125(4):563–580.

Felix, T. H. and von Eije, H. 2019. Underpricing in the cryptocurrency world: evidence from initial coin offerings. *Managerial Finance*, 45(4):563–578.

Fisch, C. 2019. Initial coin offerings (ICOs) to finance new ventures. *Journal of Business Venturing*, 34(1):1–22.

Fisch, C. and Momtaz, P. P. 2020. Institutional investors and post-ICO performance: an empirical analysis of investor returns in initial coin offerings (ICOs). *Journal of Corporate Finance*, 64.

Fisher, G. 2020. The Complexities of New Venture Legitimacy. *Organization Theory*, 1(2):263178772091388.

Fisher, G., Kotha, S., and Lahiri, A. 2016. Changing with the times: An integrated view of identity, legitimacy, and new venture life cycles. *Academy of Management Review*, 41(3):383–409.

Fisher, G., Kuratko, D. F., Bloodgood, J. M., and Hornsby, J. S. 2017. Legitimate to whom? The challenge of audience diversity and new venture legitimacy. *Journal of Business Venturing*.

Gartner, W. B., Bird, B. J., and Srarr, J. A. 2016. Acting As If: Differentiating entrepreneurial from organizational behavior. In *Entrepreneurship as Organizing: Selected Papers of William B. Gartner*, pages 108–126.

Gazni, A., Sugimoto, C. R., and Didegah, F. 2012. Mapping world scientific collaboration: Authors, institutions, and countries. *Journal of the American Society for Information Science and Technology*, 63(2):323–335.

Glaser, F. 2017. Pervasive Decentralisation of Digital Infrastructures: A Framework for Blockchain enabled System and Use Case Analysis. In *Proceedings of the 50th Hawaii International Conference on System Sciences (2017)*.

Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M. C., and Siering, M. 2014. Bitcoin - Asset or currency? Revealing users' hidden intentions. In *ECIS 2014 Proceedings - 22nd European Conference on Information Systems*.

Glied, S. and Bakken, S. 2007. Institutional Challenges of Interdisciplinary Research Centers. *Journal of Research Administration*, XXXVIII(2):28–36.

Glynn, M. A. and Navis, C. 2013. Categories, identities, and cultural classification: Moving beyond a model of categorical constraint. *Journal of Management Studies*.

Goel, R. K. and Nelson, M. A. 2020. How do firms use innovations to hedge against economic and political uncertainty? Evidence from a large sample of nations. *Journal of Technology Transfer*.

Golant, B. D. and Sillince, J. A. 2007. The constitution of organizational legitimacy: A narrative perspective. *Organization Studies*, 28(8):1149–1167.

Golosovsky, M. 2019. *Citation analysis and dynamics of citation networks*, volume 3 of *SpringerBriefs in Complexity*. Springer International Publishing, Cham.

Golub, G. H. and Loan, C. F. V. 1991. Matrix Computations. *Mathematics of Computation*, 56(193):380.

González-Teruel, A., González-Alcaide, G., Barrios, M., and Abad-García, M. F. 2015. Mapping recent information behavior research: an analysis of co-authorship and co-citation networks. *Scientometrics*, 103(2):687–705.

Goodwin, J. and Garfield, E. 1980. Citation Indexing-Its Theory and Application in Science, Technology, and Humanities. *Technology and Culture*, 21(4):714.

Grasso, M. 2011. The role of justice in the North-South conflict in climate change: The case of negotiations on the Adaptation Fund. *International Environmental Agreements: Politics, Law and Economics*, 11(4):361–377.

Gray, B. 2008. Enhancing Transdisciplinary Research Through Collaborative Leadership.

Gulati, R. and Higgins, M. C. 2003. Which ties matter when? The contingent effects of interorganizational partnerships on IPO success. *Strategic Management Journal*, 24(2):127–144.

Gupta, A. 2010. Transparency in global environmental governance: A coming of age? *Global Environmental Politics*, 10(3):1–9.

Haack, P., Pfarrer, M. D., and Scherer, A. G. 2014. Legitimacy-as-Feeling: How Affect Leads to Vertical Legitimacy Spillovers in Transnational Governance. *Journal of Management Studies*.

Haans, R. F. 2019. What's the value of being different when everyone is? The effects of distinctiveness on performance in homogeneous versus heterogeneous categories. *Strategic Management Journal*, 40(1):3–27.

Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*.

Hannan, M. T. 2010. Partiality of memberships in categories and audiences. *Annual Review of Sociology*, 36:159–181.

Hannan, M. T., Pólos, L., and Carroll, G. R. 2012. *Logics of organization theory: Audiences, codes, and ecologies*.

Hannigan, T. R., Haan, R. F., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., Kaplan, S., and Jennings, P. D. 2019. Topic modeling in management research: Rendering new theory from textual data. *Academy of Management Annals*, 13(2).

Hargadon, A. B. and Douglas, Y. 2001. When innovations meet institutions: Edison and the design of the electric light. *Administrative Science Quarterly*, 46(3):476–501.

Hart, C. 2014. *Discourse, grammar and ideology: Functional and cognitive perspectives*. Bloomsbury Publishing.

Hastie, T. and Tibshirani, R. 1998. Classification by pairwise coupling. *Annals of Statistics*, 26(2):451–471.

Hawlitschek, F., Notheisen, B., and Teubner, T. 2018. The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy. *Electronic Commerce Research and Applications*, 29:50–63.

Hewa, T. M., Hu, Y., Liyanage, M., Kanhare, S. S., and Ylianttila, M. 2021. Survey on Blockchain-Based Smart Contracts: Technical Aspects and Future Research.

Hicks, D. 1987. Limitations of co-citation analysis as a tool for science policy. *Social Studies of Science*, 17(2):295–316.

Hirschman, A. O. 1964. The Paternity of an Index. *American Economic Review*, 54(5):761.

Hirschman, E. C. 1980. Innovativeness, Novelty Seeking, and Consumer Creativity. *Journal of Consumer Research*.

Hofer, B. K. and Pintrich, P. R. 1997. The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1):88–140.

Hoffman, D. L. and Fodor, M. 2010. Can you measure the ROI of your social media marketing? *MIT Sloan Management Review*, 52(1):41–49.

Hölbl, M., Kompara, M., Kamišalić, A., and Zlatolas, L. N. 2018. A systematic review of the use of blockchain in healthcare. *Symmetry*, 10(10).

Holub, M. and Johnson, J. 2018. Bitcoin research across disciplines. *Information Society*, 34(2):114–126.

Hou, J., Yang, X., and Chen, C. 2018. Emerging trends and new developments in information science: a document co-citation analysis (2009–2016). *Scientometrics*, 115(2):869–892.

Howell, S. T., Niessner, M., and Yermack, D. 2019. Initial Coin Offerings: Financing Growth with Cryptocurrency Token Sales. *The Review of Financial Studies*, 33(9):3925–3974.

Humphreys, A. and Latour, K. A. 2013. Framing the game: Assessing the impact of cultural representations on consumer perceptions of legitimacy. *Journal of Consumer Research*, 40(4):773–795.

Islam, M. R., Rashid, M. M., Rahman, M. A., Mohamad, M. H. S. B., and Embong, A. H. B. 2022. A Comprehensive Analysis of Blockchain-based Cryptocurrency Mining Impact on Energy Consumption. *International Journal of Advanced Computer Science and Applications*, 13(4):590–598.

Jabbar, A. and Dani, S. 2020. Investigating the link between transaction and computational costs in a blockchain environment. *International Journal of Production Research*, 58(11):3423–3436.

Jacobs, J. A. 2015. The Critique of Disciplinary Silos. In *In Defense of Disciplines*, pages 13–26. University of Chicago Press.

Janssen, M., Weerakkody, V., Ismagilova, E., Sivarajah, U., and Irani, Z. 2020. A framework for analysing blockchain technology adoption: Integrating institutional, market and technical factors. *International Journal of Information Management*, 50:302–309.

Johnson, C., Dowd, T. J., and Ridgeway, C. L. 2006. Legitimacy as a social process. *Annual Review of Sociology*, 32:53–78.

Jost, L. 2006. Entropy and diversity.

Kahn, B. E. 1995. Consumer variety-seeking among goods and services. An integrative review. *Journal of Retailing and Consumer Services.*

Kahneman, D. and Tversky, A. 2018. Prospect theory: An analysis of decision under risk. In *Experiments in Environmental Economics*, volume 1, pages 143–172.

Kaplan, S. and Vakili, K. 2015. The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, 36(10):1435–1457.

Kennedy, M. T. 2008. Getting counted: Markets, media, and reality. *American Sociological Review*, 73(2):270–295.

