A topic modelling analysis of white papers in security token offerings: which topic matters for funds?

Paola Bongini, Alessia Pedrazzoli, Monica Rossolini[1]

Abstract

In the blockchain technology era, Security token offerings (STOs) are attracting increasing attention as an alternative means for ventures financing. This paper analyses the white papers' content of a unique sample of 193 STOs from 2017 to 2021 to detect which topic influences campaign success. We adopt latent Dirichlet allocation (LDA) topic modelling to identify the topics and themes in white papers. Nine topics emerge through LDA: company description, blockchain technology components, energy and green issue, financial and legal issues, AI and IT application in different industries (healthcare, manufacturing and construction, education, gaming, and financial services). We find that even if environmental disclosure is the theme with the highest coverage in the document, the disclosure about blockchain technology components is the only one that positively affects the probability of campaign success. Its effectiveness increases with the use of visual cues and a long space dedicated in the document. Results are helpful to entrepreneurs to better perform in the disclosure of their campaigns and open new questions for policymakers about investors in token offerings.

Keywords: security token; token offerings; white paper; LDA; topic modelling

[1] Paola Bongini, Full Professor, University of Milano-Bicocca,Via Degli Arcimboldi, 8 20126, Milan, Italy – paola.bongini@unimib.it
Alessia Pedrazzoli, Lecturer, University of Modena and Reggio Emilia, Via Università 4, 41121 Modena, italy - alessia.pedrazzoli@unimore.it
Monica Rossolini, Associate Professor, University of Milano- Bicocca, Via Degli Arcimboldi, 8 20126, Milan, Italy – monica.rossolini@unimib.it.

1. Introduction

Blockchain technology has been considered one of the youngest and fastest innovations in different industrial sectors and among the most valuable to future adoption (Bellavitis et al., 2021; Momtaz, 2019, 2020). Although the initial attention and successful application of blockchain were mainly driven by digital currency, its revolutionary impact is not limited to the payment services field. Crypto-assets have the potential to become an efficient and inclusive means of raising capital for small and medium-sized enterprises (Block et al., 2018; Fish et al., 2020; Simonella and Kondova, 2020), and the European Parliament (2020) recognized its importance in integrating the Capital Market Union.

Blockchain-based initial coin offerings (ICOs) and security token offerings (STOs) are increasingly considered alternative means to mainstream debt and equity fundraisings. Both ICOs and STOs are based on a direct peer-to-peer mechanism that enables an entrepreneurial project to raise funds in exchange for cryptographically secured tokens that can be publicly traded (Adhami et al., 2018; Lyandres et al., 2019). In ICOs, investors mainly acquire utility tokens linked to the right to use the company's product or service. In contrast, in STOs, investors purchase security tokens that share future earnings, voting rights, or ownership rights. Since STOs are selling security tokens, they are full-fledged financing instrument subject to regulation under securities law. The higher investor protection in STOs compared to ICOs has increased the expectation for the continuous growth of this market (Mazzorana-Kremer, 2019; Lambert et al., 2021), with a sharping 195% growth of the total volume raised from 2017($22 million) to 2018 ($442 million) (Pwc, 2020).

Literature that sheds light on ICOs is quickly expanding and has produced significant insights into the functioning of the market (Bellavitis et al., 2021; Huang et al., 2020), and venture funding determinants (Adhami et al., 2018; Ante et al., 2018; Fish, 2019; Fish and Momtaz, 2020; Howell et al., 2020; Momtaz, 2019; Zhao et al., 2020). In contrast, research focusing on STOs and their peculiarities are still scarce (Ante and Fiedler, 2019; Myalo, 2019; Lambert et al., 2021). As in ICOs, STOs issuers adopt voluntary disclosure to reduce information asymmetries. The white paper is where the token issuers promote their business idea with technical and descriptive information, with no specific requirements to follow a mandatory format. The content of the white paper is similar to the IPO prospectus or the presentation of an equity crowdfunding campaign, showing token evaluations, business idea, and risks. The white paper represents an informative means to attract prospective investors' interest, communicate the issuers' quality, and support the liquidity of the market (Cohney et al., 2018; Mazzorana-Kremer, 2019).

This paper further advances the literature on STOs by exploring the relevance of the various topics and contents in white papers that impact campaign performance. The nascent literature has focused

mainly on explaining if the white paper presence influences campaign dynamics or on mainly quantifying the content of a white in terms of its informativeness and readability (e.g., Blaseg, 2018; Samieifar and Baur, 2020; Zhang et al., 2019). We argue that also firms may provide disclosures on different topics, and these topics may have implications on investor decisions. The focus on blockchain technology but also environmental disclosure may assume high relevance during the company evaluation process. None of the previous studies examined the influence of topical features (i.e., latent semantics) mined from textual descriptions of projects on fundraising success. Therefore, as Rui Chen and Chen (2020) highlighted, there is the need to explore how investor behaviors are influenced by different information contents published during the campaign. Our research work can fill the research gap mentioned above, answering two research questions: *Which topics characterized STOs' white paper so far? which of them was able to impact campaign success?*

We use a natural language processing, Latent Dirichlet Allocation (LDA), to investigate which topic constitutes the white paper using a comprehensive sample of 193 STO issuers from 2017 to 2021. The full text of each campaign is retrieved from the six different STO portals. A topic model allows creating variables that aid in building a prediction model for determining whether a token offering would be successful (Chuanjie et al., 2019). Thus, we then examine with multivariate analysis which topics influence the STOs' success.

Nine different topics emerge in white papers of our sample: i. energy and green issues, ii. blockchain technology components, iii. company description, iv. financial and legal issues, v. AI and machine learning vi. IT in art and education, vii. IT in construction and energy industry, viii. IT in health care, and ix. IT in gaming, voting, and financial services. Our results reveal that the quantity of information dedicated to energy use and green aspects is the highest in the white paper. Still, only disclosure about blockchain technology components increases campaign success. Therefore, visual cues and an extended section dedicated to blockchain in the document increase the likelihood of successfully closing the campaign.

