

Louvain School of Management

The Informativeness of ICO Whitepapers

A Topic Modeling Approach

Author: Deveux Benoît
Supervisor: Thewissen James
Academic year: 2020-2021

The Informativeness of ICO Whitepapers

A Topic Modeling Approach

Benoît Deveux
Louvain-School of Management (LSM)
benoit.deveux@student.uclouvain.be

Promotor: James Thewissen
UCLouvain
james.thewissen@uclouvain.be

Academic year: 2020-2021

ABSTRACT

The ICO industry is growing fast and becomes increasingly integrated into business models of young companies. However, this inherently unregulated environment is plagued by information asymmetry. It is thus essential to determine the ICO success factors that will help entrepreneurs as well as investors to identify valuable token sales. We take a new angle compared to previous ICO success research by examining the importance of the thematic content in ICO whitepapers. To do so, we use a topic modeling approach to quantify the potential informational value of the thematic content. We employ a database of 1848 ICOs between 2016 and 2020 from ICO.Bench.com. We show that topics matter for the ICO success (TokenTradedHistorical), the amount raised by the ICO (AmountRaised) and its return relative to the market return (AbnormalIR). In a second step, we discuss the different topics that are of interest to investors. We show that blockchain related topics positively impact the success while law category reacts negatively.

Keywords: ICOs; whitepapers; thematic content; sentence-based topic modeling; topics.

I would like to warmly thank James Thewissen, Professor of Finance at UCLouvain (LSM), for its guidance in writing this paper. The numerous exchanges and discussions allowed me to better understand the subject of ICO whitepapers and I have been able to benefit from his experience on several occasions. His timely advice, and scientific approach have helped me to a very great extent in the achievement of my Master thesis. In addition, special thanks to Prabal Shrestha, PHD Student at KU Leuven, for his support in the realization of this article.

Contents

1	Introduction	1
2	Literature Review	3
2.1	Initial Coin Offering	3
2.1.1	ICO - History	3
2.1.2	ICO - Few Regulations	6
2.1.3	ICO - As A Fraudulent Capital Raising	6
2.2	Determinants of ICOs Success	7
2.2.1	Theoretical Framework Explaining ICO Success	7
2.2.2	Whitepaper Availability & Characteristics	7
2.2.3	Linguistic Features	8
2.2.4	The Thematic Content Motivation	8
3	Research Questions	9
4	Data, Methodology and Empirical Results	9
4.1	Data Selection	9
4.2	Methodology	10
4.2.1	Automated Text Analysis	10
4.2.2	The Sent-LDA Method	11
4.2.3	Our Process	12
4.2.4	Models	13
4.3	Empirical Results	17
4.3.1	Summary Statistics	17
4.4	Topics in ICO Whitepapers	19
4.5	The Thematic Content Matters	23
4.6	Informative Topics in Whitepapers	25
5	Additional Analysis - Simpson Diversity	28
6	Sensitivity Checks	29
7	Conclusion	30
8	Appendices	34

1. Introduction

New ventures need to raise funds in order to grow (Gompers and Lerner, 2004), and the task is part of a tough competitive environment (Chemmanur and Fulghieri, 2014). An Initial Coin Offering (ICO), allows new companies, such as startups, to raise funds by selling tokens to investors. Based on blockchain technologies, cryptocurrencies are used as a means of exchange. ICOs are innovative in entrepreneurial finance and bring new possibilities of capital raising for young entrepreneurs. In fact, it's becoming more and more integrated into the business model of companies as less efforts are requested and it provide solutions by avoiding costs and intermediaries. This allows entrepreneurs and investors to keep all the profit for themselves (Kaal and Dell'Erba, 2017). Although they are recent, ICOs are valuable. In October 2018, the amount raised globally by ICOs was more than \$20 billion¹ and the capital raising through ICOs exceeded the entire European venture capital industry (Amsden and Schweizer, 2018).

There is currently a lot of information asymmetry between investors and entrepreneurs given the fact that this industry relies on limited regulations and lack of experience as ventures are in early stages (Fisch, 2019). The truth is that there is not a lot of objective information that is shared and this environment easily leads to numerous frauds (Shifflett and Jones, 2018). Thus, determining the value of ICOs is not straightforward for investors, who only have limited and sometimes imprecise information about the project (An, Hou, and Liu, 2018). In fact, small investors such as individuals, don't even have incentive to determine whether the signals sent by the token sellers are truthful or not (Becht, Bolton, and Röell, 2003). Investors must therefore make their own judgement by evaluating the company, the founding team of the ICO, and the whitepaper explaining their project. Even if, there are many ratings that are established and cross-referenced by different websites that try to assess ICOs according to different criteria (e.g., team composition), it has been shown that these ratings often differ from one site to another, making the assessment imprecise for investors (Boreiko and Sahdev, 2018).

¹About \$15 billions of dollars in 2019 for the same time period.

However, ICO whitepapers are the only document provided in order to decrease information asymmetry (Feng, Li, Wong, and Zhang, 2019; Zhang, Aerts, Lu, and Pan, 2019) and thus, it is of the utmost importance to determine what are the factors that would indicate the success of an ICO. There are financial and textual style variables (such as readability, exaggeration, and tone) which are used in the literature as signals for ICO success. In general, they have ignored the context, but the context has been shown in the literature to provide significant predictive power either on a stand-alone basis or in a combination with financial and textual style variables (Brown, Crowley, and Elliott, 2020). Thus, the theme contains informational value and incorporating it could possibly help us to predict more of the future performance of an ICO. For this reason, we are going to investigate the thematic content of ICOs whitepapers. Particularly, we are interested to quantify, based on a topic modeling approach, the information value contained in ICOs whitepapers. It is essential to determine these factors because of the information asymmetry that plagues this industry which is unregulated and decentralized.

As of today, no research has evaluated the informational intrinsic value of ICO whitepapers by using a topic modeling approach, and most importantly, no research has been conducted on the importance of the topics assuming that not every topic is equally important to every investor. This paper investigates the effects of the thematic content on the ICO success (TokenTradedHistorical), the amount raised by the ICO (AmountRaised), the volatility (ssgarchvol) and its return minus the market return (AbnormalIR). We have collected data from ICObench.com and used a consequent dataset of 1848 ICOs from 2016 to 2020 to build our analysis. The added value of our report is that, as opposed to previous literature, we conduct our research on *what* is being released in ICO whitepapers as opposed to *how* it is released thanks to a sent-LDA approach. The key question we ask ourselves in this work is whether the thematic content of ICO whitepapers is informative to investors by following a topic modeling approach and, if so, which themes are of most interest to them.

The aim of this article is to identify the features of ICO whitepapers that allow investors to identify the future performance of an ICO. Previous literature is ambiguous on the informativeness in ICO whitepapers. There are conflicts in the literature. Taking an approach that has not been done so far, which is the thematic content, we show that there is information in the whitepaper, not that it is the most informative document, but that the thematic content has information.

2. Literature Review

In this section, we will review the ICO background and the main challenges that the industry has to overcome. The lack of regulation, fraudulent ICOs and the information asymmetry are current issues that are linked and it highlights the usefulness of identifying the determinants of ICO success. In the following part of this work, we focus on what factors have been determined in the previous literature to identify a successful ICO. After that, we will see why the thematic content matters and finally I will introduce my research questions.

2.1. Initial Coin Offering

2.1.1. ICO - History

As the name suggests, a blockchain is a series of blocks that store a range of information such as financial transactions. This technology was created by a named Satoshi Nakamoto in 2008 for the purpose of Bitcoin transaction. The advantage of the blockchain is that it facilitates electronic exchanges in a secure way that is independent of any central authority. This technology is welcomed as an alternative financing option not subject to the instability of governments and financial crises.

ICO's allow for the issuance of tokens² that are purchased by investors through the use of crypto currencies. This is an alternative to venture capital financing or business angles (for early stage companies mainly) and is made possible thanks to the blockchain technology. A cryptocurrency is different from the conventional currency on several points. Where a traditional currency is managed by a country, cryptocurrencies cannot be controlled by a single entity. Additionally, cryptocurrencies are decentralized. A second element is that each cryptocurrency is produced in limited quantities, unlike currency which can be printed by central banks.

The advantages of the blockchain are numerous, security is ensured by the decentralized aspect of the technology as well as the codes of each block which provide proof of transaction and cannot be falsified. Secondly, blockchain are not regulated by governments, there are no more intermediaries between buyers and sellers. Since the creation of Bitcoin, new cryptocurrencies have appeared and

²As we are talking about fund raising, the terms token or coin can be interchangeable. Both refer to cryptocurrency but the first emphasis that it needs a separate blockchain to operate while the second not (Amsden and Schweizer, 2018).

according to CoinMarketCap they were around 7,400 at the end of 2020. The creation of a new cryptocurrency can be registered on an already existing blockchain such as the Ethereum platform. An element that has enabled this growth is a mechanism called "smart contract". This allows everyone to create their own ICO, in a simplified way, thanks to the Blockchain Ethereum created in 2015 (Amsden and Schweizer, 2018).

When an organisation wants to raise funds, it can do so in different ways. This ranges from self-financing to Initial Public Offering (IPO). The initial coin offering (ICO), is another recent way for a new organisation to raise funds. Recent, because the first ICO took place in July 2013 by Mastercoin. There are three main reasons why ICOs are important for businesses. The first one is linked to the technology used by ICOs to raise funds, the famous blockchain³. This technology notably reduces the costs of raising funds and avoids intermediaries such as banks. Second, the fact that being able to transfer a token between two players on the internet makes tokens very liquid compared to the equity, it allows the ICO owner to create a secondary market for their investments. A final reason is that ICOs allow to develop one's own customer base and more customers lead to a positive network effect. These two final benefits are made possible thanks to the fact that ICOs leads to decentralized business and open-source project development.

Scientific literature often compares ICOs to IPOs, but the two differ clearly. IPOs are highly regulated, shareholders have a voting right and they receive a share of the operating profits of the company (Burns and Moro, 2018). As for ICOs, it is quite the opposite. There is limited regulation regarding ICOs at the moment (Zetzsche, Buckley, Barberis, and Arner, 2017), the companies are young and often have not even started their operations (Cerezo Sánchez, 2017), and the token enables its owner to benefit from the service in the future. We can therefore define Initial Coin Offering as unregulated calls for funds from companies, entrepreneurs or organisations by exchanging tokens or coins using cryptocurrencies or legal tender (Adhami, Giudici, and Martinazzi, 2018; Amsden and Schweizer, 2018).

³Here we assume that ICOs are using an already existing blockchain. It's possible for an ICO to create its own blockchain but it will increase costs. (Amsden and Schweizer, 2018)

At the time of its initiation, the ICO offers different types of tokens to investors. These come in three categories and can be purchased individually or collectively depending on the offer of the issuing organisation. The first category that we identify are the utility tokens that allow its owner to benefit from the future service set up by the company. The second, which are security tokens, make it possible to receive a share of the future profit and act as a security for investors. The last one which is the payment tokens, considered as cryptocurrencies, act as a medium of exchange for investors. Nevertheless, it is important to specify that these three types of tokens can be exchanged on dedicated platforms that are not regulated (Burns and Moro, 2018). It is not uncommon to see presales that allow large investors to buy tokens at discounted prices before other smaller investors. The aim is to stimulate investors in the project. They can thus benefit from a lower price but are encouraged to the success of the project because in the long term they will be able to use the acquired tokens or even exchange them in the hope of a higher value than the original one (Fisch, 2019). Once the presale is over, the price of the token is determined by the law of supply and demand (Kastelein, 2017). However, a presale is not always viewed positively to investors, as Amsden and Schweizer (2018) explain us, this may reflect the company's lack of confidence in the ICO's success and even a lack of funds on the part of the ICO's founders.