Kennedy, M. T., Lo, J. Y. C., and Lounsbury, M. 2010. Category currency: The changing value of conformity as a function of ongoing meaning construction. *Research in the Sociology of Organizations*, 31:369–397.

Khan, D., Jung, L. T., and Hashmani, M. A. 2021. Systematic literature review of challenges in blockchain scalability.

Kher, R., Terjesen, S., and Liu, C. 2020. Blockchain, Bitcoin, and ICOs: a review and research agenda. *Small Business Economics*.

Kostova, T. and Zaheer, S. 1999. Organizational legitimacy under conditions of complexity: The case of the multinational enterprise. *Academy of Management Review*, 24(1).

Kreitzer, M. J. and Saper, R. 2015. From silos of care to circles of collaboration.

Krzeminska, A., Lundmark, E., and Härtel, C. E. 2021. Legitimation of a heterogeneous market category through covert prototype differentiation. *Journal of Business Venturing*, 36(2).

Kudyba, S. 2020. COVID-19 and the Acceleration of Digital Transformation and the Future of Work. *Information Systems Management*, 37(4):284–287.

Kuilman, J. G. and Wezel, F. C. 2013. Taking off: Category contrast and organizational mortality in the UK airline industry, 1919-64. *Strategic Organization*, 11(1):56–77.

Kumar Bhardwaj, A., Garg, A., and Gajpal, Y. 2021. Determinants of Blockchain Technology Adoption in Supply Chains by Small and Medium Enterprises (SMEs) in India. *Mathematical Problems in Engineering*, 2021.

Kuperberg, M. 2020. Blockchain-Based Identity Management: A Survey from the Enterprise and Ecosystem Perspective. *IEEE Transactions on Engineering Management*, 67(4):1008–1027.

Kurubacak, G., Sharma, R. C., and Uğur, S. 2022. Living in the Meta Immersive Smart 21st Century and Beyond: A Digital Transformation in Open and Distance Learning (ODL). *TAM Akademi Dergisi*.

Lach, D. 2014. Challenges of interdisciplinary research: Reconciling qualitative and quantitative methods for understanding human-landscape systems. *Environmental Management*, 53(1):88–93.

Lamin, A. and Zaheer, S. 2012. Wall street vs. main street: Firm strategies for defending legitimacy and their impact on different stakeholders. *Organization Science*, 23(1):47–66.

Lampel, J., Lant, T., and Shamsie, J. 2000. Balancing Act: Learning from Organizing Practices in Cultural Industries. *Organization Science*, 11(3):263–269.

Larson, E. L., Landers, T. F., and Begg, M. D. 2011. Building interdisciplinary research models: A didactic course to prepare interdisciplinary scholars and faculty. *Clinical and Translational Science*, 4(1):38–41.

Lee, D. K. C., Guo, L., and Wang, Y. 2018. Cryptocurrency: A new investment opportunity? *Journal of Alternative Investments*.

Legendre, A., Lanusse, A., and Rauzy, A. 2017. Toward model synchronization between safety analysis and system architecture design in industrial contexts. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10437 LNCS, pages 35–49. Springer Verlag.

Leland, H. E. and Pyle, D. H. 1977. Informational Asymmetries, Financial Structure, and Financial Intermediation. *The Journal of Finance*, 32(2):371.

Levitt, C. and Gitlin, T. 1981. *"The Whole World Is Watching": Mass Media in the Making and Unmaking of the New Left*, volume 8/9. Univ of California Press.

Light, D. W., Berger, P. L., and Luckmann, T. 1967. The Social Construction of Reality: A Treatise in the Sociology of Knowledge. *Sociological Analysis*, 28(1):55.

Lipusch, N. 2018. Initial Coin Offerings A Paradigm Shift in Funding Disruptive Innovation. *SSRN Electronic Journal*.

Liu, Z., Luong, N. C., Wang, W., Niyato, D., Wang, P., Liang, Y. C., and Kim, D. I. 2019. A Survey on Blockchain: A Game Theoretical Perspective. *IEEE Access*, 7:47615–47643.

Ljungqvist, A., Nanda, V., and Singh, R. 2006. Hot markets, investor sentiment, and IPO pricing. *Journal of Business*, 79(4):1667–1702.

Lo, J. Y., Fiss, P. C., Rhee, E. Y., and Kennedy, M. T. 2020. Category viability: Balanced levels of coherence and distinctiveness. *Academy of Management Review*, 45(1):85–108.

Long, J. S. and Ervin, L. H. 2000. Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *American Statistician*, 54(3):217–224.

Loughran, T. and Ritter, J. R. 2002. Why Don't Issuers Get Upset About Leaving Money on the Table in IPOs? *Review of Financial Studies*, 15(2 SPEC.):413–443.

Lounsbury, M., Cornelissen, J., Granqvist, N., and Grodal, S. 2019. Culture, innovation and entrepreneurship. *Innovation: Management, Policy and Practice*, 21(1):1–12.

Lounsbury, M. and Glynn, M. A. 2001. Cultural entrepreneurship: Stories, legitimacy, and the acquisition of resources. *Strategic Management Journal*, 22(6-7):545–564.

Lounsbury, M. and Glynn, M. A. 2019. *Cultural entrepreneurship: A new agenda for the study of entrepreneurial processes and possibilities*. Cambridge University Press.

Lundmark, L. W., Oh, C., and Verhaal, J. C. 2017. A little Birdie told me: Social media, organizational legitimacy, and underpricing in initial public offerings. *Information Systems Frontiers*, 19(6):1407–1422.

Lyandres, E., Palazzo, B., and Rabetti, D. 2019. Do tokens behave like securities? An anatomy of initial coin offerings. *SSRN Electronic Journal*.

Lynn, T., Rosati, P., and Fox, G. 2018. Legitimizing #Blockchain: An empirical analysis of firm-level social media messaging on twitter. In *26th European Conference on Information Systems: Beyond Digitization - Facets of Socio-Technical Change, ECIS 2018*.

MacRoberts, M. H. and MacRoberts, B. R. 1989. Problems of citation analysis: A critical review. *Journal of the American Society for Information Science*.

Magurran, A. E. 2021. Measuring biological diversity. *Current Biology*, 31(19):R1174–R1177.

Marchant, G. E. 2011. The Growing Gap Between Emerging Technologies and the Law. In *International Library of Ethics, Law and Technology*, volume 7, pages 19–33. Springer Science and Business Media B.V.

Martens, M. L., Jennings, J. E., and Jennings, P. D. 2007. Do the stories they tell get them the money they need? The role of entrepreneurial narratives in resource acquisition. Technical Report 5.

Maurer, B., Nelms, T. C., and Swartz, L. 2013. "When perhaps the real problem is money itself!": The practical materiality of Bitcoin. *Social Semiotics*, 23(2):261–277.

Maz-Machado, A. and Jiménez-Fanjul, N. 2018. Collaboration and Citation Analysis Within Social Sciences: A Comparative Analysis Between Two Fields. In *Scientometrics*. IntechOpen.

McLeish, T. and Strang, V. 2016. Evaluating interdisciplinary research: The elephant in the peer-reviewers' room.

Mervis, C. B. and Rosch, E. 1981. Categorization of Natural Objects. *Annual Review of Psychology*, 32(1):89–115.

Mitchell, R. B. 2011. Transparency for governance: The mechanisms and effectiveness of disclosure-based and education-based transparency policies. *Ecological Economics*, 70(11).

Mollick, E. 2014. The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*.

Momtaz, P. P. 2018. Token Sales and Initial Coin Offerings: Introduction. *SSRN Electronic Journal*.

Momtaz, P. P. 2019. The Pricing and Performance of Cryptocurrency. *The European Journal of Finance*.

Momtaz, P. P. 2020. Initial coin offerings. *PLoS ONE*, 15(5).

Morisse, M. 2015. Cryptocurrencies and bitcoin: Charting the research landscape. In *2015 Americas Conference on Information Systems, AMCIS 2015*.

Moss, T. W., Renko, M., Block, E., and Meyskens, M. 2018. Funding the story of hybrid ventures: Crowdfunder lending preferences and linguistic hybridity. *Journal of Business Venturing*, 33(5):643–659.

Moss, T. W., Short, J. C., Payne, G. T., and Lumpkin, G. T. 2011. Dual identities in social ventures: An exploratory study. *Entrepreneurship: Theory and Practice*, 35(4):805–830.

Most, F., Conejo, F. J., and Cunningham, L. F. 2018. Bridging past and present entrepreneurial marketing research: A co-citation and bibliographic coupling analysis. *Journal of Research in Marketing and Entrepreneurship*, 20(2):229–251.