The study contributes to the literature and has practical implications in several ways. Firstly, it contributes to the literature about STOs' determinants of success (Ante and Fiedler, 2019; Lambert et al., 2021). Our findings suggest that not only the presence of technical words or the white paper itself impacts campaign funding but also the quantity of information dedicated to specific topics and the presence of visual elements that support the document readability. Secondly, we contribute to the literature on the role of disclosure in digital markets (Hornuf et al., 2021; Florysiak and Schandlbauer, 2019; Zhang et al., 2019). We apply LDA and show which themes should be effectively disclosed to increase campaign success. Finally, as more companies are starting to consider STO as a financial instrument, our study suggests that a more extensive discussion about blockchain technology and

solutions would benefit both entrepreneurs and STOs portals. Finally, the study contributes to the actual debate about token offering regulation. In particular, it gives essential insights into how investors are influenced by the content published in the campaign.

The remainder of the paper is organized as follows. Section 2 summarizes studies that explore white paper content in the literature and develop our hypotheses. Sections 3 and 4 present the research methodology and empirical findings, respectively. Section 5 concludes the paper with discussion, limitations, and future research directions.

2. Literature review and hypotheses development

2.1 Status of white paper analysis

Given their voluntary nature, white papers in STOs campaigns are perceived to reduce information asymmetry with third parties. Compared to more regulated IPOs and equity crowdfunding campaigns, the information asymmetry between fundraisers and investors is higher in STOs because of the high project variety and complicated fundamental technical details (Chan, 2019). Information asymmetry in security token offerings increases the cost of raising capital; thus, it is in the firms' interest to voluntarily disclose information to the public and signal project quality (Blaseg, 2018; Momtaz, 2020). Contrary to traditional IPO, STOs adopt much more flexible information requirements, and there is no standardized or mandatory format for white papers. As in crowdfunding platforms, STOs portals publish information regarding issuers, and then investors evaluate the quality of the offering and decide whether to invest or not. The whitepaper is considered a very persuasive marketing tool for investors in token offerings, even if its disclosure quality is generally poor (Zhao et al., 2021). Indeed, sometimes white papers did not disclose the names of their employees or listed someone who did not exist (Shifflett and Jones; 2018).

Few studies exist on STOs, whereas plenty of literature has been developed on ICOs, particularly on the role of white paper in token offerings. Previous literature in ICOs explored if the white papers presence impacts the funding success. Although few studies show that the white paper availability is not significant for campaign success (Adhami et al, 2018; Boreiko and Vidusso, 2019), the majority of studies pinpoints its positive impact on the amount raised (Boreiko and Sahdev, 2018; Cerchiello et al., 2019; Chen, 2019; Fish, 2019; Giudici et al., 2018). The importance of the disclosed contents in the white paper is higher for STOs compared to ICOs (Zhao et al., 2020), and in the absence of specific requirements, they vary dramatically in their structure, language style, and quality (Florysiak and Schandlbauer, 2019). Literature in token offerings has explored each of these aspects.

Focusing on white paper structure, the first group of studies investigates white paper readability. It appears that highly readable documents are generally valued because they make it easy for small investors to process information (Lee et al., 2017; Miller, 2010), and facilitate more accurate investment decisions (Arora and Chakraborty, 2021). Previous studies identify different readability measures, such as Fog index, which measures readability based on sentence and word length (Zhang et al., 2019), or the length of the entire document (Samieifar and Baur, 2020). In ICOs more readable white papers are likely to result in a higher initial return for ICO investors (Zhang et al., 2019), and publishing a longer white paper improves the likelihood of completing the ICO and the amount raised during the campaign (Amsden and Schweizer, 2019). Moreover, the average length of a white paper increased over time, thus suggesting that issuers understand the importance of white papers and attempt to signal their quality by publishing longer white papers (Samieifar and Baur, 2020). Florysiak and Schandlbauer, (2018) and Yen et al. (2021) investigate the white paper's content information structure. Florysiak and Schandlbauer, (2018) distinguish between the standard and informative white papers. They find that informative content is unrelated to the trading volume generated after the campaign, while Yen et al. (2021) evidence that ICOs with unique content raise more funds or are subject to more active trading and higher market values post-ICO period.

Others studies explore how the language style used in white papers influences the funding of new ventures. Zhang (2019) investigates the relationship between the level of mispricing at ICOs and the subjectivity (confidence and emotional tone). Findings reveal that if the language in a token's whitepaper shows more confidence, then the token tends to be less underpriced at its ICO. Momtaz (2020) quantifies informational exaggeration in ICO whitepapers as a manifestation of moral hazard. Exaggerated projects attract substantially more funding in significantly less time, but after a token starts trading, investors abandon excessive projects recognizing the biased signal.

In studies that explore white paper quality, this measure is often proxied by using a dichotomous categorisation for the presence of specific information and section or quantifying the white paper's information content. For example, whitepapers with a high number of "technical" words (for instance "block", "node", and "ledger") are considered of high quality (Bourveau et al., 2018; Lyandres et al., 2019). Chen (2019), Feng et al. (2019) and Fish (2019) provide evidence that offerings technical details in the white paper can be an effective way to signal the quality of an ICO project with a positive effect on the amount of funds raised. Another measurement of white paper quality is expressed by the presence of a detailed project evaluation that is also positively related to the duration of ICO (Blaseg, 2018; Boreiko and Sahdev, 2018) and its secondary market liquidity and trading volume (Bourveau, et al., 2018; Howell et al., 2020). Finally, other measurements of white

paper quality are whether a risk section (Konstantinidis et al., 2018) or the presence of a roadmap (Blaseg, 2018) are included in the text.

2.2 Which topics influence campaign performance?

Whitepaper is the primary source of voluntary disclosure in token offerings, and it has been used to proxy for the level of asymmetric information in several studies (e.g., Fisch, 2019; Howell et al., 2020; Lyandres et al., 2019). Due to information asymmetry, investors want to choose high quality business after a complex decision-making process. White papers aim at providing investors with the minimum amount of information that can facilitate an effective investment decision. Signalling theory has been used to explain voluntary disclosure in corporate reporting. Voluntary disclosure is one of the signalling means, where companies would disclose more information than the mandatory ones with the purpose to signal their quality (Campbell et al., 200; Verrecchia, 1990). Thus, the voluntary disclosure on specific topics explored in the white paper may impact the venture evaluation.