The first step in an ICO, well before the distribution of tokens, is to make it known to as many investors as possible. This is made possible thanks to an unaudited marketing document, called whitepaper, which might be distributed and available on dedicated sites such as cryptocurrency forums. The whitepaper is the document issued by ICO project owners in order to convince potential token buyers but there is no obligation to produce one. These are documents that might content information such as the proposed service, the token description and distribution, the team, technology, expected performance, etc. The whitepaper essentially describes the objective of the project and the means used. However, as we are in an industry with little or no existing regulation, there is no blueprint. Thus, not all whitepapers are identical, they do not all deal with the same topics, some go into detail while others do not (Fisch, 2019). The purpose of a whitepaper is to signal the venture quality and to differentiate themselves from other projects. The whitepaper is not the only one to share information but is a piece of the puzzle that the investor must use to evaluate the ICO project⁴.

⁴The information environment is composed of website, social media platforms, information aggregators, ICO rating providers and whitepapers Bourveau, De George, Ellahie, and Macciocchi (2018).

2.1.2. ICO - Few Regulations

The crypto-tokens market is growing fast and this is attracting the attention of regulators in an effort to protect investors. Since its creation in 2013, developers have always acted to maintain a decentralised market by avoiding all kinds of regulations. Regulations concern the creation of new tokens and their exchange on dedicated platforms. At the time of writing this report, this crypto-token market is unregulated and rumours are beginning to emerge that some tokens could be classified as securities (in the U.S.), which could lead them to respect existing securities jurisdictions. To determine the status of securities or not, the "Howey test" is used (Bourveau et al., 2018). In September 2020, the European Commission has adopted a series of legislation to regulate crypto assets, updating some financial market rules, and created a Pan-European blockchain regulatory sandbox⁵. The Commission came with a new framework to replace some EU rules concerning the creation of new tokens and their exchange on dedicated platforms. This is called the Markets in Crypto-Assets Regulation, MICA (legroju, 2020).

2.1.3. ICO - As A Fraudulent Capital Raising

This lack of regulation easily leads some ICOs to scams and it impacts the trust of the system. As an example of fraudulent ICO, we can cite REcoin. Created in 2017, REcoin presented itself as revolutionary in the world of real estate. The idea behind this ICO was to use the created coin as a means of payment in real estate transactions instead of conventional money. A month after the initiation, the U.S Securities and Exchange Commission (the SEC), noticed that this business had no employees (contrary to what had been announced) and that the money raised by the ICO was never invested in the real estate industry. To top it all off, the REcoin token did not exist (Morris, 2017). What is interesting in this example is that this fraud was possible, among other things, thanks to information asymmetry. There was few information available to investors such as a whitepaper with erroneous information and some news on social networks. Thus, it was difficult to investors to assess the fair value of the ICO and making decisions in this situation increased the risks.

⁵A sandbox is is a forum for bringing together people, regulators, businesses and experts to identify obstacles and test innovative solutions.

2.2. Determinants of ICOs Success

2.2.1. Theoretical Framework Explaining ICO Success

Amsden and Schweizer (2018) show that venture uncertainty, as they explained as having shorter whitepapers⁶ or not having an account on Github and Telegram for instance, is negatively correlated with ICO success. The opposite, venture quality, like a larger team size, is positively correlated. Furthermore, they found that having a hard cap (the goal amount) will increase the amount raised and the token tradability. Howell, Niessner, and Yermack (2020) find that ICOs having a whitepaper, promoting through social media such as Twitter, having certifications and sharing codes development on Github will have higher probability of success than those who do not. Then, Fisch (2019) explores the potential of technical signals such as patents, source code and technical whitepapers and conclude that whitepapers revealing high quality of technological information raise higher funding. Most importantly, only one fourth of their sample counts technical whitepapers⁷. Finally, Bourveau et al. (2018) find that the information environment, such as sharing source code, informative whitepapers⁸, using social media platforms, is positively correlated with the amount raised and the token tradability. In addition, they find that the length and the opacity of whitepapers are correlated to crash risk.

2.2.2. Whitepaper Availability & Characteristics

Literature is ambiguous on the relationship between whitepapers availability and the probability of success of an ICO. Adhami et al. (2018) show that there is no correlation between ICO success and the availability of a whitepaper. While Howell et al. (2020) find that ICOs having a whitepaper will have higher probability of success than those who do not. Additionally, informative whitepapers⁹ is positively correlated with the amount raised and the token tradability (Bourveau et al., 2018). Other studies take into consideration technical information into the discussion. Whitepapers revealing high quality of technological information raise higher funding (Fisch, 2019), technical details in a whitepaper reflect the quality of an ICO (Feng et al., 2019). Amsden and Schweizer (2018) show that venture uncertainty, as they explained as having shorter whitepapers results in a less sophisti-

⁶In their article, they stressed that having a shorter whitepaper results in a less sophisticated business plan and will decrease the token tradability.

⁷A technical whitepaper is defined by a detailed description of the technical implementation of the project (such as algorithms or sample code among others), and if the explanations reflect a technological expertise

⁸ICO project owners can differentiate their ICO by sharing informative disclosures to investors to facilitate the assessment of the future value of the project.

⁹ICO project owners can differentiate their ICO by sharing informative disclosures to investors to facilitate the assessment of the future value of the project.

cated business plan and will decrease the token tradability which is negatively correlated with ICO success. Additionally, [Zetzsche, Buckley, Arner, and Föhr \(2017\)](#) show that the technical information in whitepapers is uncertain¹⁰

2.2.3. Linguistic Features

There are few studies that have been carried out on the linguistic features of ICO whitepapers. We mainly retain the article from [Zhang et al. \(2019\)](#), that shows that readability of ICO whitepapers and ICO first-day returns are positively correlated. In addition, they stress that these documents are important sources of information for investors. Another article argued that the information included in the whitepapers is systematically exaggerated by the founders of the ICO. An ICO that sends biased signals perform better by collecting higher amounts in less time. However, once the token is exchanged, investors become aware of the bias and get an idea of the venture quality. ICOs that use biased signals have a lower initial return and higher volatility ([Momtaz, 2020](#)). Lastly, the tone and the uncertainty have already been studied. It has been found that the tone tends to be negative in ICO whitepapers. There is no significant relationship between whitepaper sentiment and ICO returns. The authors of the study suppose that investors don't pay attention to the information provided in ICO whitepapers ([Dittmar and Wu, 2019](#)). Note that in this study, authors don't analyse nor quantify the informational content of ICO whitepapers.

2.2.4. The Thematic Content Motivation

Previous literature has already looked at the success of an ICO by analysing linguistic features such as the exaggeration or the tone among other things as signals of ICO success. However, it has been shown that using textual analysis does not reflect the context ([Loughran and McDonald, 2016](#)). Yet, the context is essential to determine inferences that can be drawn. To overcome this problem, the thematic content of financial statement filings was studied by [Brown et al. \(2020\)](#) in order to determine the potential informational value and the predictive power of its model. In fact, they were motivated to determine whether the content may better predict misreporting than textual analysis. Interestingly, they find that the thematic content is incrementally informative and has predictive power either on a stand-alone basis or in a combination with Financial and textual variables.

¹⁰Only one fifth of their ICO whitepaper sample provide technical information about the project, one third do not communicate on the initiators and one fourth do not say how the collected capital will be used.

3. Research Questions

These studies on subjective characteristics (Zhang et al., 2019; Momtaz, 2020; Dittmar and Wu, 2019) have led to analyse *how* investors assess the value of the ICO but as of today no research has been carried out on *what* actually provide value to investors. There is a gap in the literature and this work aims to fill that gap by studying the informativeness of ICO whitepapers thanks to a topic modeling approach. This paper makes several important contributions. First, we investigate the content and go deeper into the analysis compared to previous linguistic studies by answering to **”Which topics are covered in ICO whitepapers?” (RQ1)**. Then, as compared to previous literature into 10K filings (e.g., Brown et al. (2020); Dyer, Lang, and Stice-Lawrence (2017); Rich, Roberts, Wall, and Zhang (2018)), we will examine the thematic content in the context of ICO whitepapers by providing insights on the second question **”Does the thematic content have informational value?” (RQ2)**. Finally, we will have a discussion on the most relevant topics by answering this last question **”Which topics in ICO whitepapers are value relevant to investors?” (RQ3)** assuming that not every topic is equally important to every investor. Our contributions suggest that extracting themes from ICO whitepapers help capture future performance of an ICO.

4. Data, Methodology and Empirical Results

This section presents the description of our dataset, the sent-LDA methodology and our empirical findings regarding the topics discussed in the ICO whitepapers. We aim at identifying the topics that help investors decide to contribute to ICO projects.

4.1. Data Selection

The task is not simple when it comes to having a global view of the ICO market because of their decentralised nature, and especially when it comes to conducting empirical studies. To date, there is no obligation for ICOs to be registered on a single platform, which means that they are scattered all over the web and makes research challenging (Fisch, 2019). Nevertheless, new actors took on this role by collecting and collating ICO data on their own platform.

For this reason, we collected our data from one of the most prominent platforms, ICObench.com. Which provides ratings¹¹, listing and calendar of ICOs from different countries and industries. ICObench is the number one ICO rating platform and it hosts more than 5,000 ICOs¹² at the time of this report. All data can be retrieved thanks to an Application Programming Interface (API). We collected not less than 1848 ICOs whitepapers from 2016 to 2020. [Fisch \(2019\)](#) would point out that there is a bias in our study because as stated in his article, taking data from only one site will lead to data mistakes. However, this website is a reference in the field of scientific literature on ICOs and has already been used several times by [Amsden and Schweizer \(2018\)](#), [Howell et al. \(2020\)](#) and [Benedetti and Kostovetsky \(2018\)](#) for the most detailed information and the largest number of ICOs.

4.2. Methodology

In this section, we will focus on what we have done compared to other previous studies and elaborate a bit on the frameworks they used.

4.2.1. Automated Text Analysis

Literature indicates that there are three main categories of automated text analysis, corresponding to dictionary, supervised and unsupervised method ([Bao and Datta, 2014](#)). The first is tagged as the simplest by using keywords to identify documents and to see in which category they belong. As an example, we can cite [Kravet and Muslu \(2013\)](#) studied risk disclosures by observing sentences that contained certain words from a list of predefined keywords. The negative point of this method is that it is necessary to define how the categories will be assigned to the documents, which can lead to irregularities if the list of keywords is not appropriate. The second method is similar to the previous one in the sense that it allows the categorisation of documents. The operation is a two-step process. first, humans have to categorise a sample document. Secondly, they have to train a model so that it can reproduce the classification for all the documents. In this case, the problem encountered in the dictionary has been solved. For instance, [Humpherys, Moffitt, Burns, Burgoon, and Felix \(2011\)](#) used this method to detect cases of fraud in financial performance reports. Finally, the last method, for which we opt for in this report, is the unsupervised machine learning. The

¹¹Interestingly, the way ICObench.com develops the ratings slightly differ from its competitors (e.g., ICOrating.com , ICODrops.com, etc.) Where others develop their ratings on the basis of internal analysis, icobench.com ratings are based on an algorithm that analyses almost 20 characteristics at a time. Furthermore, they include expert analysis that are independent from the website ([Bourveau et al., 2018](#))

¹²See ICObench.com

big difference with the two previous techniques is that here there are no imposed categories. Actually, it uses assumptions to categorise documents, the LDA or in our case, sentences, the Sent-LDA.

Some previous literatures have been using the LDA into ICO whitepapers. [Bian, Deng, Li, Monroe, Shi, Sun, Wu, Wang, Wang, Yuan, et al. \(2018\)](#) developed a model, a machine-learning based ICO rating system, to identify scam ICO projects. This article is interesting for our work in the sense that this team has categorized their 2,251 ICO projects into ten different topics by using a LDA method. They were able to conclude that ICO topics such as gaming and gambling have more chances to be a scam rather than smart contract or payments. A last article is that of [Chuanjie, Koh, and Griffin \(2019\)](#) which demonstrate that the LDA model is a good way to categorise ICO whitepapers. Furthermore, they demonstrated that not all topics are equivalent in the success of an ICO.