Mundale, J. 2002. Concepts of localization: Balkanization in the brain.

Murray, G. C. and Marriott, R. 1998. Why has the investment performance of technology-specialist, European venture capital funds been so poor? *Research Policy*, 27(9):947–976.

Nagy, B. G., Pollack, J. M., Rutherford, M. W., and Lohrke, F. T. 2012. The influence of entrepreneurs' credentials and impression management behaviors on perceptions of new venture legitimacy. *Entrepreneurship: Theory and Practice*, 36(5):941–965.

National Academy Of Sciences 2010. *Persistent forecasting of disruptive technologies - report 2*.

Navis, C. and Glynn, M. 2011. Legitimate distinctiveness and the entrepreneurial identity: Influence on investor judgments of new venture plausibility. *Academy of Management Review*, 36(3):479–499.

Navis, C. and Glynn, M. A. 2010. How new market categories emerge: Temporal dynamics of legitimacy, identity, and entrepreneurship in satellite radio, 1990-2005. *Administrative Science Quarterly*, 55(3):439–471.

Negro, G., Hannan, M. T., and Rao, H. 2011. Category reinterpretation and defection: Modernism and tradition in Italian winemaking. *Organization Science*, 22(6):1449–1463.

Newman, M. E. 2004. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, 101(SUPPL. 1):5200–5205.

Obradović, S. 2019. Publication pressures create knowledge silos. *Nature human behaviour*, 3(10):1028.

O'Donovan, G. 2002. Environmental disclosures in the annual report: Extending the applicability and predictive power of legitimacy theory.

Officer, M. S. 2007. The price of corporate liquidity: Acquisition discounts for unlisted targets. *Journal of Financial Economics*, 83(3):571–598.

Ofir, M. and Sadeh, I. 2019. ICO vs IPO: Empirical Findings, Information Asymmetry and the Appropriate Regulatory Framework. *SSRN Electronic Journal*.

Okamura, K. 2019. Interdisciplinarity revisited: evidence for research impact and dynamism. *Palgrave Communications*, 5(1).

Osareh, F. 1996. Bibliometrics, citation analysis and co-citation analysis: A review of literature I. *Libri*, 46(3):149–158.

Ozmel, U., Reuer, J. J., and Gulati, R. 2013. Signals across multiple networks: How venture capital and alliance networks affect interorganizational collaboration. *Academy of Management Journal*, 56(3):852–866.

Pahwa, A. 2018. Initial coin offering versus initial public offering.

Papacharissi, Z. 2009. The Virtual Sphere 2.0: The Internet, the Public Sphere and beyond. In *Handbook of Internet Politics*.

Patten, D. M. 1992. Intra-industry environmental disclosures in response to the Alaskan oil spill: A note on legitimacy theory. *Accounting, Organizations and Society*, 17(5):471–475.

Peirce, C. S. 1993. The essential Peirce: selected philosophical writings. v.1: 1867-1893. *Choice Reviews Online*, 31(01):31–0230.

Perelló-Sobrepere, M. 2017. Affective publics: sentiment, technology, and politics. *Information, Communication & Society*, 20(12):1788–1790.

Piezunka, H. and Dahlander, L. 2015. Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing. *Academy of Management Journal*, 58(3):856–880.

Pollock, T. G. and Rindova, V. P. 2003. Media legitimation effects in the market for initial public offerings. *Academy of Management Journal*, 46(5):631–642.

Pontikes, E. G. 2012. Two sides of the same coin: How ambiguous classification affects multiple audiences' evaluations. *Administrative Science Quarterly*.

Porac, J. F. and Rosa, J. A. 1996. Rivalry, industry models, and the cognitive embeddedness of the comparable firm. *Advances in Strategic Management*, 13:363–388.

Porter, A. L., Cohen, A. S., David Roessner, J., and Perreault, M. 2007. Measuring researcher interdisciplinarity. *Scientometrics*, 72(1):117–147.

Porter, A. L. and Rafols, I. 2009. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, 81(3):719–745.

Powell, W. W. and DiMaggio, P. J. 2012. *The new institutionalism in organizational analysis*. University of Chicago press.

Rafols, I. and Meyer, M. 2010. Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, 82(2):263–287.

Rafols, I., Porter, A. L., and Leydesdorff, L. 2010. Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology*, 61(9):1871–1887.

Repko, A. F. and Rick, S. 2020. *Interdisciplinary Research: Process and Theory*. Sage Publications.

Reuber, A. R. and Fischer, E. 2009. Signalling reputation in international online markets. *Strategic Entrepreneurship Journal*.

Rindova, V. P., Petkova, A. P., and Kotha, S. 2007. Standing out: How new firms in emerging markets build reputation. *Strategic Organization*.

Risius, M. and Spohrer, K. 2017. A Blockchain Research Framework. *Business & Information Systems Engineering*, 59(6):385–409.

Ritter, J. R. and Welch, I. 2002. A review of IPO activity, pricing, and allocations.

Rogers, E. 2010. *Diffusion of innovations*. Free Press, 4 edition.

Rogers, E. M., Dearing, J. W., and Bregman, D. 1993. The Anatomy of Agenda-Setting Research. *Journal of Communication*, 43(2):68–84.

Rosa, J. A., Porac, J. F., Runser-Spanjol, J., and Saxon, M. S. 1999. Sociocognitive dynamics in a Product market. *Journal of Marketing*, 63(SUPPL.):64–77.

Rosch, E. 1983. Prototype Classification and Logical Classification: The Two Systems. In *New Trends in Conceptual Representation: Challenges to Piaget ' s Theory*, pages 73–83. Lawrence Erlbaum Associates.

Ross, A., Chadwick, A., and Vaccari, C. 2021. Digital media and the proliferation of public opinion cues online: Biases and vulnerabilities in the new attention economy. In *The Routledge Companion to Political Journalism*, pages 241–251.

Ross, S. A. 1977. The determination of financial structure: the incentive-signalling approach. *The bell journal of economics*, pages 23–40.

Ruef, M. 2000. The emergence of organizational forms: A community ecology approach. *American Journal of Sociology*, 10(3):658–714.

Sackley, W. H. 2005. Why Has IPO Underpricing Changed over Time? *CFA Digest*.

Salamanca, A., Lebel, L., and Kallayanamitra, C. 2017. The governance of adaptation financing: pursuing legitimacy at multiple levels. *International Journal of Global Warming*, 11(2):1.

Santos, F. and Eisenhardt, K. 2009. Constructing markets and shaping boundaries: Entrepreneurial power in nascent fields. *Academy of Management Journal*, 52(4):643–671.

Santos, F. M. and Eisenhardt, K. M. 2005. Organizational boundaries and theories of organization. *Organization Science*, 16(5):491–508.

Schultz, F., Castelló, I., and Morsing, M. 2013. The construction of corporate social responsibility in network societies: A communication view. *Journal of Business Ethics*, 115(4):681–692.

Scott, J., Carrington, P., and Mische, A. 2015. Relational Sociology, Culture, and Agency. In *The SAGE Handbook of Social Network Analysis*, pages 80–98.

Scott, W. R. 2014. *Institutions and organizations : ideas, interests, and identities (4th edition)*. Sage publications.

Shahid, M. N. and Hahn, J. 2020. *A cross-disciplinary review of blockchain research trends and methodologies: Topic modeling approach*, volume 2020-Janua.

Shema, H. 2013. What's wrong with citation analysis? - Scientific American Blog Network.

Shen, C. and Pena-Mora, F. 2018. Blockchain for Cities - A Systematic Literature Review. *IEEE Access*, 6:76787–76819.

Shepherd, D. A. and Zacharakis, A. 2003. A new venture's cognitive legitimacy: An assessment by customers.

Shi, C., Pukthuanthong, K., and Walker, T. 2013. Does disclosure regulation work? Evidence from international ipo markets. *Contemporary Accounting Research*, 30(1).

Shoemaker, P. J. and Reese, S. D. 2013. *Mediating the message in the 21st century: A media sociology perspective.*

Short, J. C., Broberg, J. C., Cogliser, C. C., and Brigham, K. H. 2010. Construct validation using computer-aided text analysis (CATA): An illustration using entrepreneurial orientation. *Organizational Research Methods*, 13(2):320–347.

Short, J. C., Ketchen, D. J., McKenny, A. F., Allison, T. H., and Ireland, R. D. 2017. Research on Crowdfunding: Reviewing the (Very Recent) Past and Celebrating the Present.

Smith, K. 2013. Institutional filters: The translation and re-circulation of ideas about health inequalities within policy. *Policy and politics*, 41(1):81–100.