In particular, the disclosure of blockchain technology positively affects different corporate performance indicators (Chen, 2019; Yen and Wang, 2021). For instance, literature has evidenced a correlation between a company's blockchain disclosure and short-term market reaction (Akyildirim et al., 2020; Chaill et al., 2020; Cheng et al., 2019), and also with the long-term firm evaluation (Yen and Wang, 2021).

In STOs, understanding the technological component of the project and its application in the business proposed are crucial factors for STO ventures and investors (Fish, 2019). The presence of a patent, the presence of technical description of the project, or GIThub code often capture the technological capabilities (Bourveau et al., 2018; Chen, 2019; Lambert et al., 2021; Roosenboom et la.,2020). Moreover, investors in the token market require some technical expertise or the willingness to familiarize themselves with a digital wallet or the technical background of each venture (Fish, 2019; Moro and Wang, 2019; Lyandersa et al., 2019). For ICOs investors' technological motives are the most important motives that guide the investment decision, followed by financial and ideological motives (Fish et al., 2021). In the paper of Fish et al. (2021), the group of "technological motives" comprises items such as "personal enthusiasm for the technology of the ICO venture" and "personal enthusiasm for the business model/idea". In particular, the likelihood of STO success may be affected by the prevalence of blockchain concepts in the project description, especially those that improve investors' understanding of the project's underlying quality and token characteristics. In the light of these assumptions, we derive that a strong focus in STO' white paper on blockchain technology impacts campaign success.

H1: A more significant white paper's disclosure on blockchain technology positively influences the probability of campaign success.

Blockchain technology has offered considerable opportunities for advances in delivering environmental and social benefits (Adams et al., 2018; Bai and Sarkis, 2019), facilitating new means of green production, as well as monitoring and storing data-related activities responsible for pollution and environmental degradation (Hou et al., 2020; Saberi et al., 2018). At the same time, some energy-intensive design of many algorithms, processing, and computations within the blockchain pose a severe challenge in the consumption of electricity and the cut of greenhouse gas emissions (Truby, 2018). Thus, the energy consumption of blockchain technology differs significantly between different design choices, and subsequently, it is an important dimension to evaluate and to communicate to the market during the origination of a blockchain-based IT solution (Kannengießer et al. 2019; Sedlmeir et al., 2020).

The majority of large businesses volunteer information concerning the impact of their activities on the environment and how these impacts are managed within the firm (Hackston and Milne, 1996; Roberts, 1992). More specifically, environmental disclosure has commonly been viewed as a pre-emptive step to mitigate adverse regulatory or legislative pressures in the future (Brammer and Pavelin, 2008).

Since among ICOs investment motives the group of "ideological motives" comprises "social motives", sustainability or environmental impact of the company could be investigated as a valued topic disclosure in the white paper. In particular, for ICO investors the reading of a white paper carefully positively correlates with both ideological and technological motives (Fish et al., 2021). The ideological motives are not new in online financing and also in equity crowdfunding studies have evidenced the benefits of signalling project environmental or social orientation in the campaign presentation (Butticè et al., 2019; Calic and Mosakowski, 2016; Vismara, 2019).
These previous studies lead us to formulate the following hypothesis:

H2: A more significant white paper's disclosure on energy and green issues positively influences the probability of campaign success.

3. Sample and method

3.1 Sample

Following previous research (Adhami et al., 2018; Lyandres, 2019), we hand-collected information about token offerings across multiple well-known STOs aggregators: Coinitelligence.com, Tokenmarket.net, STOscope.com, STOrating.com, STOwise.com and ICObench.com. We manually matched data across various sources to generate a sample of uniquely identified offerings. Thanks to this strategy, we collected a sample of 430 STOs issued from December 2017 to February 2021. Following Lambert et al. (2021) we consider only "true" STOs, thus eliminating ICOs registered as STOs and stable coins. Then related white papers were downloaded from aggregators or from the official company's website, when present. Since some aggregators delate some campaign information affecting the availability of the campaigns, we used Google to search for white papers not available, and selected only white papers written in the English language. Our final sample consists of 193 STOs. Table 1 shows some descriptive characteristics of our sample.

The average duration of each offering is 102 days and only 35% of the issues provide for a pre-sale offering. In terms of geographical area, 56% of the issues refer to extra European countries, 34% to European countries and United Kingdom, and 10% of the sample is made up of Swiss companies. The majority of companies that decide to issue a STO belong to Finance Insurance and real estate (38% of the sample) and Services (36%) sectors. Manufacturing firms represent 10% of the sample, firms operating in media and communication sector represent the 9%. A smaller percentage is represented by mining and construction firms (3%) and retail and wholesale trade (4%). The larger part of the STOs issues are concentrated in the years 2018 and 2019 (90% of our sample).

Table 1
Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Offering days	193	102.25	82.01	6	548
Presale(dummy)	192	0.35	0.48	0	1
ExtraEU	193	0.56	0.50	0	1
Eu_UK	193	0.34	0.47	0	1
Swiss	193	0.10	0.31	0	1
Agricolture, foresty and fishing	193	0.00		0	1
Finance, Insurance, real estate	193	0.38	0.49	0	1
Manufacturing	193	0.10	0.30	0	1
Mining and Construction	193	0.03	0.16	0	1
Retail and wholesale trade	193	0.04	0.20	0	1
Services	193	0.36	0.48	0	1
Media and Communication	193	0.09	0.28	0	1

Year 2017	193	0.02		0	1
Year 2018	193	0.34	0.47	0	1
Year 2019	193	0.56	0.49	0	1
Year 2020	193	0.07	0.26	0	1
Year 2021	193	0.01	0.08	0	1

Method

The paper develops the analysis applying two main methodologies. The first step is addressed to answer the first research question, "*Which topics characterized STOs' white paper so far?".* We apply a LDA methodology described in Section 3.2. The second step is to verify which are the characteristics of the white paper able to affect the campaign success; in this case, we apply a Probit regression model described in Section 3.3.