4.2.2. The Sent-LDA Method

To determine the informational value of ICO whitepapers, we used the Sent-LDA proposed by [Bao and Datta \(2012\)](#) instead of the classical LDA developed by [Blei, Ng, and Jordan \(2003\)](#). The first LDA study carried out by [Aral, Ipeirotis, and Taylor \(2011\)](#) analysed stock selection bias. Subsequently, [Bao and Datta \(2012\)](#) improved the classic LDA model to Sent-LDA which allows one single topic on words in the same sentence. Then, they proved that the sent-lda method is more efficient than the classic LDA ([Bao and Datta, 2012](#)). The reason is that, the results of the LDA are less accurate than the one of the sent-LDA. In the first one, a certain number and a certain set of topics will be defined for each document by assigning a topic on words and the order of the words is not retained. While for the second, topics are assigned to words corresponding to the same sentence. We are thus able to obtain more accurate results because each word in a same sentence is dependent and words in different sentences are not interchangeable. The two methods are different but have the same advantage, which is that they reduce researcher bias. This is made possible as it relies on the fact that topics are formed on the principle that words frequently appearing together share the same meaning instead of being selected based on a predefined dictionary. Notice that the sent-LDA varies across time and some modifications are made to improve the algorithm as it has already been done by [Zhu, Yang, and Moazeni \(2016\)](#) and [Bao and Datta \(2014\)](#) for better analysing risk factors in corporate disclosure for instance.

As explained by [Bao and Datta \(2014\)](#), the sent-LDA model is parametric, and the process can be resumed easily. Each document is composed of a set of topics and the list of the words (known as vocabulary) composing each topic, follow a Dirichlet distribution as well as the proportion of topics in each document. Furthermore, the particularity of the sent-LDA is that for each sentence in a document, the assignment of topic follows a Multinomial as well as the words in a sentence.

4.2.3. *Our Process*

The aim of this algorithm is to analyse the thematic content of our sample containing 1848 ICOs whitepapers. The output is basically composed of several lists of words containing our different subjects. Certain conditions must be met in order to obtain satisfactory results and this is what we are going to develop.

The first step is the lemmatisation¹³ of our ICOs whitepapers. It consists of grouping different form of a word into a single item [Malik \(2019\)](#). Subsequently, we needed to clean the data by removing punctuation, stop words (special character), white space and non-English words¹⁴ that could be found in the sample. The whitepapers list a lot of 'non-words', like website names, token names and other technical terms. As we are trying to determine the main topics of these whitepapers, we want words that have general meaning so sensible patterns can be spotted. So, first the input data needs to get rid of these words.

Then, we specify some inputs such as the number of topics. It can be determined automatically but the model is very sensitive to this parameter. Literature informs us that there are two approaches, a statistical fit or a substantive fit. The first use perplexity while the second focus on human judgment (semantic coherence). Using perplexity is criticised by [Grimmer and Stewart \(2013\)](#) highlighting that there is a negative relation between perplexity results and the substantive information provided. So, we opt for the second approach and we fix the optimal number of topics at 50. As [Bao and Datta \(2014\)](#) explain, the perplexity is steady when the number of chosen topics is higher than 30. For our analysis, we opt for 50 because of the clarity, meaning and sharpness of our given topics. Once these

¹³We use lemmatisation instead of stemming because the problem with stemming is that the words can become nonsensical. A better alternative to stemming is lemmatization.

¹⁴By using a dictionary with all the list of all English words, we were able to check if the words in the document are English words or not. Then, we get rid of the non-English words.

assumptions are drawn, the number of topics defined and few others running parameters in place, we can run the process. The model identifies and categorizes one topic per sentence and we get 50 lists of vocabulary words that need to be labelled. It is possible to label them automatically as described by [Mei, Shen, and Zhai \(2007\)](#) but it's common in the topic model research to do it manually to obtain a higher quality [Chang and Chien \(2009\)](#). Our labelled topics can be viewed in table 10.

4.2.4. Models

In this paper, we used two equations that explain our thematic analysis and we draw regressions based on our four dependent variables. The model with topics (see equation 1) allows us to determine the added value of the thematic content while the model without topics (see equation 2) is used as a point of comparison.

$$y = \sum_{k=1}^K \text{Topics } k + \text{ICOControls} + \varepsilon \quad (1)$$

$$y = \text{ICOControls} + \varepsilon \quad (2)$$

where ε are Newey–West standard errors, corrected for heteroskedasticity and autocorrelation.

To obtain unbiased estimates we include control variables based on past literatures (e.g., [Adhami et al. \(2018\)](#); [Amsden and Schweizer \(2018\)](#)). All of the variables are described as follow:

a) *Dependent variables*

TokenTradedHistorical: is our binary variable for ICO success which represent the token tradability post ICO as [Amsden and Schweizer \(2018\)](#) recommend. Furthermore, the wish of all token investor is that the token is listed on an exchange platform (e.g., coinmarketcap.com) and is actively traded. Determining the success of an ICO is not as simple as in the case of crowdfunding. We cannot simply take the fact that the desired amount (the soft cap) has been reached as it is not mandatory for an ico to define one.

AmountRaised (log.): It is the total amount raised in USD by the ICO project for the time period defined. This variable is usually used in financial research [Mollick \(2014\)](#). As in previous studies, we use a logarithm transformation for the skewness of the variable [Anglin, Short, Drover, Stevenson, McKenny, and Allison \(2018\)](#).

ssgarchvol: It is the price volatility of the tokens on secondary exchange. As most of the tokens can be traded against other tokens or fiat currencies on secondary platforms, the price volatility is an interesting element to take into consideration [Benedetti and Kostovetsky \(2018\)](#).

AbnormalIR: Calculated by subtracting the market cap weighted price returns trading on the same day, with the market cap weighted price returns of all other tokens trading on the same day (e.g., returns - market returns).

b) *Control variables*

PreICODummy: It indicates if there is a presale of the token before the official launch of the ICO. We have seen in the introduction that a presale can be beneficial because it allows large investors to buy tokens at discounted prices before other smaller investors with the aim of stimulating them into the project [Fisch \(2019\)](#). However, a presale is not always viewed positively to investors, as [Amsden and Schweizer \(2018\)](#) explain, this may reflect the company's lack of confidence in the ICO's success and even a lack of funds on the part of the ICO's founders.

BonusDummy: Some ICOs offer bonuses to investors which is translated by discount tokens. The idea behind that is to attract more easily investors for the pre ICO or even during the main offering and thus in fine, gathering higher amounts. However, [Amsden and Schweizer \(2018\)](#) and [Adhami et al. \(2018\)](#) didn't find any relation between the amount raised and the availability of bonuses.

EthereumBlockDummy: Since the creation of the first ico in 2013, an ICO development has been increasingly simplified¹⁵. The creation of Ethereum in 2015 and Waves in 2016, let anyone to create easily their ICO thanks to smart contracts. This ERC20 is a revolution which allow the expansion of the industry due to the reduction of technical requirements. For instance, when an investor wants to

¹⁵ICO teams have the choice to create their own blockchain or to use code as a starting point to build their own.

buy a token on the Ethereum platform, he can use standard wallet. As the literature demonstrates, there is a significant relationship between ICO success, the amount raised and the Ethereum based ICOs [Amsden and Schweizer \(2018\)](#) and [Fisch \(2019\)](#).

WhiteListKYCDummy: Here we refer to two elements which are whitelists¹⁶ and Know your customer, KYC¹⁷. Both of them indicates regulatory compliance with the aim of reducing fraud (e.g., money laundering). This goes in conflict with the definition of cryptocurrencies, which allows buyers to remain anonymous. This in any case does not seem to affect the success of the ICO as demonstrated by [Amsden and Schweizer \(2018\)](#) who don't spot any relationship between ICO success and the availability of whitelists and KYC processes.

number_of_team: We look at the number of persons included in the ICO team. The characteristics of the team is important for the quality of the ICO as already demonstrated in the scientific literature [Muzyka, Birley, and Leleux \(1996\)](#). What also counts is the number of people who make it up because an ICO requires a lot of work and operations. In addition, it has been documented that ICOs comprising several members perform better than that achieved by a single individual [Chandler, Honig, and Wiklund \(2005\)](#). The more the people in a team, the larger the confidence in the project [Amsden and Schweizer \(2018\)](#).

TaxHaven: This variable is used to check if ICOs come from a country categorised as tax haven according to the list of 52 countries defined by [Hines \(2010\)](#). [Amsden and Schweizer \(2018\)](#) didn't find any significant relationship between the variable tax haven and the amount raised but they used the OECD list which has been criticized. The one that we use is less influenced by internal biases as cited by [Palan \(2009\)](#).

softcapDummy and *hardcapDummy*: An ICO can be capped or uncapped and the difference between them is that a limit on the amount received is defined for capped projects. If too many cryptocurrencies are raised, the excess will be returned to the investors. Interestingly, an uncapped is not always seen positively. In fact, as the ICO has an unlimited goal amount, this reduces the token value. However, in

¹⁶A whitelist is a registration requirement in order to participate in the ICOs [Kumar \(2019\)](#)

¹⁷KYC is a process where the ICO team verify the identity of the token buyers

both cases a minimum amount (soft cap) is specified and if that minimum amount is not reached, the ico is cancelled and the cryptocurrencies are returned to the investors. The hard cap is the maximum amount that the ICO can raise and therefore only makes sense for capped projects [Bourveau et al. \(2018\)](#).

NumCurrencies: ICOs may accept different currencies for the token exchange. This may lead to help the process by reducing the number of steps needed. As [Howell, Niessner, and Yermack \(2018\)](#) indicate, it sends a positive signal of technical capacity of the team because it requires some blockchain expertise. In our analysis, we thus include the number of accepted currencies by the ICO.

BitcoinPriceAtStart: We use the Bitcoin price at the start of the ICO as it is a market condition. The Bitcoin is the first cryptocurrency and has a market cap of \$616 billion in Jan 2021 which is the highest market cap compared to others. As a point of comparison, Ethereum, the second largest, is around \$200 billion. The amount raised in an ICO have a significant relationship with the value of the Bitcoin. The reason is that this initial cryptocurrency is often used as the main source of payment when purchasing the token. In addition, if the Bitcoin price increase, it will positively impact the overall market and thus in fine increase the amount raised as cited by [Fisch \(2019\)](#).

FiatAcceptingDummy: This refers to the fact that when the tokens are exchanged, the ICO team receives cryptocurrencies against the token. However, it is also possible that in some cases the ico team will accept common currencies produced by central banks (as known as fiat currencies). The reason behind that is to reach a higher number of investors than only crypto investors. The risk is that it sends a negative signal as it may translate a lack of confidence in the offering [Amsden and Schweizer \(2018\)](#). Another issue is that if the soft cap is not reached, the smart contract will not be willing to reimburse the investor.

4.3. Empirical Results

4.3.1. Summary Statistics

We show the statistics of our dependent and control variables in table 1,2, including the number of observations, the mean, the standard deviation, the minimum and the maximum as well as the median value.

Table 1.: Table of Descriptive Statistics

Dependent Variables	N	Mean	St. Dev.	Min	Median	Max
TokenTradedHistorical	1,848	0.256	0.437	0	0	1
AmountRaised	888	19,549,732	157,596,489	420	5,000,000	4,197,956,135
ssGarchVol	416	1.117	3.651	0.00000	0.227	70.223
AbnormalIR	425	0.184	0.927	-0.859	0.017	15.935
<hr/>						
Control Variables						
PreICODummy	1,848	0.467	0.499	0	0	1
BonusDummy	1,848	0.141	0.348	0	0	1
EthereumBlockDummy	1,848	0.871	0.335	0	1	1
WhiteListKYCDummy	1,848	0.437	0.496	0	0	1
number_of_team	1,737	12.790	7.553	0.000	11.000	73.000
TaxHaven	1,848	0.312	0.463	0	0	1
hardcapDummy	1,848	0.759	0.428	0	1	1
softcapDummy	1,848	0.534	0.499	0	1	1
NumCurrencies	1,848	1.926	1.718	0	1	12
BitcoinPriceAtStart	1,848	7,909.006	3,197.770	422.843	7,755.490	19,475.800
FiatAcceptingDummy	1,848	0.016	0.126	0	0	1

The total number of observations is 1,848 for the sample as well as for the TokenTradedHistorical variable while we observe lower values for the other dependent variables. When running regressions, it automatically drops observations with missing value. The model with AmountRaised as dependent variable has smaller sample compared to TokenTradedHistorical because AmountRaised values are not available for all observations. By looking at ssGarchVol and AbnormalIR, we only have details for the tokens that were issued, so the sample size in those regressions is even smaller.