Snihur, Y. 2016. Developing optimal distinctiveness: organizational identity processes in new ventures engaged in business model innovation. *Entrepreneurship and Regional Development*, 28(3-4):259–285.

Somers, M. R. 1995. What's Political or Cultural about Political Culture and the Public Sphere? Toward an Historical Sociology of Concept Formation. *Sociological Theory*, 13(2):113.

Sotiriadis, M. D. and van Zyl, C. 2013. Electronic word-of-mouth and online reviews in tourism services: The use of twitter by tourists. *Electronic Commerce Research*.

Stanley Budner, N. Y. 1962. Intolerance of ambiguity as a personality variable. *Journal of Personality*.

Star, S. L. and Griesemer, J. R. 1989. Institutional ecology,translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social studies of science*, 19(3):387–420.

Steele, L. G., Bostic, A., Lynch, S. M., and Abdelaaty, L. 2022. Measuring Ethnic Diversity.

Stock, J. H. and Watson, M. W. 2012. *Introduction to Econometrics Third Edition*.

Stuart, T. E., Hoang, H., and Hybels, R. C. 1999. Interorganizational endorsements and the performance of entrepreneurial ventures. *Administrative Science Quarterly*, 44(2):315–349.

Suchman, M. C. 1995. Managing Legitimacy: Strategic and Institutional Approaches. *Academy of Management Review*, 20(3):571–610.

Suddaby, R., Bitektine, A., and Haack, P. 2017. Legitimacy. *Academy of Management Annals*, 11(1):451–478.

Surwase, G., Sagar, A., Kademani, B. S., and Bhanumurthy, K. 2011. Co-citation analysis: An overview.

Swan, M. 2015. *Blockchain: Blueprint for a new economy*.

Taeuscher, K. 2019. Reputation and new venture performance in online markets: The moderating role of market crowding. *Journal of Business Venturing*.

Taeuscher, K., Bouncken, R. B., and Pesch, R. 2020. Gaining Legitimacy by Being Different: Optimal Distinctiveness in Crowdfunding Platforms. *Academy of Management Journal*.

Talukder, K. and Kuzma, J. 2008. Evaluating technology oversight through multiple frameworks: A case study of genetically engineered cotton in India. *Science and Public Policy*, 35(2):121–138.

Tapscott, D. and Tapscott, A. 2016. *Blockchain Revolution: How the Technology Behind Bitcoin is Changing Money, Business, and the World.*, volume 25. Penguin.

Tost, L. P. 2011. An integrative model of legitimacy judgments. *Academy of Management Review*, 36(4):686–710.

Tripsas, M. 2009. Technology, identity, and inertia through the lens of "The Digital Photography Company". *Organization Science*, 20(2):441–460.

Tschorsch, F. 2017. Bitcoin and Beyond: A Technical Survey on Decentralized Digital Currencies Florian. *Deloitte University Press*.

Tushman, M. L. and Anderson, P. 1986. Technological Discontinuities and Organizational Environments. *Administrative Science Quarterly*, 31(3):439.

Vacca, A., Di Sorbo, A., Visaggio, C. A., and Canfora, G. 2021. A systematic literature review of blockchain and smart contract development: Techniques, tools, and open challenges. *Journal of Systems and Software*, 174.

Vieira, E. S. 2023. The influence of research collaboration on citation impact: the countries in the European Innovation Scoreboard. *Scientometrics*, 128(6):3555–3579.

Villanueva, J. 2012. *Does it matter how you tell it ? On how Entrepreneurial Storytelling Affects the Opportunity Evaluations of Early-Stage investors A DISSERTATION SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF MINNESOTA BY Jaume Villanueva IN PARTIAL*. PhD thesis.

Wang, Y., Han, J. H., and Beynon-Davies, P. 2019. Understanding blockchain technology for future supply chains: a systematic literature review and research agenda.

Wastl, J. and Fane, B. 2024. Fragmentation: The Divided Research World – Part Two, Siloed Knowledge.

Weinberg, B. H. 1974. Bibliographic coupling: A review. *Information Storage and Retrieval*, 10(5-6):189–196.

Welch, I. 1989. Seasoned Offerings, Imitation Costs, and the Underpricing of Initial Public Offerings. *The Journal of Finance*, 44(2):421.

Welch, I. 1992. Sequential Sales, Learning, and Cascades. *The Journal of Finance*, 47(2):695–732.

Westerman, D., Spence, P. R., and Van Der Heide, B. 2014. Social Media as Information Source: Recency of Updates and Credibility of Information. *Journal of Computer-Mediated Communication*.

White, H. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4):817.

Woods, W. F. and Fisher, W. R. 1989. Human Communication as Narration: Toward a Philosophy of Reason, Value and Action. *College Composition and Communication*, 40(2):236.

Worrall, J. L. and Cohn, E. G. 2023. Citation Data and Analysis: Limitations and Shortcomings. *Journal of Contemporary Criminal Justice*.

Wry, T., Lounsbury, M., and Jennings, P. D. 2014. Hybrid vigor: Securing venture capital by spanning categories in nanotechnology. *Academy of Management Journal*.

Xie, P., Chen, H., and Hu, Y. J. 2020. Signal or Noise in Social Media Discussions: The Role of Network Cohesion in Predicting the Bitcoin Market. *Journal of Management Information Systems*, 37(4):933–956.

Xu, M., Chen, X., and Kou, G. 2019. A systematic review of blockchain.

Xu, X., Bandara, H. M., Lu, Q., Zhang, D., and Zhu, L. 2020. Understanding and Handling Blockchain Uncertainties. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 12404 LNCS, pages 108–124. Springer Science and Business Media Deutschland GmbH.

Yeoh, P. 2017. Regulatory issues in blockchain technology. *Journal of Financial Regulation and Compliance*, 25(2):196–208.

Yin, W. and Ran, W. 2021. Theoretical exploration of supply chain viability utilizing blockchain technology. *Sustainability (Switzerland)*, 13(15).

Yli-Huumo, J., Ko, D., Choi, S., Park, S., and Smolander, K. 2016. Where Is Current Research on Blockchain Technology?-A Systematic Review. *PloS one*, 11(10):e0163477.

Yu, D. and Pan, T. 2021. Tracing the main path of interdisciplinary research considering citation preference: A case from blockchain domain. *Journal of Informetrics*, 15(2):101136.

Zaghloul, E., Li, T., Mutka, M. W., and Ren, J. 2020. Bitcoin and Blockchain: Security and Privacy. *IEEE Internet of Things Journal*, 7(10):10288–10313.

Zhao, E. Y., Fisher, G., Lounsbury, M., and Miller, D. 2017. Optimal distinctiveness: Broadening the interface between institutional theory and strategic management. In *Strategic Management Journal*, volume 38, pages 93–113. John Wiley and Sons Ltd.

Zhao, E. Y., Ishihara, M., Jennings, P. D., and Lounsbury, M. 2018. Optimal distinctiveness in the console video game industry: An exemplar-based model of proto-category evolution. *Organization Science*, 29(4):588–611.

Zhu, X., Turney, P., Lemire, D., and Vellino, A. 2015. Measuring academic influence: Not all citations are equal. *Journal of the Association for Information Science and Technology*, 66(2):408–427.

Zimmerman, M. A. and Zeitz, G. J. 2002. Beyond survival: Achieving new venture growth by building legitimacy.

Zuckerman, E. W. 1999. The categorical imperative: Securities analysts and the illegitimacy discount. *American Journal of Sociology*, 104(5).

Appendix A

Figures & Tables

A.1 Preview Data Table

Doc No	Year	Title	Abstract	Text
0	2013	Information Propagation in the Bitcoin Network	Bitcoin is a digital currency that unlike traditional currencies does not rely on a centralized authority...	'inform', 'propag', 'bitcoin', 'network', 'bitcoin', 'digit', 'currenc', 'unlik', 'tradit', 'currenc', 'reli', 'central', 'author'...

Discipline	Discipline Cluster	Author	Topic1 Perc Contrib	Topic2 Perc Contrib	Topic3 Perc Contrib	Topic4 Perc Contrib	Topic5 Perc Contrib	Topic6 Perc Contrib	Topic7 Perc Contrib	Topic8 Perc Contrib	Dominant Topic
Computer Science	Computer Science	Decker, Christian; Wattenhofert, Roger	0.834	0.019	0.019	0.019	0.019	0.052	0.019	0.019	Topic 1

Doc No	Year	Title	Abstract	Text
1	2013	An Inquiry into Money Laundering Tools in the Bitcoin Ecosystem	We provide a first systematic account of opportunities and limitations of anti-money laundering (AML) in Bitcoin..	'inquiri', 'money', 'launder', 'tool', 'bitcoin', 'ecosystem', 'provid', 'systemat', 'account', 'opportun', 'limit', 'anti', 'money'...