3.2 LDA methodology

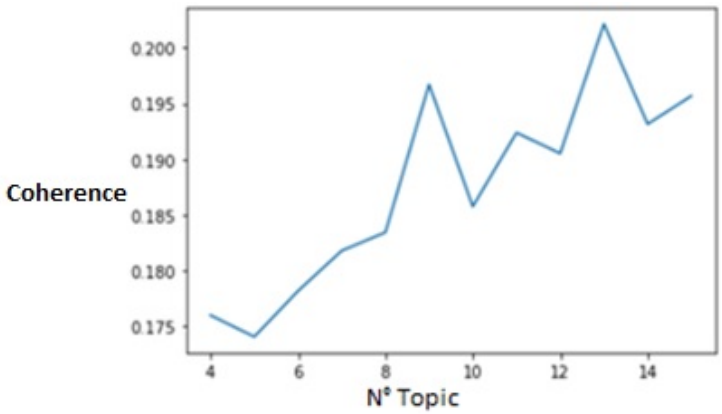
To verify which topics characterized the STOs' white papers we apply LDA. LDA is a Bayesian topic modelling that fully automates the process of producing a probability distribution of the topics in a corpus (Hannigan et al., 2019). Similar to factor or cluster analysis, LDA uses statistical associations of words to generate latent thematic structure within a huge amount of text (Blei et al., 2010; Chen, 2011). The latent topics are clusters of bag-of-words that jointly represent a theme in the text that emerge without the aid of pre-defined, explicit dictionaries or interpretive rules (Hannigan et al., 2019). Indeed, one of the advantages of LDA is that it does not impose on the researcher dictionaries or interpretative restrictions. It relies on the fact that words frequently appearing together tend to be semantically related (Stroropoli, 2019). This process reduces researcher bias, as foreknowledge of document content does not affect the topic classifications (Zhang et al., 2021). The LDA topic model is widely used in patent content analysis (Wang et al., 2015; Zhang et al., 2021) and technology topics evaluation (Wang et al., 2020, Yen and Wang, 2021). The application of LDA is a very recent technique in the categorization of ICO whitepapers. It has been used only in two recent papers. Chuanjie et al. (2019) applied data science techniques to categorize ICO whitepapers topics, Bian et al. (2018) lightly touched on using LDA to categorize ICOs into ten different topics based on manual labelling process. The novelty of our work is that we explore the relationship between LDA emerging topics and STO funding success.

To develop LDA method we extract the text from the white paper. First, the texts are pre-processed such that only words remain and these are lemmatized. We eliminate stopwords and filter terms that are very common such as "blockchain," "token," "paper," and "white," that appear in almost all documents. Moreover, we delete the company's name and other words with a higher

frequency (60%) present in the text and of little significance that could influence the analysis. Hence, these words would generate minimal value in differentiating one category of whitepapers from another. Following Chuanjie et al. (2019) other words such as “platform,” “user,” “data,” “service,” “system,” “network,” “contract,” “exchange,” “business,” and “company” have been included in the list of stop words in the pre-processing of the text documents to achieve more disparity in the topic words.

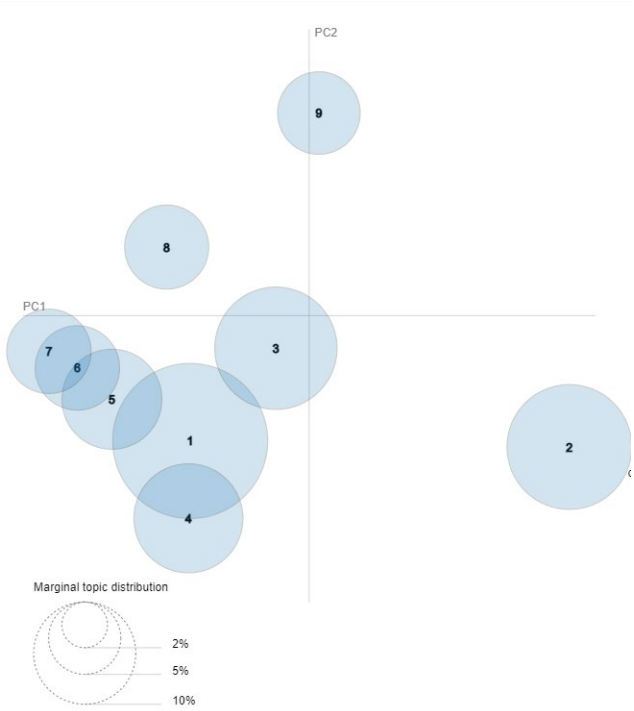
To achieve the optimal number of topics two measures have been computed: cosine distance and topic coherence. The cosine distance has been discussed by Blei and Lafferty (2007) and applied to intend to maximize similarity within clusters and minimize similarity between clusters. Following Chuang et al. (2012), we use saliency as the second criterion for evaluating the quality of the model. In Figure 1, it is possible to see that the coherence increases as the number of topics increase until the number of nine topics; after that point, a marginal increase in the number of topics decreases the coherence. Subsequently, the coherence rises again, reaching a new maximum value considering 13 topics. We have to consider a trade-off between coherence and interpretability of the topic (Zhang et al., 2021). Therefore, the number of topics cannot be so large that the model becomes unusable for practical usage. We choose the model with nine topics to facilitate topics interpretability.

Fig. 1. Number of topics



We use LDAMulticore package to visualize the results of LDA topic models (Chuang et al., 2012; Sievert and Shirley, 2014) in Figure 2. This method represents each different topic in a circle, where numbers on the circles represent the serial numbers of the topics, and the topics are arranged according to the quantity of related text data. The distance between the circles indicates the degree of discrimination be-tween different topics. The Figure 2 denotes that the topics have minimal overlap that is due to the repeat of keywords between these topics.