Table 2.: Table of Descriptive Statistics Cont.

Independent Variables	N	Mean	St. Dev.	Min	Median	Max
Successful & non-successful ICOs						
Top five topics						
Team experience (36)	1,848	12.686	17.181	0	5	161
Token sale (11)	1,848	11.115	11.744	0	8	97
ICO development (32)	1,848	10.865	14.053	0	7	125
Blockchain revolution (8)	1,848	9.944	14.591	0	5	147
Community platform (21)	1,848	9.695	17.611	0	3	273
Bottom five topics						
Performance uncertainty (30)	1,848	1.376	4.435	0	0	100
Tone at the top (25)	1,848	2.162	10.540	0	0	362
Future performance statement (37)	1,848	2.423	6.378	0	0	76
Risk understanding (23)	1,848	3.022	7.960	0	0	112
Legal disclosure (35)	1,848	3.064	6.615	0	0	96
Successful ICOs						
Top five topics						
Blockchain transaction (1)	473	16.101	45.614	0	1	435
Team experience (36)	473	15.139	19.828	0	7	136
Platform development (9)	473	14.725	31.661	0	3	326
Platform content creation (13)	473	12.082	41.128	0	0	384
Blockchain solution (24)	473	11.968	18.082	0	5	182
Bottom five topics						
Performance uncertainty (30)	473	1.076	2.783	0	0	41
Product & service production (47)	473	1.634	9.655	0	0	131
Tone at the top (25)	473	2.086	5.927	0	0	76
Blockchain for film industry (29)	473	2.558	12.162	0	0	212
Future performance statement (37)	473	2.640	6.654	0	0	50
Non-Successful ICOs						
Top five topics						
Team experience (36)	1,375	11.842	16.091	0	5	161
Token sale (11)	1,375	11.348	11.709	0	8	97
ICO development (32)	1,375	10.560	13.528	0	6	110
Blockchain revolution (8)	1,375	9.496	13.686	0	5	106
Trade platform (15)	1,375	9.143	29.273	0	0	270
Bottom five topics						
Performance uncertainty (30)	1,375	1.479	4.871	0	0	100
Tone at the top (25)	1,375	2.188	11.715	0	0	362
Future performance statement (37)	1,375	2.348	6.280	0	0	76
Legal disclosure (35)	1,375	2.960	6.159	0	0	60
Risk understanding (23)	1,375	3.072	8.371	0	0	112

This table shows the TokenTradedHistorical variable only. Table 9 provides the details

As we can see in table 2, we divide our sample into successful and non-successful ICOs. To establish the ranking, we sort the topics based on their mean. At this time of the analysis, it's too early to decide which topics are the most important. Naturally, we will have to run regressions to know their importance but there are still some insights that we can already point out. We are now able to analyse the most as well as the less discussed topics. The number of observation (N) refers to the number of ICOs that are successful & non-successful, successful, and non-successful. The mean is the number of sentences that are referring to that particular topic divided by the total number of sentences. The last column represents the maximum number of sentences present in a whitepaper in a particular topic.

We can therefore observe that there are 473 successful and 1,375 non-successful ICOs in our sample. Team experience (36) is the most favourite topic discussed into successful as well as non-successful ICO. Interestingly, all the less discussed topics refers to the risk category. Concerning successful ICOs, Blockchain transactions (1) is the most cited. This topic provides technical information about the creation of blocks. For the non-successful, Team experience (36) is the most used. In both cases, Performance uncertainty (30) is the less popular.

4.4. Topics in ICO Whitepapers

Our first research question (**RQ1**) deals with the identification of the topics in ICO whitepapers. To determine them, we label the 50 topics and as our number of topics is quite large, we classify them into seven relevant categories. On the basis of the words, it is possible to associate certain topics and similar subjects are combined together. Then, we give a sub heading to each group of topics. The categories allow us to facilitate the interpretation.

In order to classify the different topics, we had to label them beforehand. To do this, we use different set of information. First, we employ lists of words representing each topic which provide quick and clear meanings. To be sure of the identification, we control with another spreadsheet using 300 most represented sentences, a third using bigrams (a combination of two words that are most relevant in the 100 sentences for each topics) and a last using cosine similarities (the top 100 sentences associated with each specific topics). Basically, the idea is that with these four different types of representation, we tried to identify what are the most relevant meaning of each topic. Table 3 provide a quick overview and a more detailed description of each topic is available in table 10.

Table 3.: Categories.

Blockchain:

This category relates to blockchain as a decentralized system, we include among others its innovative aspect, the solutions it provides for transaction, its technical aspects and its security.

Blockchain transaction (1)	<i>Related to Blockchain transaction via block creation.</i>
Blockchain revolution (8)	<i>Related to blockchain revolution</i>
Loan agreement (12)	<i>Related to loan availability.</i>
Smart contract description (17)	<i>Related to transaction via Smart contract.</i>
Blockchain solution (24)	<i>Related to Blockchain solutions</i>
Blockchain based offering (26)	<i>Related to Blockchain-based offerings</i>
User data encryption (31)	<i>Related to user data encryption</i>
Smart contract technology (33)	<i>Related to the smart contract technology</i>
Blockchain dispute resolution (34)	<i>Related to blockchain dispute resolution</i>
Blockchain security (38)	<i>Related to blockchain security</i>
Energy consumption (48)	<i>Related to energy consumption concerns</i>

Token:

This category relates to tokens as digital assets that can be transferred without duplication between two parties. We include token description, token utility and exchange method.

Digital wallet (2)	<i>Related to currency exchange via digital wallet.</i>
Crypto Token usage & value (3)	<i>Related to functions a token serve in order to gain value.</i>
Payment system (4)	<i>Related to payment of token exchange</i>
Token sale (11)	<i>Related to token sale</i>
Token (14)	<i>Related to token exchange</i>
Trade platform (15)	<i>Related to platform trading digital currency.</i>
Transaction fee (20)	<i>Related to transaction fees</i>
Future performance statement (37)	<i>Related to future performance disclosure</i>
Token exchange (45)	<i>Related to possibility of token exchange.</i>
Token trans. terms & cond. (50)	<i>Related to terms and condition for token transaction.</i>

Team:

This category relates to the team characteristics of ICO projects. Based on thematic content, token issuers can be distinguished by their experience, their educational level, etc.

Educational level (19)	<i>Related to team educational level</i>
Tone at the top (25)	<i>Related to describing top management</i>
Team experience (36)	<i>Related to team experience</i>

Law & regulation:

This category relates to ICOs potential regulations. While the industry is unregulated, it is not impossible to find some disclosure related to legal aspects.

Terms and conditions (16)	<i>Related to terms and conditions of whitepapers.</i>
Legal disclosure (35)	<i>Related to legal disclosure</i>
Intellectual property (39)	<i>Related to Intellectual Property Interface</i>
Law protection disclosure (46)	<i>Related to law protection disclosure</i>
ICOs regulation (49)	<i>Related to customer protection, fraudulent ICOs, etc.</i>

Blockchain in specific sector:

This category relates to blockchain technology used in specific sector. The Blockchain technology revolutionized some industry such as the real estate, the medical sector, etc.

Real estate (7)	<i>Related to real estate investment.</i>
Blockchain health (27)	<i>Related to Blockchain technology in the health sector.</i>
Blockchain based telecom (28)	<i>Related to marketplace enabling stakeholders to transact and exchange value with trust based upon smart provenance.</i>
Blockchain for film industry (29)	<i>Related to distribution complexity, management of creative rights and digital piracy.</i>
Blockchain in energy industry (44)	<i>Related to Blockchain technology in the energy sector.</i>

Risk:

This category relates to potential risks occurring with ICOs. This relies on uncertain future performance, and risk prevention.

Risk prevention (6)	<i>Related to risk</i>
Risk understanding (23)	<i>Related to acknowledgement of risk understanding</i>
Performance uncertainty (30)	<i>Related to performance uncertainty and risk disclosure</i>

ICO development:

This category relates to ICO creation and expansion. It includes the business offer, the platform development and the community creation and loyalty building.

Market desc. & expect. (5)	<i>Related to market description and expectation.</i>
Platform development (9)	<i>Related to platform development.</i>
Mobile app (10)	<i>Related to mobile app development</i>
Platform content creation (13)	<i>Related to content creation for users on main platform.</i>
Product/ service delivery (18)	<i>Related to business offer and customer satisfaction</i>
Community platform (21)	<i>Related to community development</i>
User review (22)	<i>Related to user review (ratings) and potential reward</i>
ICO development (32)	<i>Related to Marketing, Community Building</i>
Crypto gaming (40)	<i>Related to Crypto Gaming as a means to make money.</i>
AI and ML (41)	<i>Related to platforms using artificial intelligence</i>
Connect people (43)	<i>Related to connect people</i>
Product & service production (47)	<i>Related to product and services production</i>

4.5. The Thematic Content Matters

Our second research question (**RQ2**) aims at showing that our models based on TokenTradedHistorical, AmountRaised, ssGarchVol and AbnormalIR are potentially improved by topic variables. AmountRaised is the ICO outcome. Volatility is what happened after finishing the ICO when the tokens are raised. It's more post ICO performance.

By including topic variables in our regressions, we observe that some of our models significantly improve (see table 4). We compute the different Adjusted R squared for the linear model and Pseudo R squared for the logistic model to see the difference between the two models, one with the 50 topics and one without them. By including this topic information, we can see that Adjusted R²/Pseudo R² on TokenTradedHistorical has been improved by 90.77%, on AmountRaised by 57.33%, on ssGarchVol by 164.71% and on AbnormalIR by 533.33% .

Table 4.: Multiple R-squared

		Adjusted R ² /Pseudo R ²	Loglikelihood	lr test(Chisq)
TokenTradedHistorical	w/o topic	0.065	-929.32	
	w/ topic	0.124	-870.07	118.5***
AmountRaised	w/o topic	0.075	-1691.9	
	w/ topic	0.118	-1645.5	92.672***
ssGarchVol	w/o topic	-0.017	-1084.8	
	w/ topic	-0.045	-1062.6	44.407
AbnormalIR	w/o topic	0.006	-530.64	
	w/ topic	0.038	-496.59	68.114*

0 (***) 0.001 (**) 0.01 (*) 0.05 (.) 0.1 (.) 1

We are reporting goodness of fit, adjusted R squared for the linear model, Pseudo R squared for the logistic model, loglikelihood and model comparison based on loglikelihood ratio test.

To demonstrate the impact of the topics on the dependent's variables TokenTradedHistorical, AmountRaised, ssGarchVol and AbnormalIR, we focus on the adjusted/pseudo R Squared. We can see on the table above that adding topics to our regression model has an impact on those dependents variables. The aim is not to predict what the most important topics are, but to

see if our topic variables add value. We see that yes, it has an impact and in a second time, we use Loglikelihood tests to determine the influence that independent variables have on TokenTradedHistorical, AmountRaised, ssGarchVol and AbnormalIR thanks to a regression study. We want the probability of Chisq to be less than zero. Chisqr is the statistic of the difference between two models. If the probability of Chisq is greater than '*' 0.05 or '.' 0.1, then there is no significant difference between the two models and there is no point adding topic variables. We have linear models and logistic models. One of the variables is a binary dependent variable. Our logistic model is with the dependent TokenTradedHistorical variable which is a binary dependent variable.

For the first model, we observe that log likelihood results are less important (-929.32) for the model without topics than for the model with topics (-870.07). The larger the number, the better it is. Our second model is thus better. Concerning the Chisq, our degree of freedom is 50. Our Chisq for the model is 118.5. if we plot that model, the probability of that 118.5 is 1.727e-07 ***. Which means that the probability that it's actually zero is like that small. There is very strong evidence of rejecting the null hypothesis of no fit. It means good fit.