Discipline	Discipline Cluster	Author	Topic1 Perc Contrib	Topic2 Perc Contrib	Topic3 Perc Contrib	Topic4 Perc Contrib	Topic5 Perc Contrib	Topic6 Perc Contrib	Topic7 Perc Contrib	Topic8 Perc Contrib	Dominant Topic
Computer Science	Computer Science	Moeser, Malte; Boehme, Rainer; Breuker, Dominic	0.203	0.015	0.015	0.015	0.015	0.032	0.666	0.039	Topic 7

Figure A.1: Preview of data table

Notes: This table shows a preview of the two rows from the actual data in the dataset. This table consists of 17 columns. The columns, “Year”, “Title”, “Abstract”, “Author”, and “Discipline” (partially), are collated directly from the source databases. All the other columns were generated based on the LDA output.

A.2 Discipline Clusters (Hierarchical Model)

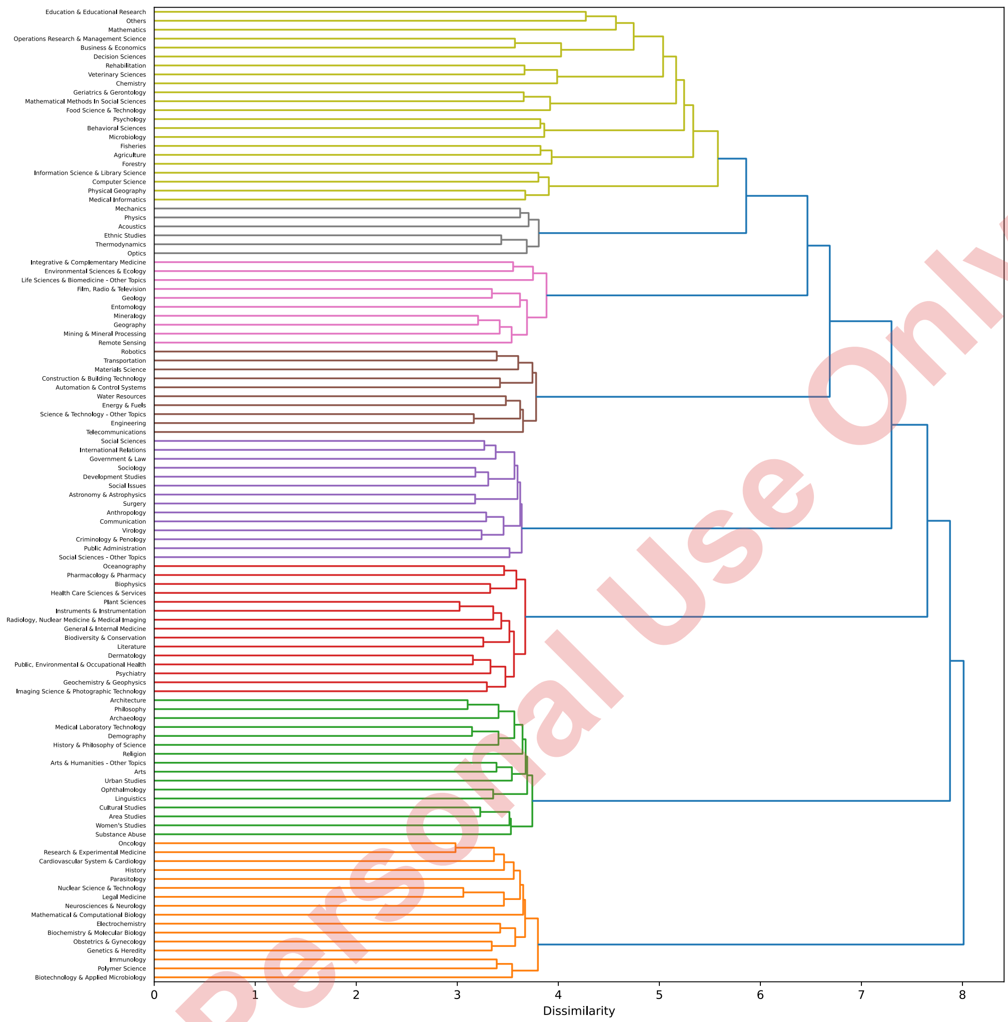

Figure A.2: Dendrogram of discipline clusters

A.3 Discipline Clusters

Table A.1: Discipline clusters

Discipline Cluster	**Disciplines**
Agriculture, Forestry & Fisheries	Agriculture; Food Science & Technology; Forestry; Fisheries
Arts & Humanities	Philosophy; Arts; Religion; Arts & Humanities - Other Topics; History & Philosophy of Science; Literature; Film; Radio & Television; History
Business, Economics, & Decision Sciences	Decision Sciences; Business & Economics; Transportation; Operations Research & Management Science
Computer Science	Computer Science; Information Science & Library Science
Cultural, Area, and Development Studies	Development Studies; Area Studies; Archaeology; Anthropology; Cultural Studies; Urban Studies
Earth and Environmental Sciences	Remote Sensing; Energy & Fuels; Environmental Sciences & Ecology; Geography; Mineralogy; Geology; Physical Geography; Oceanography; Water Resources
Engineering & Technology	Engineering; Science & Technology - Other Topics; Automation & Control Systems; Construction & Building Technology; Telecommunications; Mechanics; Robotics; Architecture; Mining & Mineral Processing; Instruments & Instrumentation; Imaging Science & Photographic Technology
Life Sciences & Biomedicine	Genetics & Heredity; Medical Informatics; Biochemistry & Molecular Biology; Biotechnology & Applied Microbiology; Pharmacology & Pharmacy; Neurosciences & Neurology; Life Sciences & Biomedicine - Other Topics; Immunology; Research & Experimental Medicine; Integrative & Complementary Medicine; Microbiology; Plant Sciences; Biodiversity & Conservation; Virology; Biophysics
Linguistics & Communication	Communication; Linguistics
Mathematics	Mathematics; Mathematical & Computational Biology; Mathematical Methods In Social Sciences
Medical Sciences & Healthcare	Health Care Sciences & Services; General & Internal Medicine; Radiology; Nuclear Medicine & Medical Imaging; Cardiovascular System & Cardiology; Surgery; Dermatology; Legal Medicine; Obstetrics & Gynecology; Medical Laboratory Technology; Psychiatry; Ophthalmology; Rehabilitation; Oncology; Substance Abuse; Geriatrics & Gerontology
Others	Others
Physical Sciences	Materials Science; Optics; Thermodynamics; Physics; Acoustics; Chemistry; Nuclear Science & Technology; Polymer Science; Astronomy & Astrophysics; Electrochemistry; Geochemistry & Geophysics
Social Sciences	Government & Law; Criminology & Penology; International Relations; Social Issues; Public; Environmental & Occupational Health; Social Sciences - Other Topics; Sociology; Education & Educational Research; Public Administration; Social Sciences; Women's Studies; Behavioral Sciences; Psychology; Demography; Ethnic Studies
Veterinary & Animal Sciences	Parasitology; Veterinary Sciences; Entomology