Fig. 2. STO White paper Topic Model



LDA results

The LDA modelling outputted nine sets of terms based on the input texts, where each set of words refers to a topic. The content of each topic is represented in Table 2. For interpretation, we then assigned each topic a label according to the top terms in each topic. To provide the most suitable labels, we followed the suggestion of Sievert and Shirley (2014), we ranked the terms in each topic by their relevance, which is a weighted average of a term's probability within a topic and that to its marginal probability among the whole corpus. Then we follow previous studies that explore blockchain technology (Chuanjie et al. 2019; Zhang et al., 2021) to support our label interpretation. According to the terms, five of the nine topics are related to the company and blockchain technology adopted: energy and green issue, blockchain technology components, company description, AI and machine learning, and financial and legal issues. While others four topics introduce the application of IT components in specific businesses: art and education, construction and energy industry, health care system and gaming, voting and financial services.

Next, we calculated the posterior probabilities of the nine topics based on the LDA algorithm for each input paragraph. In this case, each whitepaper has a distribution among nine topics where the probabilities sum up to one.

Table 2
Extracted topics from STOs white paper

Topic	Top 7 words	Meaning
1	Energy, green, mining, electricity, production, capacity, wind	Energy and green issue related to the business
2	Start-up, corporation, production, series, venture, CEO, exit	Company description
3	Compute, film, equipment, machine, learn, ai, device	AI and machine learning
4	Algor, art, learn, ai, chain, education, artwork	IT for art and education
5	Chain, quantum, node, block, Algor, hash, protocol	Blockchain technology components
6	Plant, gas, material, construction, oil, turbine, engineer	IT for construction and manufacturing
7	Patient, healthcare, medical, data, care, compute, app	IT for healthcare
8	Issuer, bond, tax, portfolio, prospectus, purchaser, income	Financial and legal issues
9	Game, player, vote, app, banking, credit, easy, web	IT for gaming, voting, and financial services

3.3 Econometric models

To detect the impact of white paper's topics focusing on the blockchain technology components (*H1*) and environmental disclosure (*H2*) on the likelihood of a STO success, we perform a probit model. The likelihood of STO success is a dummy defined as whether the total fundraised in the STO is greater than the minimum target (Moro and Wang, 2019). In this model, we estimate the probability that firm *i* achieves a successful issue using the following equation:

$$Prob(Success_i = 1) = F\big(\alpha + \beta_1(Readability) + \beta_2(Structure)_i + \beta_3(Topics)_i + \beta_4(Industry) + \beta_5(Geo_{area}) + \beta_6(Year) + \varepsilon_{i,k,t}\big) \quad \textbf{(1)}$$

where *Success* represents the closing of a successful campaign for firm *i* and α is a constant term. In the above equation, "Readability" as in previous literature (Samieifar and Baur, 2020; Zhang et al., 2019) is a vector of variables that include: Fog rate[2], the number of pages in the document (*Length*), the number of files attached to the white paper (*File*), and the size of the document in KB *(Size*). "Structure" is a vector of variables that include dummy variables that capture the presence of pictures (*Pictures*), tables (*Tables*), charts (*Charts*), table of content (*Table of content*), and the presence of other documents published apart to the whitepaper (*Extra documents*).

[2] Calculated as ((average number of words per sentence) + (number of words of 3 syllables or more)) ∗ 0.4)

"Topic" is a vector of variables that represents the nine topics identified through the LDA. Accordingly, we create nine variables: *Topic_1* to *Topic_9*, the percentage of words in the document related to a specific topic. We include all nine indicator variables simultaneously in our analysis. As in Fish (2019), we also control for sectorial, geographical, and time-fixed effects, where the locational status is where the token is issued.

To control for other potential aspects that could influence the campaign success we considered: the variable *Offering days,* which captures the campaign duration (Roosenboom et al., 2020), and a dummy variable equals one if the offering provides for a pre-sale (*Pre-sale offering*) and zero otherwise. We follow the evidence by Adhami et al. (2018) and, Lyandres et al. (2019), who showed that campaigns with a pre-sale are more successful. Table A.1 in the Appendix shows the variable definitions.

Table 3 shows the main characteristics of our sample. On average, the white papers have a length of 38 pages, a higher number compared to ICOs white papers (33 pages, Samieifar and Baur , 2020), and a size of 5,371 KB. The readability measured by Fog rate is about 15; a high score indicates less readable text. The data is higher compared to ICO white papers (13.70, Samieifar and Baur, 2020). The number of attached documents to the whitepaper is 1.35, and only 22% of issuers present additional documents. In terms of white papers' structure, it includes on average six pictures, three charts, and about two tables, indicating the use of visual elements to support the text. The 76% of white papers present at the beginning of the document a table of contents.

One of the most interesting results of our descriptive analysis relates to the content distribution in the white papers. As described above, we detected nine different Topics. The Topic 1 - *Energy and green issue* related to the business represents on average the 27% of the total text. This means that environmental disclosure is one of the main topics that are covered in the text. Words related to AI/machine learning, company presentation, and blockchain technology components represent 16%, 11%, 9%, respectively. The other topics related to the application of IT in the specific industry have a low percentage (between 5% and 9%), as well as the financial and legal issues (8%).

Table 3
All sample

	Obs	Mean	Std. Dev.	Min	Max
Lenght (number of pages)	193	38.10	24.98	2	208
Size (KB)	193	5371.80	9250.23	75	100558
Fog rate	193	15.28	2.71	8.33	33.43
Files (number)	193	1.35	0.95	1	7
Extra documents (dummy)	193	0.22	0.42	0	1

Pictures (number)	193	6.37	5.65	0	26
Charts (number)	193	3.02	4.02	0	22
Tables (number)	193	1.92	2.59	0	20
Table of contents (dummy)	193	0.76	0.43	0	1
Topic1 in %- Energy and green issue	193	0.27	0.31	0	1
Topic2 in %- Company description	193	0.11	0.24	0	1
Topic3 in %-AI and machine learning	193	0.16	0.29	0	1
Topic4 in %- IT for art and education	193	0.09	0.22	0	1
Topic5 in %- Blockchain technology components	193	0.09	0.25	0	1
Topic6 in %- IT for construction and manufacturing	193	0.05	0.17	0	1
Topic7 in %- IT for Healthcare	193	0.08	0.23	0	1
Topic8 in %- Financial and legal issue	193	0.08	0.22	0	1
Topic9 in %- IT for gaming, voting and financial services	193	0.07	0.20	0	1

4. Empirical results

The results of this first step are helpful to develop our second step aimed at verifying which white paper's themes are able to influence the STO's success and in particular, the role of blockchain technology disclosure (*H1*), and environmental disclosure (*H2*). Table 4 shows the main differences between successful and unsuccessful STOs. The difference in means underlies that successful STO has a higher number of files and additional documents and is more frequently characterized by a pre-sale offering. This means that a high disclosure represents a positive signal for token investors. In terms of theme, successful STO seem to dedicate more text in the white paper to explaining blockchain technology components.