The Second model comparing Amount Raised with or without topics, reinforces us in the idea that there is strong evidence of rejecting the null hypothesis. The log likelihood is -1691.9 and for the second it is -1645.5. Concerning the Chisq, our degree of freedom is also 50. Our Chisq for the model is 92.67. The probability of that 92.67 is 0.0002318***.

The third model concerning volatility is not significant. The probability of Chisq, which is 0.70 is greater than '*' 0.05 or '.' 0.1, so there is no significant difference between the two models and there is no point adding topic variables. For volatility it is normal that we don't get much result in the sense that volatility is a measure with a lot of errors in it. We also have fewer variables included in our model, so we get a poorer fit from our model. Secondly, there are more measurement uncertainties. From then on, it is more complicated to get meaningful results for volatility. However, we still observe a slight increase, but it is not statistically significant.

Finally, the AbnormalIR model is significant at 0.01*. The loglikelihood is -530.64 and for the second it is -496.59. Concerning the Chisq, our degree of freedom is also 50. Our Chisq for the model

is 68.114. The probability of that 68.114 is 0.04507*. We see that the topics do not really have huge impact on this variable but it still slightly improves the model.

The first conclusion we can therefore draw from our analysis is that the thematic content of ICO whitepapers is informative except for volatility due to the two reasons mentioned above. This conclusion reinforces us in the idea that the thematic content of ICO white papers potentially contains informational value to decrease information asymmetry to investors.

4.6. Informative Topics in Whitepapers

Our last research question (**RQ3**) focus on communicating specific results by answering which topics are important for TokenTradedHistorical, AmountRaised, Volatility and AbnormalIR. During our analysis, we collected 50 themes that we reorganized into seven clusters. There are three main categories that we have been able to identify: positive significant, negative significant and non-significant. The objective here is to discuss the various topics that help to reduce information asymmetry and in the same way reduce volatility. Specific topics analysis be viewed in table 11.

We are reporting in table 5 coefficients and standard errors from our regression analysis based on subject categories. We observe that the main topics of interest according to our results distinguish themselves in two parts. First, we note that when referring to blockchain, it impacts positively the success, the amount raised, and it reduces the volatility. Secondly, we observe that whenever ICO white papers refer to the law category, it decreases the volatility. Interestingly, the success variable is negatively impacted as well. As ICOs are decentralised, we assume that laws are not easily welcome by some investors which consider them as hurdles. This result is quite questioning because on the one hand we notice that legal aspects in whitepapers have negative impacts on success and on the other side a company must be transparent. We therefore have a delicate situation where the control of this decentralized industry is made impossible. Then, it also depends to what extent it is integrated into business activities. Companies do not want to invest in scams or even worse be linked to them.

Table 5.: Statistical models based on categories

Category	TokenTradedHistorical	log(1+AmountRaised)	ssGarchVol	AbnormalIR
(Intercept)	-0.282 (0.872)	13.448*** (0.487)	1.458*** (0.517)	0.230* (0.119)
Blockchain	0.004*** (0.001)	0.003*** (0.001)	-0.003** (0.001)	-0.001 (0.001)
Token	0.001 (0.001)	0.002** (0.001)	0.002* (0.001)	0.000 (0.001)
Team	-0.000 (0.002)	-0.003 (0.002)	0.010 (0.010)	-0.001 (0.001)
Law	-0.005*** (0.002)	0.001 (0.002)	-0.007** (0.003)	-0.002 (0.001)
Risk	-0.001 (0.004)	0.004 (0.003)	-0.002 (0.005)	0.001 (0.002)
ICodevelopment	0.001 (0.001)	0.001 (0.001)	0.000 (0.001)	-0.000 (0.000)
Blockchainsector	-0.002 (0.002)	0.001 (0.001)	0.001 (0.002)	0.002 (0.001)
Control variables	yes	yes	yes	yes
Year fixed effects	yes	yes	yes	yes
AIC	1824.572			
BIC	1950.030			
Log Likelihood	-889.286			
Deviance	1778.572			
Num. obs.	1728	851	397	404
R ²		0.155	0.027	0.052
Adj. R ²		0.132	-0.030	-0.003
RMSE		1.716	3.779	0.916

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 11 and table 12 provide the details.

Not all topic categories are value relevant to investors. In fact, there is only one, related to blockchain security and technology which positively impact the success of an ICO. This is in line with previous literature that emphasise the importance of technical whitepapers. The blockchain and token category contain several technical information such as smart contracts, token description and security. When referring to technology, as [Fisch \(2019\)](#) has already demonstrated, it reassures investors and it impact positively the amount raised. However, we point out that technical aspect also impacts the success. Also, [Fisch \(2019\)](#) find that shorter whitepapers don't focus on technical aspect. We are thus further interested in the thematic content to spot any ICOs that could be categorize as scam. In table 6, we point out that there is no significance between our independent variables and our scam related variable. The thematic content is thus not appropriate to detect any fraudulent ICO.

Table 6.: ICO Scam

	Scam Dummy
(Intercept)	- 2.523** (1.075)
Blockchain	0.001 (0.001)
Token	0.002 (0.001)
Team	0.001 (0.004)
Law	-0.003 (0.003)
Risk	0.001 (0.005)
ICOdevelopment	-0.000 (0.001)
Blockchainsector	0.001 (0.002)
Control variables	yes
Year Fixed effects	yes
AIC	1108.889
BIC	1234.348
Log Likelihood	-531.445
Deviance	1062.889
Num. obs.	1728

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

5. Additional Analysis - Simpson Diversity

In this section, we analyse the distribution of the topics to determine if whitepapers are more global by discussing different topics or more centralized. We use the so-called Simpson Diversity, which is an aggregate measure for the 50 topics. For instance, a whitepaper composed of N sentences will have a higher diversity if the N sentences are related to different topics. Conversely, a low diversity score would indicate that sentences talk about the same topic. By running regressions including this diversity variables, a positive value would indicate that higher diversity scores react positively with the success.

Table 7.: Diversity

	TokenTradedHistorical	$\log(1+\text{AmountRaised})$	ssGarchVol	AbnormalIR
Diversity	-1.004** (0.447)	-1.249** (0.524)	1.323** (0.545)	0.153 (0.186)
Control variables	yes	yes	yes	yes
Year fixed effects	yes	yes	yes	yes
AIC	1822.294			
BIC	1953.207			
Log Likelihood	-887.147			
Deviance	1774.294			
Num. obs.	1728	851	397	404
R ²		0.161	0.029	0.052
Adj. R ²		0.138	-0.030	-0.005
RMSE		1.711	3.780	0.917

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In the above table, we determine the effect of diversity on our dependent variables. We note that Diversity reacts negatively with the success and the amount raised variables which translates that higher diversity scores in whitepapers are hurdles for investors. Not surprisingly, this aggregate variable positively impacts the volatility as a higher score would indicate that whitepapers are more global than focused on few topics.

6. Sensitivity Checks

In this last section, we illustrate that the results do not change significantly without grouping. We know from our analysis that technical topic categories are value relevant to investor and we are interested in the variation that we could have induced in the research. This is why we control the topics only. There are few topics for the success variable that are significant and positive in the "Blockchain" category. The first refers to Blockchain & transaction properties (1), an illustrative sentence of this topic provided from the Quant-Network ICO is "*Blockchain always guarantees atomicity since a transaction only be present or absent in a block.*" The second topic entitled smart contract description (17) focus on security as On.Live's whitepaper illustrates "*Transactions between all parties are secured by the blockchain and Smart Contracts.*"

Table 8.: Significant Topics

	TokenTradedHistorical	AmountRaised	ssGarchVol	AbnormalIR
Blockchain transaction(1)	0.007** (0.003)	0.003** (0.002)	-0.002 (0.002)	0.000 (0.001)
Market desc. and expect.(5)	0.014** (0.006)	-0.006 (0.006)	0.006 (0.018)	0.029 (0.026)
Platform development(9)	0.008*** (0.002)	0.006** (0.003)	-0.001 (0.002)	-0.001 (0.001)
Loan agreement(12)	0.005** (0.002)	0.004** (0.002)	-0.002 (0.004)	-0.001 (0.001)
Trade platform(15)	0.003* (0.002)	0.003 (0.002)	-0.002 (0.005)	-0.000 (0.001)
Terms & cond.(16)	-0.018*** (0.007)	0.001 (0.007)	-0.039 (0.032)	-0.008 (0.005)
Smart contract desc.(17)	0.012*** (0.004)	0.007*** (0.002)	0.004 (0.004)	-0.000 (0.001)
Telecom(28)	-0.007** (0.004)	0.000 (0.004)	0.040 (0.036)	-0.007 (0.006)
Legal disclosure(35)	0.022* (0.012)	0.005 (0.011)	-0.001 (0.019)	-0.003 (0.008)
Team experience(36)	0.008** (0.004)	0.002 (0.004)	0.029 (0.030)	0.002 (0.003)
Token exchange(45)	-0.012* (0.007)	-0.004 (0.004)	-0.004 (0.013)	0.000 (0.004)
Token transact.(50)	-0.010* (0.006)	-0.005 (0.006)	0.011 (0.009)	-0.003 (0.003)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

We are reporting significant topics for success only. Table 11 and table 12 provide the details.

Concerning the "law" category there are only two topics (entitled Terms and conditions (16) and Legal disclosure (35)) that are significant for the success variable. It refers to a notice that invite the reader to take some consideration about the whitepaper content. It warns the investors that the whitepaper shall not constitute an offer to sell or to buy any securities or tokens described in it and that there shall be any sale of these securities in a state where it could possibly be unlawful. In addition, the company indemnifies itself against any failure to meet future performance as described in the whitepaper. In fact, it warns that potential investors should conduct their own due diligence and not just rely on this document.

We notice that not all significant topics are technical. We even observe contradictory elements such as "token exchanges"(45) and "Token transaction"(50) which negatively impact the success while we consider them as technical topics. Interestingly, we find that "Team experience"(36) is positive significant which is linked to previous studies on team composition (Burns and Moro, 2018; Ahlers, Cumming, Günther, and Schweizer, 2015). Lastly, by analysing the volatility by topics (see table 11) we observe that there are only two topics that are significant and negative which highlights previous results (table 4) where volatility models with topics are not significantly higher.

7. Conclusion

There is substantial information asymmetry in the field of ICOs which is new and characterized as decentralized and unregulated. Thus, the task is not simple for investors to distinguish between promising and unpromising ICOs and for entrepreneurs to communicate on value relevant topics. We took a brand-new approach in the field of ICO success by analysing the context rather than financial and textual style variables as previous studies already carried out. The first conclusion is that the thematic content has information value. We then dig deeper and find that value relevant categories concern the security and the technical aspects of the blockchain related topics. In addition, we find that legal aspects of ICO whitepapers negatively impact the success of the project which is comprehensible as the industry is decentralized. We further indicate that while the thematic content has informational value for the success, this is not the case for detecting any fraudulent cases. Then, one variable combining all the 50 topic variables is diversity. It allows us to determine how diverse the elements are in the population. Results of the Simpson diversity indicate that a higher value of this

variable negatively impacts the success of the ICO and increases the volatility. Finally, by conducting sensitivity checks we notice that technical topics as well as non-technical impact the success. This translates that our independent variables do not explain completely the success of an ICO but provides value relevant insights to token owners as well as investors.