A.4 Topics Overlap for Discipline Clusters

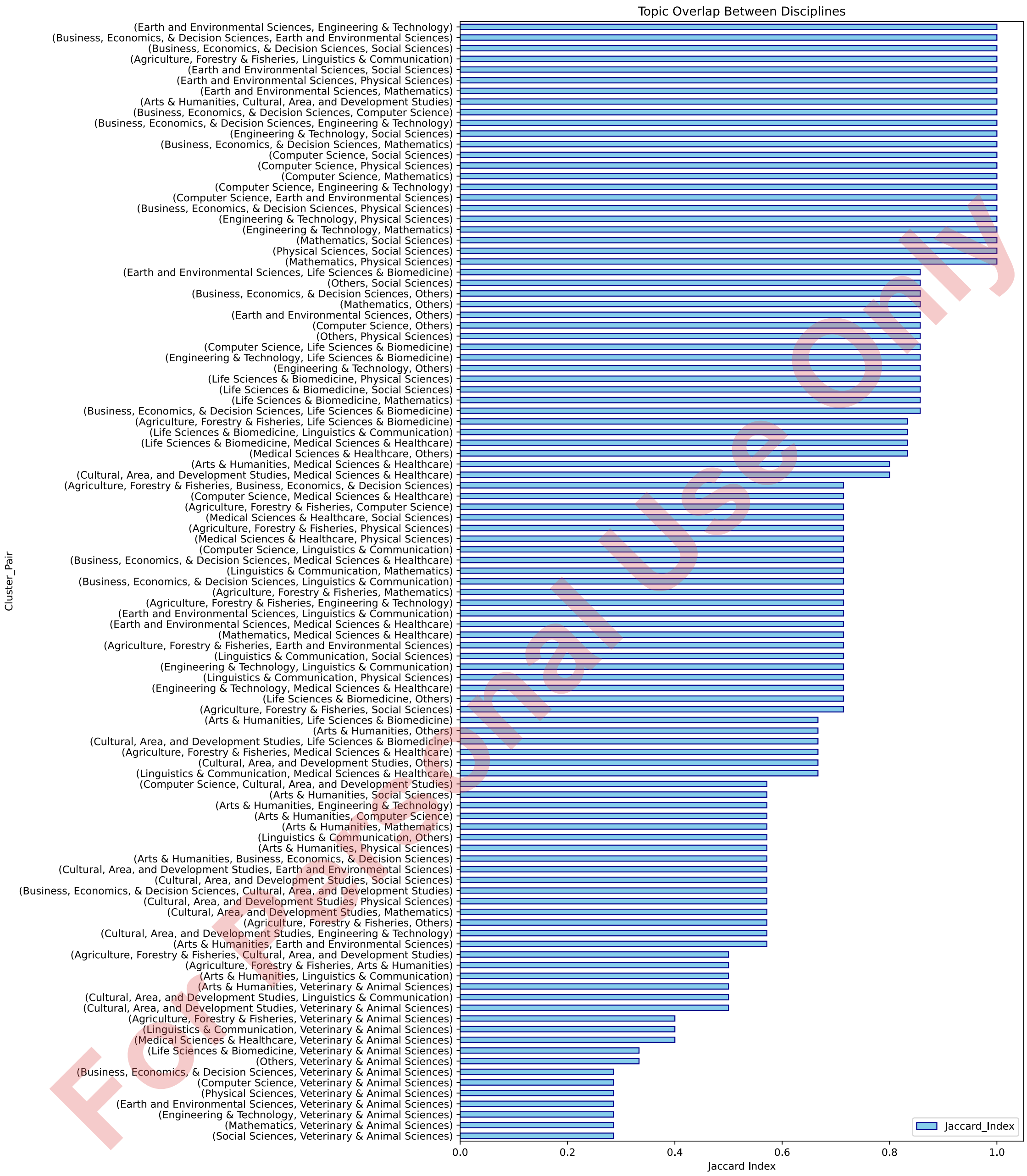

Figure A.3: Topics overlap for discipline clusters

Notes: The figure shows the Jaccard Index for topic overlap between different discipline clusters. This index provides a quantitative measure of similarity between them. The values range between 0 and 1, which shows the degree of overlap between discipline clusters. A value closer to 1 indicates greater similarity.

Appendix B

Research Focus Concentration

B.1 Illustrative Example for Research Focus Concentration

We provide following example to illustrate how the "HHI" is used in our context to measure interdisciplinarity.

Suppose that we have publications from three academic disciplines: "Computer Science", "Economics", and "Law". These publications were published in "Topic 1" in 2023. In particular, "Computer Science" has 4, "Economics" has 3, and "Law" has 3 publications, respectively. We use the following three steps to determine the Average HHI.

Step 1: Calculate the "Total Number of Papers on the Topic for the Year".

$\text{Total Papers} = 4(\text{Computer Science}) + 3(\text{Economics}) + 3(\text{Law}) = 10$

Step 2: Calculate the "Proportion of Papers from Each Discipline".

Computer Science: $p_{\text{CS}} = \frac{4}{10} = 0.4$
Economics: $p_{\text{Eco}} = \frac{3}{10} = 0.3$
Law: $p_{\text{Law}} = \frac{3}{10} = 0.3$

Step 3: Calculate the "HHI for Topic 0 in the Year 2023".

$\text{HHI}_{\text{Topic 1, 2023}} = (0.4)^2 + (0.3)^2 + (0.3)^2 = 0.16 + 0.09 + 0.09 = 0.34$

Appendix C

Regression Tables

C.1 Alternative Operationalisation of Distinctiveness

Variables	Model 2[a]		Model 2[b]		Model 2[c]		Model 2[d]		Model 2[e]	
	β	ρ	β	ρ	β	ρ	β	ρ	β	ρ
Distinctiveness	0.683***	0.000	0.96***	0.000	0.46***	0.000	0.016***	0.000	0.687***	0.000
Contribution claims	0.16***	0.000	0.18***	0.000	0.10**	0.000	0.26***	0.000	0.048***	0.000
Tweets	0.57*	0.000	0.35*	0.100	-0.001*	0.002	0.0161*	0.000	0.34*	0.000
Followers	0.66	0.002	0.336	0.000	0.026*	0.012	0.034**	0.000	0.045	0.010
Threads	0.056*	0.013	0.564	0.000	0.079***	0.027	0.003	0.000	0.004	0.100
Market cat. v. index	-0.18***	0.000	0.16***	0.000	0.11**	0.000	0.017	0.000	0.163***	0.000
Venture age	0.14***	0.000	0.14***	0.000	-0.007	0.0071	0.654	0.000	0.14***	0.000
Raised amount	-2.356**	0.000	0.454	0.000	0.001	0.001	0.556	0.000	0.021**	0.034
Valuation	0.665**	0.000	0.012	0.000	-0.000	0.000	0.543	0.000	0.565***	0.017
Oversubscribed	-0.256	0.000	0.665	0.000	1.000***	0.428	0.054*	0.000	-0.045	0.000
ICO duration	0.115	0.000	0.432	0.000	0.030**	0.0091	0.1121	0.000	0.001***	0.000
Min cap	0.432	0.000	0.212	0.000	-0.035	0.138	0.003	0.000	0.021	0.001
Max cap	0.212*	0.000	0.657	0.000	0.424**	0.237	0.554	0.000	0.542**	0.065
Length	0.21***	0.000	0.21***	0.000	0.06	0.000	0.055	0.000	0.21***	0.000
U-Token	0.562	0.000	0.573	0.000	0.324	0.000	0.652	0.000	0.651	0.000
S-Token	0.866	0.000	0.781	0.000	0.799	0.000	0.901	0.000	0.931	0.000
Bonus	0.076*	0.000	0.061*	0.000	0.076*	0.000	0.581**	0.000	0.551**	0.000
GitHub	0.255	0.000	0.371	0.000	0.169	0.000	0.191	0.000	0.197	0.000
Team	0.026	0.000	0.0266	0.000	0.034	0.000	0.034*	0.000	0.036*	0.000
Advisor	0.059	0.000	0.554	0.000	0.055	0.000	0.441**	0.000	0.573**	0.000
Projects-backed	0.01***	0.000	0.01***	0.000	0.01***	0.000	0.01**	0.000	0.01**	0.000
Creator-projects	-0.01	0.369	-0.00	0.543	-0.00	0.643	-0.00	0.254	-0.00	0.131
Insider	0.600	0.000	0.631	0.000	0.698	0.000	0.573	0.000	0.451	0.000
Rating	0.544	0.000	0.636	0.000	0.064**	0.000	0.671	0.000	0.356	0.000
Dist. x cont.claims	-0.06***	0.000	-0.11***	0.001	-0.06**	0.000	-0.11***	0.000	-0.12***	0.000
R^2	0.581		0.501		0.567	0.533	0.494	0.0001	0.417	0.006
P	0.000		0.000		0.000		0.000		0.000	

N=272. * $p<0.05$. ** $p<0.01$. *** $p<0.001$.

[a] Distinctiveness vis-a-vis market category prototype; as presented in the main models.

[b] Distinctiveness vis-a-vis ICO prototype.

[c] Distinctiveness vis-a-vis basic category prototype.

[d] Distinctiveness vis-a-vis market category prototype in the respective quarter.

[e] Distinctiveness vis-a-vis market category prototype in the respective year.

Notes: Regression models for the dependent variable *ICO underpricing*. Standard errors based on the *272* samples are presented in separate columns next to the coefficient estimates. The models differ only in the operationalisation of *Distinctiveness*.