Table 4
Successful and Unsuccessful STOs

	Successful STO					Unsuccessful STO					Equality of means
Variable	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max	
Length (number of pages)	156	37.37	24.66	2	208	37	41.19	26.42	6	122	-3.82
Size (KB)	156	4899.97	6187.96	75	39155	37	7361.14	16925.16	158	100558	-2461.17
Fog rate	156	15.31	2.85	8.33	33.43	37	15.18	2.04	9.40	19.82	0.13
Files (number)	156	1.40	1.03	1	7	37	1.11	0.39	1	3	0.30*
Extra documents (dummy)	156	0.25	0.43	0	1	37	0.11	0.31	0	1	0.14*
Pictures (number)	156	6.58	5.75	0	26	37	5.49	5.20	0	26	1.09
Charts (number)	156	2.93	3.67	0	20	37	3.38	5.29	0	22	-0.45
Tables (number)	156	2.02	2.76	0	20	37	1.49	1.64	0	6	0.53
Table of contents (dummy)	156	0.75	0.43	0	1	37	0.78	0.42	0	1	-0.03
Offering days(number)	156	100.16	77.96	6	366	37	111.08	97.98	17	548	-10.92
Pre-sale(dummy)	155	0.39	0.49	0	1	37	0.22	0.42	0	1	0.17*
Topic1in %- Energy and green issue	156	0.27	0.32	0	1	37	0.28	0.32	0	1	-0.01
Topic2 in %- Company description	156	0.11	0.23	0	1	37	0.14	0.28	0	1	-0.04
Topic3 in %- AI and machine learning	156	0.16	0.29	0	1	37	0.15	0.29	0	1	0.02
Topic4 in %- IT for art and education	156	0.09	0.22	0	1	37	0.09	0.25	0	0.94	0.00
Topic5 in %- Blockchain technology components	156	0.10	0.26	0	1	37	0.03	0.14	0	0.82	0.07*
Topic6 in %- IT for construction and manufacturing	156	0.05	0.16	0	1	37	0.05	0.21	0	1	0.00
Topic7 in %- IT for Healthcare	156	0.08	0.22	0	1	37	0.09	0.27	0	1	-0.01
Topic8 in %- Financial and legal issue	156	0.08	0.22	0	1.00	37	0.08	0.24	0	0.94	-0.01
Topic9 in %- IT for gaming, voting and financial services	156	0.06	0.19	0	1	37	0.08	0.21	0	1.00	-0.01

The Table represents the descriptive statistics of a sample of 156 STOs from 2017-2021. T- test indicate the statistical differences between successful and unsuccessful campaigns, ***, **, * at the 1%, 5% and 10% levels, respectively

Table 5 and 6 shows the results of the probit regression models. Model I shows that an increase in the number of pages decreases the probability of success, as well as, the size of the white paper. An increase of one page in the white paper decreases by 0.2% the probability of success. In contrast, an increase of one hundred KB of the withe paper's size decreases by 0.004% the likelihood of success. Compared to ICOs' white paper, where the length of the document is positively related with campaign funds raised (Samieifar and Baur, 2021) STOs present longer and less readable documents. This means that to publish a long white paper decreases campaign success probability. Conversely about the structure, a high number of non-text elements and in particular, pictures and tables, increases the probability of success by 1.29% and 2.46% respectively. The results are confirmed also controlling for industry fixed effect (Model II).

Model III adds to the analysis the themes of the white paper. It confirms the previous results relative to the structure and readability, add an important information relative to the Topics: the unique content able to influence the success seem to be the information relative to blockchain technology. This result confirms our *H1*, while we do not find evidence about the impact of environmental disclosure (*H2)*. An increase by 1% of the words of this topic increases by 40.60% the probability of success.

In Models IV-VI we add some interaction for verifying if other structural elements of the white paper are able to emphasize the effect of Topic 5 on the STO's success. Model IV focuses on the role of the pictures. The use of images seems to be very efficacy in communicate the blockchain technology, an increase in the number of pictures generates a relevant increase (equal to 77.61%) of the Topic 5's effect on STO's success. Conversely, the communication of the Topic 5 without the use of Pictures seem to be not effective, it even reduces the probability of success (-50%). In particular Model V focuses on length. An increase in the number of pages dedicated to the blockchain increases the probability of success by 3.99%. Finally, Model VI considers the number of Tables. Generally, an increase in the number of Table increases the probability of success but Tables dedicated to blockchain are not efficacy, in contrast to the relevant role of pictures described in model IV.