References

- Adhami, S., G. Giudici, and S. Martinazzi (2018). Why do businesses go crypto? an empirical analysis of initial coin offerings. *Journal of Economics and Business* 100, 64–75.
- Ahlers, G. K., D. Cumming, C. Günther, and D. Schweizer (2015). Signaling in equity crowdfunding. *Entrepreneurship theory and practice* 39(4), 955–980.
- Amsden, R. and D. Schweizer (2018). Are Blockchain Crowdsales the New ‘Gold Rush’? Success Determinants of Initial Coin Offerings. Available at SSRN 3163849.
- An, J., W. Hou, and X. Liu (2018). Initial coin offerings: Investor protection and disclosure. *Technical report, Working paper*.
- Anglin, A. H., J. C. Short, W. Drover, R. M. Stevenson, A. F. McKenny, and T. H. Allison (2018). The power of positivity? the influence of positive psychological capital language on crowdfunding performance. *Journal of Business Venturing* 33(4), 470–492.
- Aral, S., P. G. Ipeirotis, and S. J. Taylor (2011). Content and context: Identifying the impact of qualitative information on consumer choice. Available at SSRN 1784376.
- Bao, Y. and A. Datta (2012). Summarization of corporate risk factor disclosure through topic modeling.
- Bao, Y. and A. Datta (2014). Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science* 60(6), 1371–1391.
- Becht, M., P. Bolton, and A. Röell (2003). Corporate governance and control. In *Handbook of the Economics of Finance*, Volume 1, pp. 1–109. Elsevier.
- Benedetti, H. and L. Kostovetsky (2018). Digital tulips? Returns to investors in initial coin offerings. *Returns to Investors in Initial Coin Offerings (May 20, 2018)*.
- Bian, S., Z. Deng, F. Li, W. Monroe, P. Shi, Z. Sun, W. Wu, S. Wang, W. Y. Wang, A. Yuan, et al. (2018). Icorating: A deep-learning system for scam ico identification. *arXiv preprint arXiv:1803.03670*.
- Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *the Journal of machine Learning research* 3, 993–1022.
- Boreiko, D. and N. K. Sahdev (2018). To ico or not to ico—empirical analysis of initial coin offerings and token sales. Available at SSRN 3209180.

-
- Bourveau, T., E. T. De George, A. Ellahie, and D. Macciocchi (2018). Initial coin offerings: Early evidence on the role of disclosure in the unregulated crypto market. *Available at SSRN 3193392*.
- Brown, N. C., R. M. Crowley, and W. B. Elliott (2020). What are you saying? using topic to detect financial misreporting. *Journal of Accounting Research* 58(1), 237–291.
- Burns, L. and A. Moro (2018). What makes an ico successful? an investigation of the role of ico characteristics, team quality and market sentiment. *An Investigation of the Role of ICO Characteristics, Team Quality and Market Sentiment (September 27, 2018)*.
- Cerezo Sánchez, D. (2017). An optimal ico mechanism. *Available at SSRN 3040343*.
- Chandler, G. N., B. Honig, and J. Wiklund (2005). Antecedents, moderators, and performance consequences of membership change in new venture teams. *Journal of Business Venturing* 20(5), 705–725.
- Chang, Y.-L. and J.-T. Chien (2009). Latent dirichlet learning for document summarization. In *2009 IEEE international conference on acoustics, speech and signal processing*, pp. 1689–1692. IEEE.
- Chemmanur, T. J. and P. Fulghieri (2014). Entrepreneurial finance and innovation: An introduction and agenda for future research. *The Review of Financial Studies* 27(1), 1–19.
- Chuanjie, F., A. Koh, and P. Griffin (2019). Automated theme search in ico whitepapers. *The Journal of Financial Data Science* 1(4), 140–158.
- Dittmar, R. F. and D. A. Wu (2019). Initial coin offerings hyped and dehyed: an empirical examination. *Available at SSRN 3259182*.
- Dyer, T., M. Lang, and L. Stice-Lawrence (2017). The evolution of 10-k textual disclosure: Evidence from latent dirichlet allocation. *Journal of Accounting and Economics* 64(2-3), 221–245.
- Feng, C., N. Li, M. Wong, and M. Zhang (2019). Initial coin offerings, blockchain technology, and white paper disclosures. *Mingyue, Initial Coin Offerings, Blockchain Technology, and White Paper Disclosures (March 25, 2019)*.
- Fisch, C. (2019). Initial coin offerings (icos) to finance new ventures. *Journal of Business Venturing* 34(1), 1–22.
- Gompers, P. A. and J. Lerner (2004). *The venture capital cycle*. MIT press.
- Grimmer, J. and B. M. Stewart (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis* 21(3), 267–297.
- Hines, J. R. (2010). Treasure islands. *Journal of Economic Perspectives* 24(4), 103–126.
- Howell, S. T., M. Niessner, and D. Yermack (2018). Initial coin offerings: Financing growth with cryptocurrency token sales. Technical report, National Bureau of Economic Research.
- Howell, S. T., M. Niessner, and D. Yermack (2020). Initial coin offerings: Financing growth with cryptocurrency token sales. *The Review of Financial Studies* 33(9), 3925–3974.

-
- Humpherys, S. L., K. C. Moffitt, M. B. Burns, J. K. Burgoon, and W. F. Felix (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems* 50(3), 585–594.
- Kaal, W. A. and M. Dell’Erba (2017). Initial coin offerings: emerging practices, risk factors, and red flags. *Verlag CH Beck (2018), U of St. Thomas (Minnesota) Legal Studies Research Paper* (17-18).
- Kastelein, R. (2017). What initial coin offerings are, and why vc firms care. *Harvard Business Review* 24.
- Kravet, T. and V. Muslu (2013). Textual risk disclosures and investors’ risk perceptions. *Review of Accounting Studies* 18(4), 1088–1122.
- Kumar, A. (2019). What is the Whitelist registration in ICO? - Quora.
- legroju (2020, September). Legal and regulatory framework for blockchain.
- Loughran, T. and B. McDonald (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research* 54(4), 1187–1230.
- Malik, F. (2019, June). NLP: Text Processing Via Stemming And Lemmatisation In Data Science Projects.
- Mei, Q., X. Shen, and C. Zhai (2007). Automatic labeling of multinomial topic models. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 490–499.
- Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of business venturing* 29(1), 1–16.
- Momtaz, P. P. (2020). Entrepreneurial finance and moral hazard: evidence from token offerings. *Journal of Business Venturing*, 106001.
- Morris, D. Z. (2017). The SEC Filed Fraud Charges Against 2 Bitcoin-Inspired ICOs.
- Muzyka, D., S. Birley, and B. Leleux (1996). Trade-offs in the investment decisions of european venture capitalists. *Journal of Business Venturing* 11(4), 273–287.
- Palan, R. (2009). The history of tax havens. *History & Policy*.
- Rich, K. T., B. Roberts, J. Wall, and J. X. Zhang (2018). Toward an understanding of year-over-year changes in municipal management discussion and analysis disclosures. *Available at SSRN 3217301*.
- Shifflett, S. and C. Jones (2018). Buyer beware: hundreds of bitcoin wannabes show hallmarks of fraud. *Wall Street Journal* (May 17).
- Zetsche, D. A., R. P. Buckley, D. W. Arner, and L. Föhr (2017). The ico gold rush: It’s a scam, it’s a bubble, it’s a super challenge for regulators. *University of Luxembourg Law Working Paper* (11), 17–83.
- Zetsche, D. A., R. P. Buckley, J. N. Barberis, and D. W. Arner (2017). Regulating a revolution: From regulatory sandboxes to smart regulation. *Fordham J. Corp. & Fin. L.* 23, 31.
- Zhang, S., W. Aerts, L. Lu, and H. Pan (2019). Readability of token whitepaper and ico first-day return. *Economics Letters* 180, 58–61.

Zhu, X., S. Y. Yang, and S. Moazeni (2016). Firm risk identification through topic analysis of textual financial disclosures. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–8. IEEE.

8. Appendices

Table 9.: Table of Descriptive Statistics

Independent Variables	N	Mean	St. Dev.	Min	Median	Max
Blockchain transaction(1)	1,848	8.834	30.986	0	0	477
Digital wallet(2)	1,848	9.236	17.900	0	3	198
Crypto Token(3)	1,848	6.464	14.740	0	2	329
Payment system(4)	1,848	6.567	21.263	0	0	263
Market description(5)	1,848	6.824	13.032	0	3	345
Risk prevention(6)	1,848	4.119	12.214	0	0	136
Real estate(7)	1,848	5.961	22.496	0	0	335
Block. rev.(8)	1,848	9.944	14.591	0	5	147
Platf. dev.(9)	1,848	9.441	22.602	0	2	326
Mobile app(10)	1,848	6.107	13.394	0	2	170
Token sale(11)	1,848	11.115	11.744	0	8	97
Loan agreement(12)	1,848	4.514	22.289	0	0	364
Content creation(13)	1,848	8.644	30.617	0	0	387
Token(14)	1,848	7.339	16.377	0	3	269
Trade platform(15)	1,848	9.518	30.265	0	0	293
Terms and cond.(16)	1,848	6.165	10.716	0	2	128
Smart contract desc.(17)	1,848	7.126	18.333	0	1	339
Prod.& Serv. deliv.(18)	1,848	6.215	21.260	0	0	273
Team charact.(19)	1,848	6.749	13.203	0	1	166
Transaction fees(20)	1,848	5.587	17.235	0	0	251
Community platf.(21)	1,848	9.695	17.611	0	3	273
User review(22)	1,848	7.302	22.469	0	0	283
Risk underst.(23)	1,848	3.022	7.960	0	0	112
Blockchain solution(24)	1,848	9.055	15.638	0	3	182
Tone at the top(25)	1,848	2.162	10.540	0	0	362
Blockchain based offer.(26)	1,848	6.177	16.089	0	1	338
Blockchain health(27)	1,848	5.622	32.807	0	0	569
Blockchain telecom(28)	1,848	5.879	15.625	0	0	199
Blockchain film(29)	1,848	3.278	13.597	0	0	336
Performance uncertainty(30)	1,848	1.376	4.435	0	0	100
Data encrypt.(31)	1,848	8.566	20.422	0	2	290
ICO develop.(32)	1,848	10.865	14.053	0	7	125
Smart contr. tech.(33)	1,848	5.499	14.705	0	0	236
Blockchain dispute. (34)	1,848	4.123	16.906	0	0	313
Legal disclosure(35)	1,848	3.064	6.615	0	0	96
Team experience(36)	1,848	12.686	17.181	0	5	161
Future perform.(37)	1,848	2.423	6.378	0	0	76
Blockchain secu.(38)	1,848	3.360	18.976	0	0	365
Intellectual prop.(39)	1,848	4.075	14.954	0	0	221
Crypto gaming(40)	1,848	7.946	31.498	0	0	322
A.I. and M.L. (41)	1,848	7.107	24.438	0	0	286
Node descr.(42)	1,848	5.295	26.965	0	0	546
Connect people(43)	1,848	5.496	24.500	0	0	554
Blockchain energy(44)	1,848	4.693	25.209	0	0	426
Token exchange(45)	1,848	5.400	14.432	0	1	316
Law protection(46)	1,848	4.825	18.512	0	1	505
Prod.& serv. (47)	1,848	4.222	29.741	0	0	959
Energy consumpt.(48)	1,848	4.854	26.166	0	0	675
ICO's reg.(49)	1,848	6.474	15.517	0	2	240
Token transaction(50)	1,848	7.877	13.322	0	3	167

Table 10.: Labelling.