C.2 Regression Models with Alternative Topic Modelling Parameter (50 Topics) for ICO Underpricing

Variables	Model 1		Model 2		Model 3		Model 4		Model 5	
	β	ρ	β	ρ	β	ρ	β	ρ	β	ρ
Distinctiveness	0.553***	0.000	0.76***	0.000	0.27	0.000	0.12	0.000	-0.213	0.221
Contribution claims	0.11***	0.000	0.19***	0.000	0.09**	0.001	0.24***	0.000	0.045***	0.000
Tweets	0.57*	0.000	0.35*	0.100	-0.001*	0.002	0.0161*	0.000	0.34*	0.000
Followers	0.66	0.002	0.336	0.000	0.026*	0.012	0.034**	0.000	0.045	0.010
Threads	0.056*	0.013	0.564	0.000	0.079***	0.027	0.003	0.000	0.004	0.100
Market cat. v. index	0.079***	0.000	0.19***	0.000	0.11**	0.000	0.017	0.000	0.133***	0.000
Venture age	0.14***	0.000	0.14***	0.000	-0.007	0.0071	0.654	0.000	0.14***	0.000
Raised amount	-2.316**	0.000	0.454	0.000	0.001	0.001	0.556	0.000	0.021**	0.034
Valuation	0.668**	0.000	0.012	0.000	-0.000	0.000	0.543	0.000	0.565***	0.017
Oversubscribed	-0.257	0.000	0.665	0.000	1.000***	0.428	0.054*	0.000	-0.045	0.000
ICO duration	0.115	0.000	0.432	0.000	0.030**	0.0091	0.1121	0.000	0.001***	0.000
Min cap	0.432	0.000	0.212	0.000	-0.035	0.138	0.003	0.000	0.021	0.001
Max cap	0.212*	0.000	0.657	0.000	0.424**	0.237	0.554	0.000	0.542**	0.065
Length	0.21***	0.000	0.21***	0.000	0.06	0.000	0.055	0.000	0.21***	0.000
U-Token	0.562	0.000	0.573	0.000	0.324	0.000	0.652	0.000	0.651	0.000
S-Token	0.866	0.000	0.781	0.000	0.799	0.000	0.901	0.000	0.931	0.000
Bonus	0.076*	0.000	0.061*	0.000	0.076*	0.000	0.581**	0.000	0.551**	0.000
GitHub	0.255	0.000	0.371	0.000	0.169	0.000	0.191	0.000	0.197	0.000
Team	0.026	0.000	0.0266	0.000	0.034	0.000	0.034*	0.000	0.036*	0.000
Advisor	0.059	0.000	0.554	0.000	0.055	0.000	0.441**	0.000	0.573**	0.000
Projects-backed	0.01***	0.000	0.01***	0.000	0.01***	0.000	0.01**	0.000	0.01**	0.000
Creator-projects	-0.00	0.428	-0.00	0.543	-0.00	0.643	-0.00	0.254	-0.00	0.131
Insider	0.600	0.000	0.631	0.000	0.698	0.000	0.573	0.000	0.451	0.000
Rating	0.544	0.000	0.636	0.000	0.064**	0.000	0.671	0.000	0.356	0.000
Dist. x cont.claims			-0.16***	0.000						
Dist. x cont.claims x omav[a] x p.sent.					0.067***	0.000				
Cont.claims x n.sent.							-0.29***	0.000		
Dist. x Market cat. v. index									0.119***	0.000
R^2	0.619		0.513		0.616	0.104	0.464	0.0001	0.418	0.006
P	0.000		0.000		0.000		0.000		0.000	

N=272. * p<0.05. ** p<0.01. *** p<0.001.

[a] Overall media activity volume.

Notes: Regression models for the dependent variable *ICO underpricing*. Standard errors based on *272* samples are reported in separate columns next to the coefficient estimates. The models differ from the main models in the operationalisation of the topics (25 topics instead of 20) and the resulting measure of *distinctiveness*.

C.3 Regression Models with Alternative Topic Modelling Parameter (100 topics) for ICO Underpricing

Variables	**Model 1**		**Model 2**		**Model 3**		**Model 4**		**Model 5**	
	β	ρ	β	ρ	β	ρ	β	ρ	β	ρ
Distinctiveness	0.558***	0.000	0.79***	0.000	0.27	0.000	0.12	0.000	-0.213	0.221
Contribution claims	0.14***	0.000	0.17***	0.000	0.10**	0.001	0.26***	0.000	0.041***	0.000
Tweets	0.56*	0.000	0.35*	0.100	-0.001*	0.002	0.0161*	0.000	0.34*	0.000
Followers	0.64	0.002	0.336	0.000	0.026*	0.012	0.034**	0.000	0.045	0.010
Threads	0.051*	0.013	0.564	0.000	0.079***	0.027	0.003	0.000	0.005	0.100
Market cat. v. index	0.07***	0.000	0.18***	0.000	0.13**	0.000	0.018	0.000	0.116***	0.000
Venture age	0.14***	0.000	0.14***	0.000	-0.007	0.0071	0.654	0.000	0.14***	0.000
Raised amount	-2.316**	0.000	0.454	0.000	0.001	0.001	0.556	0.000	0.021**	0.034
Valuation	0.668**	0.000	0.012	0.000	-0.000	0.000	0.543	0.000	0.565***	0.017
Oversubscribed	-0.257	0.000	0.665	0.000	1.000***	0.428	0.054*	0.000	-0.045	0.000
ICO duration	0.115	0.000	0.432	0.000	0.030**	0.0091	0.1121	0.000	0.001***	0.000
Min cap	0.432	0.000	0.212	0.000	-0.035	0.138	0.003	0.000	0.021	0.001
Max cap	0.212*	0.000	0.657	0.000	0.424**	0.237	0.554	0.000	0.542**	0.065
Length	0.21***	0.000	0.21***	0.000	0.06	0.000	0.055	0.000	0.21***	0.000
U-Token	0.562	0.000	0.573	0.000	0.324	0.000	0.652	0.000	0.651	0.000
S-Token	0.866	0.000	0.781	0.000	0.799	0.000	0.901	0.000	0.931	0.000
Bonus	0.077*	0.000	0.061*	0.000	0.076*	0.000	0.581**	0.000	0.551**	0.000
GitHub	0.255	0.000	0.371	0.000	0.169	0.000	0.191	0.000	0.197	0.000
Team	0.023	0.000	0.0266	0.000	0.034	0.000	0.034*	0.000	0.036*	0.000
Advisor	0.059	0.000	0.554	0.000	0.055	0.000	0.441**	0.000	0.573**	0.000
Projects-backed	0.01***	0.000	0.01***	0.000	0.01***	0.000	0.01**	0.000	0.01**	0.000
Creator-projects	-0.00	0.428	-0.00	0.543	-0.00	0.643	-0.00	0.254	-0.00	0.131
Insider	0.600	0.000	0.631	0.000	0.698	0.000	0.573	0.000	0.451	0.000
Rating	0.544	0.000	0.636	0.000	0.064**	0.000	0.671	0.000	0.356	0.000
Dist. x cont.claims			-0.16***	0.000						
Dist. x cont.claims x omav[a] x p.sent.					0.069***	0.000				
Cont.claims x n.sent.							-0.31***	0.000		
Dist. x Market cat. v. index									0.172***	0.000
R^2	0.633		0.613		0.518	0.105	0.601	0.0001	0.519	0.006
P	0.000		0.000		0.000		0.000		0.000	

N=272. * p<0.05. ** p<0.01. *** p<0.001.

[a] Overall media activity volume.

Notes: Regression models for the dependent variable *ICO underpricing*. Standard errors based on *272* samples are reported in separate columns next to the coefficient estimates. The models differ from the main models in the operationalisation of the topics (100 topics instead of 50) and the resulting measure of *distinctiveness*.

Appendix D

ICO Stories

D.1 Excerpts of ICO stories

"Explore the infinite potential of artificial intelligence within a wide selection of proprietary [ICO name] solutions. By catering to a diverse range of users, [ICO name] meets the ever-growing needs of artists, content creators, asset traders, and more, all within a sleek, comfortable, and intelligent platform. Bring your productivity to the next level by utilising the [ICO name] token in a vast array of real-world use cases, and discover how [ICO name] can optimise your workflow today.

[ICO name] is the premier suite of Artificial Intelligence products that makes it easy to optimise and automate workflow. Its range of AI solutions includes trade bots, chatbots, and image generators that are all powered by blockchain technology.

With [ICO name], you can trust that your data is secured and your transactions are facilitated by the latest, most reliable on-chain solutions. Its powerful AI tools are easy to use, so you can quickly get your workflow up and running. Utilise its AI products to make the most of your crypto investments, content creation, copywriting, and more, all within a single platform."

"[ICO name] is a free, social-to-earn network for community wherein users and content creators earn cryptocurrency and NFTs by learning (through short-form content), engaging with the community, and growing and improving the network. [ICO name]'s mission is to democratise social network and pave new pathways to financial freedom through our community-owned platform. Our platform provides opportunities for ownership and residual income to users and creators through our native token, [ICO name].”