Table 5
Main Analysis part A

VARIABLES	Model I	Margins	Model II	Margins	Model III	Margins
Lenght	-0.019**	-0.002**	-0.024***	-0.003***	-0.012**	-0.002**
	(0.009)	(0.001)	(0.009)	(0.001)	(0.005)	(0.001)
Size (KB)	-0.000*	-0.000*	-0.000*	-0.000*	-0.000*	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Extra documents (dummy)	0.867	0.120	1.050*	0.133*	0.543	0.126*
	(0.607)	(0.083)	(0.622)	(0.077)	(0.332)	(0.076)
Pictures (number)	0.092*	0.012*	0.108**	0.013**	0.057**	0.013**
	(0.049)	(0.006)	(0.049)	(0.006)	(0.025)	(0.005)
Charts (number)	-0.071	-0.009	-0.054	-0.006	-0.045	-0.010
	(0.059)	(0.008)	(0.058)	(0.007)	(0.030)	(0.007)
Tables (number)	0.177*	0.024*	0.197*	0.025*	0.114*	0.026*
	(0.106)	(0.014)	(0.107)	(0.013)	(0.059)	(0.013)
Table of contents (dummy)	-0.060	-0.008	-0.191	-0.024	-0.043	-0.010
	(0.477)	(0.066)	(0.516)	(0.065)	(0.275)	(0.063)
T1-Energy and green issue (%)					0.506	0.117
					(0.528)	(0.122)
T3-AI and machine learning(%)					0.629	0.146
					(0.549)	(0.126)
T4-IT for art and education(%)					0.694	0.161
					(0.651)	(0.150)
T5-Blockchain technology components(%)					1.750**	0.406**
					(0.735)	(0.168)
T6-IT for construction and manufacturing(%)					0.310	0.071
					(0.738)	(0.171)
T7-IT for Healthcare(%)					0.523	0.121
					(0.659)	(0.152)
T8-Financial and legal issue(%)					0.603	0.140
					(0.636)	(0.148)
T9-IT for gaming, voting and financial services(%)					0.451	0.105
					(0.747)	(0.173)
Constant	18.36***		31.64***		6.956***	
	(1.093)		(1.627)		(0.591)	
Time FE	Yes		Yes		Yes	
Country FE	Yes		Yes		Yes	
Sector FE	No		Yes		No	
Observations	192		192		192	
Pseudo R2	0.112		0.179		0.149	
ROC AUC	0.7221		0.7824		0.7576	
Prob > chi2	0.4887				0.4669	

VIF	2.9	6.06	3.21

Table V shows the results of the probit regression model with with '***', '**' and '*' representing the 1%, 5% and 10% levels, respectively. The dependent variable is a dummy variable, which equals 1 when the STO closes with success. Time, regional and sector fixed effects are three vectors that capture differences at chronological, geographical, and sectorial levels. The 'Coeff' columns report the coefficient of the probit regression; the 'Margins' columns report the marginal effects of the probit regression.

Table 6
Main Analysis part B

VARIABLES	Model IV	Margins	Model V	Margins	Model VI	Margins
Lenght	-0.013**	-0.002**	-0.014**	-0.003**	-0.013**	-0.003**
	(0.005)	(0.001)	(0.005)	(0.001)	(0.005)	(0.001)
Size (KB)	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Extra documents (dummy)	0.600*	0.133*	0.562	0.127*	0.547	0.129
	(0.343)	(0.074)	(0.343)	(0.076)	(0.336)	(0.082)
Pictures (number)	0.052**	0.011**	0.059**	0.013**	0.059**	0.013**
	(0.024)	(0.005)	(0.025)	(0.005)	(0.026)	(0.006)
Charts (number)	-0.044	-0.009	-0.047	-0.010	-0.047	-0.010
	(0.030)	(0.006)	(0.030)	(0.006)	(0.030)	(0.007)
Tables (number)	0.117**	0.025**	0.117**	0.026**	0.110*	0.025*
	(0.058)	(0.012)	(0.059)	(0.013)	(0.059)	(0.013)
Table of contents (dummy)	-0.102	-0.022	-0.127	-0.028	-0.071	-0.012
	(0.288)	(0.063)	(0.289)	(0.065)	(0.276)	(0.065)
T1-Energy and green issue (%)	0.487	0.108	0.505	0.114	0.507	0.115
	(0.519)	(0.114)	(0.519)	(0.116)	(0.524)	(0.121)
T3-AI and machine learning(%)	0.652	0.144	0.666	0.150	0.635	0.144
	(0.545)	(0.120)	(0.545)	(0.122)	(0.547)	(0.127)
T4-IT for art and education(%)	0.682	0.151	0.713	0.161	0.707	0.162
	(0.651)	(0.143)	(0.653)	(0.146)	(0.651)	(0.154)
T5-Blockchain technology components(%)	-2.258*	-0.500**	-1.237	-0.279	1.088	0.246
	(1.153)	(0.251)	(1.594)	(0.358)	(0.834)	(0.204)
T6-IT for construction and manufacturing(%)	0.401	0.0887	0.367	0.0827	0.326	0.0673
	(0.734)	(0.162)	(0.734)	(0.166)	(0.735)	(0.172)
T7-IT for Healthcare(%)	0.494	0.109	0.509	0.115	0.511	0.125
	(0.651)	(0.143)	(0.652)	(0.146)	(0.654)	(0.156)
T8-Financial and legal issue(%)	0.491	0.109	0.561	0.126	0.592	0.123
	(0.627)	(0.139)	(0.629)	(0.142)	(0.632)	(0.147)
T9-IT for gaming, voting and financial services(%)	0.434	0.0960	0.466	0.105	0.432	0.106
	(0.734)	(0.162)	(0.743)	(0.167)	(0.744)	(0.178)
T5-Blockchain technology components x Pictures	3.507*	0.776*				
	(1.960)	(0.427)				
T5-Blockchain technology components x Lenght			0.177*	0.0399*		
			(0.103)	(0.0231)		

T5-Blockchain technology components x Tables			2.033	0.509
			(2.387)	(0.582)
Constant	6.793***	6.917***	6.988***	
	(0.612)	(0.621)	(0.597)	
Time FE	Yes	Yes	Yes	
Country FE	Yes	Yes	Yes	
Sector FE	No	No	No	
Observations	192	192	192	
Pseudo R2	0.1891	0.174	0.158	
ROC AUC	0.777	0.7705	0.7595	
Prob > chi2	0.746	0.6804	0.5822	
VIF	3.33	3.35	3.22	

Table VI shows the results of the probit regression model with with '***', '**' and '*' representing the 1%, 5% and 10% levels, respectively. The dependent variable is a dummy variable, which equals 1 when the STO closes with success. Time, regional and sector fixed effects are three vectors that capture differences at chronological, geographical, and sectorial levels. The 'Coeff' columns report the coefficient of the probit regression; the 'Margins' columns report the marginal effects of the probit regression.