Topic	Label	Words
Topic 1	Blockchain transaction	[block, transaction, node, network, blockchain, consensus, time, proof, chain, algorithm] <i>ICO's relating to Blockchain transaction via block creation.</i> 731, hash signature block create transaction 1176 process create block record transaction 2462, proof work bitcoin block generation
Topic 2	Digital wallet	[currency, exchange, wallet, token, user, fiat, payment, platform, cryptocurrency, crypto] <i>ICO's relating to currency exchange via digital wallet.</i> 1254, send receive user currency crypto 1233, token standard easily wallet exchange 897, token main currency easily exchange
Topic 3	Crypto Token Usage and Value	[token, sale, platform, reward, fund, bounty, event, program, user, team] <i>ICO's relating to what functions a token serve in order to gain value.</i> 3111, participant reward eligible token 408, circulation token foundation ecosystem 2693, bonus program bounty campaign
Topic 4	Payment system	[payment, card, merchant, transaction, service, fee, token, system, customer, user] <i>ICOs related to payment of token exchange</i> 1790, easy simple convenient payment credit card 1104, minimize transaction fee global payment channel 826, outcome business accept token payment loyalty

Topic 5	market description and expectation	[market, billion, million, global, growth, industry, revenue, world, according, total] <i>ICO's related to market description and expectation.</i> 963, market value billion expect reach 2980, yearly revenue market currently billion 1334, global trade industry revenue trillion
Topic 6	risk prevention	[token, risk, platform, attack, business, including, development, technology, company, blockchain] <i>ICOs related to risk</i> 40, risk lack adoption application view investment value time 3091, risk hacker attack platform smart contract token 1175, lack interest impact development platform potential value
Topic 7	real estate investment	[investment, real, market, estate, investor, fund, property, asset, risk, project] <i>ICOs related to real estate investment</i> 3411, vision estate property especially foreign investor 1027, investment income fund activity reinvest system 2061, firm manage diverse real estate portfolio
Topic 8	blockchain revolution	[blockchain, technology, world, people, platform, market, business, cryptocurrency, industry, digital] <i>ICOs related to blockchain revolution</i> 3225, blockchain aspect technology market cryptocurrencies 1946, increase adoption blockchain technology cryptocurrency 2241, blockchain token economy internet value
Topic 9	Platform development	[application, platform, service, user, network, blockchain, system, data, decentralized, software] <i>ICO's related to platform development.</i> 3743, core platform blockchain development easy 2702, platform network blockchain base ecosystem 1608, platform decentralize network user blockchain

Topic 10	Mobile app - development and download	[user, app, mobile, application, wallet, platform, version, android, store, available] <i>ICOs related to mobile app development</i> 857, internet network app development android 436, user interact mobile wallet smartphones 1627, available platform apple google android
Topic 11	Token sale	[token, sale, total, bonus, million, coin, amount, cap, period, price] <i>ICOs related to token sale</i> 3770, token sale bonus 1220, token sale period 2929, token sale price
Topic 12	Loan agreement	[loan, credit, fund, interest, borrower, token, risk, platform, lender, bank] <i>ICO's related to loan availability.</i> 1032, neutralize default risk credit diversification lender loan 2082, loan usually build benefit allow investor money 2071, company seek conduct token sale provide loan
Topic 13	Platform - Con- tent creation	[content, platform, user, advertising, medium, video, social, creator, ad- vertiser, digital] <i>ICO's related to content creation for users on main platform.</i> 72, user create experience monetize content application 1688, payment option platform allow stream content 2241, platform distribute digital entertainment resource user
Topic 14	Token	[token, value, price, market, coin, currency, exchange, rate, time, in- crease] <i>ICOs related to token exchange</i> 3272, increase token price produce real currency 1384, low sell offer price token exchange 843, constant price matter coin exchange value

Topic 15	Trade platform	[exchange, trading, market, platform, trader, user, trade, crypto, asset, cryptocurrency] <i>ICO's related to platform allowing to buy, sell and trade digital currency.</i> 1747, plan future offer buy sell trade 2276, combine unique solution allow exchange market 4299, create effective trade investment exchange program
Topic 16	Terms and conditions	[information, token, paper, white, whitepaper, document, term, legal, sale, please] <i>ICO's related to terms and conditions of whitepapers.</i> 270, legal consideration please read section carefully 2514, note participate please read term condition 1025, legal information section whitepaper
Topic 17	Smart contract description	[contract, smart, token, blockchain, code, platform, asset, protocol, system, transaction] <i>ICO's related to transaction via Smart contract.</i> 3829, protocol unique asset blockchain token 1417, incorporate smart contract blockchain business 2486, random generation smart contract technology
Topic 18	Product/service delivery	[product, data, service, customer, consumer, information, chain, platform, user, business] <i>ICOs related to business offer and customer satisfaction</i> 3825, item feature meet quality requirement return seller 3706, client check leave feedback service allow receive 458, transaction seller buyer receive item arrive destination
Topic 19	Team characteristics - educational level	[university, business, director, science, company, degree, technology, international, management, member] <i>ICOs related to team educational level</i> 3549, hold communication medium study university 3425, graduate computer software engineer university 1688, professional experience work international law

Topic 20	transaction fees	[cost, service, fee, customer, business, platform, user, travel, time, price] <i>ICOs related to transaction fees</i> 1502, low fee charge result substantial saving 2787, flat fee guarantee low transaction cost 1622, obvious fee intermediary charge service high
Topic 21	Community platform	[platform token, user, community, ecosystem, network, project, business, service, value] <i>ICOs related to community development</i> 922, ecosystem create enable developer extend platform 3050, technical company create loyalty program platform 1317, allocate token affiliate system grow community
Topic 22	User review	[user, content, reward, platform, token, review, system, social, community, network] <i>ICOs related to user review (ratings) and potential reward</i> 123 trust review reward high amount standard 2423 rating reward system rate writer pay 3104 pay increase rating receive positive review
Topic 23	risk understanding	[token, loss, risk, including, party, purchaser, damage, liability, company, warranty] <i>ICO related to acknowledgement of risk understanding</i> 1929, risk lose access token due loss private key store wallet vault password 2117, extent website information service whitepaper provide free charge liable loss damage nature 2295, purchaser token acknowledge understand agree purchase obtain transfer usage platform tax consequence

Topic 24	Blockchain solutions	<p>[blockchain, transaction, data, technology, system, network, decentralized, distributed, ledger, platform]</p> <p><i>ICOs related to Blockchain solutions</i></p> <p>3126 blockchain solution ensure transaction detail check</p> <p>3865 currently create decentralize platform blockchain solution</p> <p>900 transparency blockchain technology ensure great consumer</p>
Topic 25	Tone at the top	<p>[officer, chief, executive, security, van, paper, white, information, person, director]</p> <p><i>ICOs related to describing top management</i></p> <p>3091, chief executive officer blockchain developer operation</p> <p>1334, chief executive officer mind set business</p> <p>3329, chief technology officer act head development</p>
Topic 26	Blockchain-based offerings	<p>[market, blockchain, company, investment, investor, technology, business, project, capital, fund]</p> <p><i>ICO's related to Blockchain-based offerings as effective capital-raising platforms.</i></p> <p>933, access capital consequently opportunity finance idea business</p> <p>3681, solution secondary market liquidity alternative finance platform</p> <p>1782, fund embrace opportunity challenge offer blockchain revolution</p>
Topic 27	Blockchain health	<p>[health, data, medical, patient, healthcare, service, information, care, user, record]</p> <p><i>ICO's related to secure transfer of patient medical records, manages the medicine supply chain and helps healthcare researchers unlock genetic code.</i></p> <p>4023, solution healthcare provider send appointment detail electronically patient</p> <p>2476, patient identity verify provider instantly access coverage information</p> <p>216, medical office network share patient record hospital</p>

Topic 28	Blockchain- based Telecom Infrastructure Marketplace	[service, telecom, business, company, industry, platform, country, global, customer, solution] <i>ICO's related to marketplace enabling stakeholders to transact and exchange value with trust based upon smart provenance and governance.</i> 2094, smart city offer solution single platform 456, opportunity market currently sector economy telecom 1994, market mobile operator enable telecom cryptocurrency
Topic 29	Blockchain Tech- nology for the Film Industry	[game people, idea, project, film, author, time, plan, work, world] <i>ICO's related to innovative opportunities to address issues related to distribution complexity, management of creative rights and digital piracy.</i> 74 bridge help author idea readable reach target audience 2584 world global animation studio generation independent investor 3064, profit sale time high sell common license
Topic 30	performance un- certainty	[statement, result, risk, actual, uncertainty, cause, future, performance, materially, event] <i>ICOs related to performance uncertainty and risk disclosure</i> 2117, statement involve unknown risk uncertainty factor cause actual future result performance achievement 1871, statement base current estimate projection involve risk uncertainty cause actual result materially 1846, risk uncertainty cause actual future result performance achievement render payment materially expect
Topic 31	user data encryp- tion	[data user, key, information, private, security, blockchain, access, wallet, identity] <i>ICOs related to user data encryption</i> 433, identity create generate public private key 187, database public user client identify key 2037, blockchain technology encryption ensure document security

Topic 32	ICO development	<p>[development, project, platform, fund, token, team, marketing, product, roadmap, plan]</p> <p><i>ICO development related to Token Development, ICO Marketing, Community Building</i></p> <p>819, chapter stage development roadmap</p> <p>2524 project website application ecosystem</p> <p>3566 market campaign target platform</p>
Topic 33	Smart contract technology	<p>[contract smart, transaction, service, blockchain, payment, platform, party, system, token]</p> <p><i>ICOs related to the smart contract technology</i></p> <p>1502, legal requirement blockchain smart contract nascent technology</p> <p>3613, power smart contract blockchain form decentralize application</p> <p>510, platform create decentralize application base smart contract</p>
Topic 34	blockchain dispute resolution	<p>[vote, contract, dispute, event, party, user, transaction, token, time, voting]</p> <p><i>ICO's related to blockchain dispute resolution provide a quick, cost-effective and efficient method of resolving disputes.</i></p> <p>1836, transaction promise resolve return receipt object value</p> <p>122, agreement record blockchain fund contract approve shareholder</p> <p>432, propose vote issue dispute resolution system</p>
Topic 35	Legal Disclosure	<p>[whitepaper, offer, security, token, investment, constitute, paper, jurisdiction, information, white]</p> <p><i>ICO's related to legal disclosure</i></p> <p>2054, whitepaper constitute prospectus document sort intend security solicitation investment jurisdiction</p> <p>1042, whitepaper constitute offer invitation person subscribe purchase share security company</p> <p>2624, whitepaper offer document prospectus intend provide basis investment decision contract</p>

Topic 36	Team experience	[experience, team, development, blockchain, business, project, marketing, company, technology, management] <i>ICO's related to team experience</i> 1592, development project security expert 1688, blockchain expert lead developer 1551, digital market professional experience
Topic 37	future performance statement	[statement, token, risk, future, business, information, whitepaper, change, operation, paper] <i>ICOs related to futur performance disclosure</i> 5072, information contain whitepaper promise commitment guarantee future performance outing distributor management policy 367, matter document future performance include limitation revenue look earnings strategy prospect statement 2931, despite intend reach describe document person party involve purchase token own risk
Topic 38	Blockchain security	[token, data, car, blockchain, art, service, owner, vehicle, platform, asset] <i>ICOs related to blochckain security</i> 1791, token help confirm identity track population transparent blockchain 819, gallery buyer guarantee project artistic safety authenticity art 2279, proof token owner access obtain service provide
Topic 39	Intellectual Property	[token, foundation, copyright, network, global, member, platform, reserved, governance, company] <i>ICO's related to Intellectual Property Interface</i> 616, regulation foundation management company 2601, protection project intellectual property 3901, software copyright protection function

Topic 40	Crypto Gaming	[game, player, platform, gaming, user, token, sport, betting, online, bet] <i>ICO's related to Crypto Gaming as a means to make money.</i> 2657, value bet mean high price token 3217, game play stake native token cryptocurrency 398, token entry ticket participate lounge game
Topic 41	Artificial intelligence and machine learning	[data, learning, network, model, system, machine, intelligence, analysis, market, user] <i>ICOs related to platforms using artificial intelligence</i> 152 research development lot time various design algorithm 1704 report analytic company insight market artificial intelligence 333 advance level platform integrate artificial intelligence method
Topic 42	Node Description	[data, node, network, file, storage, user, message, contract, blockchain, key] <i>ICO's related to node as critical component of a blockchain's infrastructure.</i> 709, datum server storage block receive store retrieve 1597, encrypt message transfer randomly select node blockchain 731, ledger availability node write datum read time
Topic 43	connect people	[platform, job, service, user, system, student, education, work, time, company] <i>ICOs related to connect people</i> 2899 own select application tool personal professional development 1615 search available talent term easy job freelance 2899 create criterion system assessment certification skill demand