“[ICO name] is a secure and safest place for your data centre. The power of nodes protects every thread of information. [ICO name] develops multiple systems from payment processors to the token. The payment processor of the company charges a negligible fee and takes microseconds for every transaction. The features of the chain are highly efficient like decentralised e-commerce development SDK, instant payment transfer, and decentralised payment Dapps.”

Appendix E

Contribution Claims (Keywords)

E.1 Keywords Representing Contribution Claims

Table E.1: Contributions claims' keywords

Keywords	Category
High ROI (Return on Investment), Speculative gains, Token appreciation, Liquidity, Innovative financial instruments, Trading/exchange opportunities	Financial
Community-driven, Decentralised governance, DAO (Decentralised Autonomous Organisation), Voting rights, Stakeholder participation, Social justice, Empowerment of creators/marginalised communities, Charity, Microfinance, Democratisation of access/financial inclusion	Non-Financial - Social
Environmental sustainability, Reduced carbon footprint, Sustainable resource management, Green technology/clean energy, Ethical supply chains	Non-Financial - Ecological
Open-source, Security audit, Reputable team/advisors, Fair distribution, Responsible development, Code of conduct, Anti-money laundering (AML), Know Your Customer (KYC)	Ethical & Transparency

Notes: The table displays contribution claim keywords categorised into three high-level areas for clarity. These keywords were theoretically validated following Short et al. (2010). Six experts with knowledge in both the theoretical and professional aspects of ICOs rated each keyword for its representativeness of contribution claims. A contribution claim was defined as a statement reflecting a company's potential to meet the financial and non-financial expectations (inclusive of economic, social, and ecological aspects) of a community and provide value for its members. The strong interrater agreement of 0.81 suggests high reliability and supports the internal validity of the selected keywords.

Appendix F

Venture Distinctiveness, Contribution Claims, & Market Category Viability

F.1 Conceptual Example for Venture Distinctiveness

Imagine a bustling street market teeming with vendors, each stall overflowing with an assortment of fruits. These stalls represent "ventures," and the fruit they offer symbolise the "topics" central to their entrepreneurial story. The market average – the common mixture of fruits displayed in most stalls – establishes a baseline for comparison.

Now, you are interested in finding the most unique and standout stalls. To do this, you will examine each stall's fruit selection compared to the typical market offering. Let us say that there are 10 varieties of fruits on the market. For each fruit type, you can note how much more or less a stall features compared to the average stall. A stall mirroring the average selection would have low or zero "distinctiveness" – nothing to distinguish it. However, a stall with a distinct mix – perhaps it specialises in rare fruits or focuses exclusively on berries – stands out dramatically from the rest.

Our distinctiveness formula mirrors this analogy. We measure each venture's "distinctiveness" by quantifying how its entrepreneurial story differs from the average of the market category. The calculation accounts for these differences in all relevant topics. A venture with an identical thematic profile to the average scores zero; however, the more its story deviates from the norm, the higher its distinctiveness score, signalling uniqueness within the market.

Like in a way where you would be drawn to distinctive fruit stalls, investors may be attracted to ventures with high distinctiveness scores.

F.2 Illustrative Example for Venture Distinctiveness Calculation

Let us illustrate the distinctiveness calculation using three ICO ventures (A, B, and C) within the same market category. We will focus on these elements of their entrepreneurial stories:

	Entrepreneurial story
T1	Innovation and technological novelty
T2	Community building and participation
T3	Environmental and social impact

Assume the following topic weights (ϑ) for each venture, along with the averages of the market category (M):

Venture	**T1**	**T2**	**T3**
A	0.40	0.30	0.30
B	0.25	0.50	0.25
C	0.35	0.15	0.50
Market (M)	0.30	0.30	0.40

We use Equation 3.2 to calculate the distinctiveness of each venture: For Venture A, (D_A) is 0.20; Venture B (D_B) is 0.40, and Venture C (D_C) is 0.30.

Here, we can conclude that Venture B has the most distinctive story (0.40), primarily due to its emphasis on community building. Venture C is also relatively distinctive (0.30), with a strong focus on social and environmental impact. Venture A is least distinctive (0.20), reflecting a more balanced emphasis across topics, closer to the average of the market category.

F.3 Example to Identifying Contribution Claims

Consider an excerpt from an ICO narrative, "Our platform aims to foster a collaborative ecosystem for impact-driven projects, driving positive change within local communities." First, our original dictionary could identify words such as "change" and "communities" during a *DICTION* scan. Second, the word embedding model suggests adding terms such as, "collaborative," "impact," and "ecosystem" due to their semantic similarity to our core concept. We add these to our *DICTION* word list after confirming their relevance. Third, *DICTION* now scans the narrative and flags the sentence due to the terms "collaborative," "impact," and "communities." Fourth, the sentence embedding likely demonstrates a high similarity score to an "ideal contribution statement" embedding. This provides a combined signal that indicates a strong contribution claim.

F.4 Illustrative Example for Market Category Viability Calculation

To assess the market viability of these ICO project categories:

1. Platform/Infrastructure
2. Finance/DeFi
3. Gaming/Metaverse

We collect data from online platforms such as Reddit and Bitcointalk.

- Step 1: Topic Modelling & Relating to Data

We use LDA to uncover the underlying themes present in all our collected ICO discussions. These topics are not the same as pre-defined categories on the platform where we collected the data.

After careful LDA analysis, consider these topics:

Topic No.	Topic Keywords
Topic 1	blockchain, protocol, scalability, security, smart contracts
Topic 2	investment, returns, tokenomics, market analysis, regulation
Topic 3	gaming, NFTs, virtual worlds, user experience, play-to-earn
Topic 4	team, vision, roadmap, partnerships, community engagement

- Step 2: Calculating Intra-Category Coherence

Analysing Topic Focus within Categories: To assess coherence, we examine how strongly discussions within each ICO category centre on specific LDA topics.

Example: If "Platform/Infrastructure" discussions heavily emphasise Topic 1, it indicates high coherence around technical themes.

- Step 3: Calculating Inter-Category Distinctiveness

Comparing Topic Emphasis Across Categories: To measure distinctiveness, we compare how dominant the LDA topics are within each category's discussions.

Example: If "Finance/DeFi" focuses on Topic 2, while "Gaming/Metaverse" centres on Topic 3, this suggests a clear distinction between them.

- Step 4: Combining for Market Category Viability

The coherence and distinctiveness scores derived from our LDA analysis are combined into our market category viability index. This helps identify categories that may reside within that optimal "zone of viability".

Appendix G

Software Tools and Utilisation

G.1 Programming and Analysis

PyCharm was utilised for writing, debugging, and testing the code for the project. It provided an integrated development environment (IDE) that facilitated the analysis process through its powerful coding assistance, error detection, and debugging tools.

For example, PyCharm was used to develop Python scripts for data preprocessing, which involved importing datasets, cleaning data, and performing exploratory data analysis.

G.2 Statistical Analysis

Stata was employed for statistical analysis, including running regressions for Study 2, generating summary statistics, and visualising data.

For example, regression analysis was conducted in Study 2 to determine the relationship between the independent and dependent variables, where multiple regression models were run.

G.3 Data Management

OpenRefine was used for data cleaning, annotation, and organisation. It helped detect and correct data inconsistencies, transform data formats, and ensure the data was properly structured for analysis.

For example, the names of authors were inconsistent across multiple publications. OpenRefine was instrumental in streamlining these names, removing duplicate entries, and correcting misspelled words.

G.4 Writing and Editing

Tools such as Wordvice, Paperpal, Writefull for Overleaf, and Grammarly were used to enhance the quality of writing by checking grammar, improving readability, and ensuring adherence to academic writing standards.

Wordvice and Grammarly were particularly useful for proofreading, while Writefull was integrated with Overleaf to assist in improving the clarity and accuracy of LaTeX document.

For example, these tools were interchangeably used to proofread all sections of this dissertation, helping to identify and correct passive voice constructions and improve sentence structure for better clarity.

G.5 Miscellaneous

Artificial Intelligence (AI) tools such as Perplexity were employed to search for relevant publications and assist in writing sections of this dissertation, particularly in refining paragraphs written by the author into an academic tone. These tools provided suggestions for rephrasing sentences and generating content that aligns with academic standards.

For example, the author would write a paragraph in their own words and then use Perplexity or Wordvice to proofread and improve it in an academic tone. This approach was applied in the introduction, methodology, and results sections, helping to ensure that the writing was clear, coherent, and formally aligned with academic expectations.

-THE END-