Topic 44	Blockain in the Energy Industry	[mining, energy, token, market, power, project, electricity, coin, company, platform] <i>ICO's related to Blockchain technology has the potential to transform the energy sector.</i> 2883, mine software miner solve save time easy 900, facilitate community fund energy asset regional pool 2102, gold mine company infrastructure blockchain system
Topic 45	Token exchange	[token, holder, platform, exchange, value, share, company, investment, coin, market] <i>ICO's related to possibility of token exchange.</i> 3010, token value provide service crypto 885 investor able exchange token acquire 142, transferable grant token holder platform
Topic 46	law protection disclosure	[token, law, security, jurisdiction, regulation, united, person, sale, act, tax] <i>ICOs related to law protection disclosure</i> 2184 legal token sale regulatory opinion information qualification security dividend interest 3026 company issue token purchase resident country legislation forbid transaction type 866 purchase sale token subject protection law govern type financial instrument
Topic 47	product & service production	[production, plant, material, product, energy, project, construction, cost, technology, market] <i>ICOs related to product and services production</i> 208 investment project review entail construction installation cycle million 1787 solar system cell form panel convert sunlight electricity 745 fresh ingredient local organic farm food preparation process

Topic 48	energy consumption	<p>[energy, system, technology, power, carbon, device, data, water, waste, cost]</p> <p><i>ICOs related to energy consumption concerns</i></p> <p>1787, renewable energy byproduct capable offer service increase load efficiency</p> <p>1176, provide rational management resource allow minimize level industrial emission</p> <p>2668, mine energy intensive process power mean heat cool expense</p>
Topic 49	ICOs regulation	<p>[legal, security, regulation, financial, system, data, compliance, risk, law, process]</p> <p><i>ICOs related to ICOs regulation such as customer protection, fraudulent ICOs, etc.</i></p> <p>976, compliance measure implement plan worry security issue sec</p> <p>1317, requirement anti money launder control apply activity cryptocurrency</p> <p>2514, market regulation aim eradicate scam fraud criminal activity</p>
Topic 50	token transaction terms and conditions	<p>[token, fee, user, amount, transaction, wallet, account, contract, address, exchange]</p> <p><i>ICO's related to terms and condition for token transaction.</i></p> <p>1456, zero transaction fee trade profit commission</p> <p>1597, transaction require commission payment fee execution</p> <p>3770, list fee transaction network free charge</p>

Table 11.: Summary of statistical models

Variables	TokenTraded	Historical	$\log(1+\text{AmountRaised})$	ssGarchVol	AbnormalIR
Blockchain transaction(1)		■	■		
Digital wallet(2)			■		
Crypto Token(3)				■	
Payment system(4)					
Market description(5)	■				
Risk prevention(6)					
Real estate(7)					
Block. revo.(8)					
Platf. dev.(9)	■		■		
Mobile app(10)					
Token sale(11)					
Loan agreement(12)	■		■		
Content creation(13)			■		
Token(14)					
Trade platform(15)	■				
Terms and cond.(16)	■				
Smart contract desc.(17)	■		■		
Product/service deliv.(18)					
Team charact.(19)					
Transaction fees(20)					
Community platf.(21)					
User review(22)					
Risk underst.(23)					
Blockchain solution(24)					
Tone at the top(25)					
Blockchain based offer.(26)			■		
Blockchain health(27)					■
Blockchain telecom(28)	■				
Blockchain film(29)					■
Performance uncertainty(30)					
Data encrypt.(31)					
ICO develop.(32)			■		■
Smart contr. tech.(33)					
Blockchain dispute. (34)				■	
Legal disclosure(35)	■				
Team experience(36)	■				
Future perform.(37)					
Blockchain secu.(38)					
Intellectual prop.(39)			■		
Crypto gaming(40)			■		
A.I. and M.L. (41)					
Node descr.(42)			■		
Connect people(43)					
Blockchain energy(44)			■		
Token exchange(45)	■				
Law protection(46)					
Prod.& serv. (47)					
Energy consumpt.(48)					
ICO's reg.(49)					
Token transaction(50)	■				

■ Pos.& Sign. ■ Neg.&Sign.

Table 12.: Statistical models.

	glm(TokenTradedHistorical)	lm(log(1+AmountRaised))	lm(ssGarchVol)	lm(AbnormalIR)
(Intercept)	-1.736 *** (0.518)	14.151 *** (0.368)	1.413 *** (0.522)	0.166 (0.122)
Topic1	0.007 ** (0.003)	0.003 ** (0.002)	-0.002 (0.002)	0.000 (0.001)
Topic2	0.004 (0.004)	0.015 *** (0.003)	-0.008 (0.011)	-0.001 (0.003)
Topic3	0.003 (0.004)	0.001 (0.002)	-0.025 * (0.015)	0.003 (0.003)
Topic4	0.001 (0.003)	0.000 (0.002)	0.006 (0.007)	-0.001 (0.003)
Topic5	0.014 ** (0.006)	-0.006 (0.006)	0.006 (0.018)	0.029 (0.026)
Topic6	0.003 (0.006)	0.000 (0.004)	-0.010 (0.021)	0.003 (0.005)
Topic7	-0.001 (0.003)	-0.000 (0.002)	-0.006 (0.013)	0.004 (0.004)
Topic8	-0.001 (0.005)	-0.004 (0.005)	-0.010 (0.014)	-0.002 (0.003)
Topic9	0.008 *** (0.002)	0.006 ** (0.003)	-0.001 (0.002)	-0.001 (0.001)

Continuation of Table 10

	glm(TokenTradedHistorical)	lm(log(1+AmountRaised))	lm(ssGarchVol)	lm(AbnormalIR)
Topic10	-0.000 (0.004)	0.001 (0.004)	0.016 (0.013)	-0.004 (0.005)
Topic11	-0.006 (0.007)	0.006 (0.006)	0.068 (0.066)	0.005 (0.005)
Topic12	0.005** (0.002)	0.004** (0.002)	-0.002 (0.004)	-0.001 (0.001)
Topic13	0.002 (0.002)	0.003** (0.002)	-0.001 (0.002)	-0.002 (0.002)
Topic14	0.002 (0.003)	0.005 (0.004)	0.001 (0.006)	0.004 (0.004)
Topic15	0.003* (0.002)	0.003 (0.002)	-0.002 (0.005)	-0.000 (0.001)
Topic16	-0.018*** (0.007)	0.001 (0.007)	-0.039 (0.032)	-0.008 (0.005)
Topic17	0.012*** (0.004)	0.007*** (0.002)	0.004 (0.004)	-0.000 (0.001)
Topic18	0.003 (0.004)	0.001 (0.002)	-0.020 (0.016)	0.000 (0.001)
Topic19	-0.005 (0.005)	-0.007 (0.005)	-0.029 (0.032)	-0.003 (0.003)
Topic20	-0.002	-0.001	-0.018	0.004

Continuation of Table 10

	glm(TokenTradedHistorical)	lm(log(1+AmountRaised))	lm(ssGarchVol)	lm(AbnormalIR)
Topic21	(0.004) -0.000	(0.003) 0.002	(0.013) 0.005	(0.005) 0.000
Topic22	(0.004) 0.001	(0.002) -0.000	(0.011) -0.004	(0.002) -0.003
Topic23	(0.003) 0.007	(0.002) 0.001	(0.009) 0.022	(0.002) 0.002
Topic24	(0.008) 0.003	(0.007) -0.001	(0.022) 0.005	(0.007) 0.002
Topic25	(0.004) -0.002	(0.003) -0.001	(0.008) -0.006	(0.002) -0.007
Topic26	(0.005) 0.005	(0.002) 0.011**	(0.035) -0.005	(0.007) -0.002
Topic27	(0.005) -0.000	(0.005) 0.003	(0.008) 0.007	(0.004) 0.003*
Topic28	(0.002) -0.007**	(0.002) 0.000	(0.007) 0.040	(0.002) -0.007
Topic29	(0.004) -0.002	(0.004) -0.003	(0.036) -0.034	(0.006) 0.010**
Topic30	(0.008) -0.042	(0.003) 0.024	(0.021) 0.031	(0.005) -0.008
	(0.029)	(0.018)	(0.042)	(0.015)

Continuation of Table 10

	glm(TokenTradedHistorical)	lm(log(1+AmountRaised))	lm(ssGarchVol)	lm(AbnormalIR)
Topic31	-0.001 (0.003)	-0.002 (0.003)	-0.004 (0.004)	-0.001 (0.001)
Topic32	0.006 (0.005)	-0.014 *** (0.004)	-0.021 (0.019)	-0.007 * (0.004)
Topic33	-0.007 (0.005)	0.005 (0.003)	0.008 (0.011)	-0.003 (0.004)
Topic34	0.007 (0.005)	0.003 (0.002)	-0.011 * (0.006)	-0.001 (0.001)
Topic35	0.022 * (0.012)	0.005 (0.011)	-0.001 (0.019)	-0.003 (0.008)
Topic36	0.008 ** (0.004)	0.002 (0.004)	0.029 (0.030)	0.002 (0.003)
Topic37	0.003 (0.011)	0.006 (0.010)	0.019 (0.029)	0.005 (0.007)
Topic38	-0.001 (0.003)	0.002 (0.002)	-0.008 (0.005)	-0.001 (0.001)
Topic39	-0.004 (0.004)	0.008 *** (0.003)	-0.011 (0.010)	0.002 (0.003)
Topic40	0.001 (0.002)	0.002 * (0.001)	-0.005 (0.007)	-0.003 (0.002)
Topic41	-0.003	0.002	-0.004	-0.001

Continuation of Table 10

	glm(TokenTradedHistorical)	lm(log(1+AmountRaised))	lm(ssGarchVol)	lm(AbnormalIR)
Topic42	(0.003) 0.001	(0.001) 0.007***	(0.005) -0.002	(0.002) -0.001
Topic43	(0.002) -0.000	(0.002) 0.003	(0.002) 0.023	(0.001) 0.001
Topic44	(0.003) 0.003	(0.002) 0.004**	(0.024) -0.003	(0.002) -0.001
Topic45	(0.002) -0.012*	(0.002) -0.004	(0.002) -0.004	(0.001) 0.000
Topic46	(0.007) 0.001	(0.004) 0.000	(0.013) -0.018	(0.004) -0.005
Topic47	(0.004) -0.009	(0.003) -0.000	(0.013) 0.014	(0.006) -0.002
Topic48	(0.007) 0.001	(0.001) 0.000	(0.014) -0.009	(0.003) 0.001
Topic49	(0.003) -0.008	(0.003) -0.005	(0.012) -0.005	(0.002) -0.003
Topic50	(0.006) -0.010*	(0.005) -0.005	(0.006) 0.011	(0.003) -0.003
PreICODummy	(0.006) -0.242	(0.006) 0.022	(0.009) 0.151	(0.003) -0.079
	(0.149)	(0.128)	(0.382)	(0.125)

Continuation of Table 10

	glm(TokenTradedHistorical)	lm(log(1+AmountRaised))	lm(ssGarchVol)	lm(AbnormalIR)
BonusDummy	-0.287 (0.190)	-0.327* (0.195)	0.657 (0.446)	0.234 (0.344)
EthereumBlockDummy	0.021 (0.194)	0.022 (0.212)	-0.169 (0.278)	0.039 (0.087)
WhiteListKYCDummy	-0.589*** (0.204)	-0.256** (0.128)	0.224 (0.278)	-0.092 (0.121)
number_of_team	0.063*** (0.010)	0.043*** (0.008)	-0.029 (0.036)	-0.008 (0.006)
TaxHaven	0.335*** (0.113)	0.455*** (0.133)	0.306 (0.673)	-0.037 (0.063)
hardcapDummy	0.235 (0.199)	-0.172 (0.209)	0.424 (0.668)	0.103 (0.110)
softcapDummy	-0.405*** (0.146)	-0.366*** (0.139)	-0.890 (0.827)	0.040 (0.079)
NumCurrencies	0.029 (0.044)	0.016 (0.037)	-0.059 (0.079)	0.011 (0.041)
BitcoinPriceAtStart	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
FiatAcceptingDummy	0.390 (0.506)	0.109 (0.412)	-0.577 (0.767)	-0.160 (0.215)
AIC	1864.142			

Continuation of Table 10

	glm(TokenTradedHistorical)	lm(log(1+AmountRaised))	lm(ssGarchVol)	lm(AbnormalIR)
BIC	2202.657			
Log Likelihood	-870.071			
Deviance	1740.142			
Num. obs.	1737	853	398	405
R ²		0.181	0.115	0.183
Adj. R ²		0.118	-0.045	0.038
RMSE		1.730	3.802	0.896

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