Robustness

To verify the stability of our results we carry out some robustness checks (Table 7). In Model VII we substitute the variable *Extra documents* with the variable *Number of Files* that appears not significant, however the results achieved in the main analysis are confirmed. Model VIII controls the robustness of the main results adding to the model two other variables: pre-sale offering and offering days. Neither is significant, however the results achieved in the main analysis are confirmed. Finally, in Model IX we measure the relevance of each topic using the logarithm of the new variable is calculated as ln (1+ previous variable) as suggested by Yen and Wang, (2021). All the previous results are confirmed and especially also the relevance of content relative to blockchain technology.

Table 7
Robustness test

VARIABLES	Model VII	Margins	Model VIII	Margins	Model IX	Margins
Lenght	-0.019*	-0.0025*	-0.018*	-0.002*	-0.021**	-0.002**
	(0.010)	(0.001)	(0.010)	(0.001)	(0.010)	(0.001)
Size (KB)	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Extra documents (dummy)			0.978	0.129	0.994	0.132
			(0.673)	(0.086)	(0.628)	(0.083)
Pictures (number)	0.099**	0.013**	0.086*	0.011*	0.101**	0.013**
	(0.044)	(0.006)	(0.045)	(0.005)	(0.046)	(0.006)
Charts (number)	-0.083	-0.011	-0.066	-0.008	-0.076	-0.010
	(0.052)	(0.007)	(0.053)	(0.007)	(0.055)	(0.007)
Tables (number)	0.206**	0.027**	0.165	0.021	0.205*	0.027**
	(0.104)	(0.013)	(0.107)	(0.013)	(0.107)	(0.013)
Table of contents (dummy)	-0.104	-0.014	-0.268	-0.035	-0.052	-0.006
	(0.502)	(0.072)	(0.507)	(0.066)	(0.495)	(0.066)
T1-Energy and green issue _ln	0.629	0.084	0.702	0.092	1.219	0.162
	(0.870)	(0.122)	(0.963)	(0.126)	(1.280)	(0.169)
T3_ AI and machine learning _ln	0.855	0.115	0.779	0.102	1.509	0.201
	(0.944)	(0.126)	(0.993)	(0.130)	(1.292)	(0.170)
T4_ IT for art and education_ln	0.810	0.109	1.066	0.140	1.633	0.217
	(1.114)	(0.143)	(1.147)	(0.150)	(1.527)	(0.202)
T5_ Blockchain technology components_ln	2.943*	0.395**	2.793*	0.368*	4.308**	0.574**
	(1.555)	(0.199)	(1.515)	(0.197)	(2.069)	(0.272)
T6_ IT for construction and manufacturing_ln	0.181	0.024	-0.191	-0.025	0.915	0.168
	(1.290)	(0.175)	(1.337)	(0.176)	(1.798)	(0.194)
T7_ IT for Healthcare_ln	0.707	0.095	0.795	0.105	1.469	0.122
	(1.156)	(0.140)	(1.247)	(0.162)	(1.638)	(0.239)

T8_ Financial and legal issue_ln	0.583	0.078	1.045	0.138	1.262	0.196
	(1.187)	(0.173)	(1.130)	(0.149)	(1.448)	(0.215)
T9_ IT for gaming, voting and financial services_ln	0.531	0.071	0.487	0.064	1.082	0.144
	(1.238)	(0.163)	(1.397)	(0.184)	(1.692)	(0.225)
Files	0.523	0.070				
	(0.374)	(0.057)				
Presale			0.787	0.104		
			(0.521)	(0.064)		
Offering days			-0.001	-0.000		
			(0.002)	(0.000)		
Constant	17.11***		18.93***		17.60***	
	(1.356)		(1.329)		(1.305)	
Time FE	Yes		Yes		Yes	
Country FE	Yes		Yes		Yes	
Sector FE	No		No		No	
Observations	192		191		192	
Pseudo R2	0.143		0.16		0.148	
ROC AUC	0.7548		0.7761		0.7602	
Prob > chi2	0.5383		0.1389		0.4358	
VIF	3.35		3.23		3.2	

Table VII shows the results of the probit regression model with with '***', '**' and '*' representing the 1%, 5% and 10% levels, respectively. The dependent variable is a dummy variable, which equals 1 when the STO closes with success. Time, regional and sector fixed effects are three vectors that capture differences at chronological, geographical, and sectorial levels. The 'Coeff' columns report the coefficient of the probit regression; the 'Margins' columns report the marginal effects of the probit regression.

5. Conclusion and limitations

The aims of this work are twofold: to explore which themes characterize STO's white paper and to verify if blockchain technology or environmental disclosure impacts the probability of campaign success. Our results evidence that environmental disclosure relative to energy use and green issues is the most covered in the text. Still, it does not impact the probability of campaign success (*H2*). Otherwise, the theme of blockchain technology components, even if it has a lower coverage in the document, assumes a primary role in the campaign funding (*H1*).

These results support previous literature about determinants of token offering success, and in particular studies that provide evidences on specific technological signals such as patent, GIThub code, and the number of technical words in the document on the funds collected in the campaign (e.g., Florysiak and Schandlbauer, 2018; Howell et al., 2020; Samieifar and Baur, 2020). Blockchain technology components disclosed in the white paper are a hot topic for investors. They may be perceived as a quality signal by the market that reduces information asymmetry between parties. As

in Bhattacharya and Ritter (1983), disclosing firms' technological progress is relevant for their financing choices, assuming it provides a credible signal about their innovation process. This opens up the possibility that in token offering the disclosure about blockchain technology serves multiple purposes besides business technological characteristics. In this vein, we demonstrate the importance of technological disclosure through the white paper, building on the idea that this topic is a credible signal to infer the quality of otherwise hard-to-value company innovation. For investors, blockchain topic may also represent the characteristics of the tokenized digital securities and how blockchain is employed to keep and validate token transactions.

Therefore, our findings strengthen the predictions about the leading role of technological motives for investors in token offerings (Fish et al., 2021) confirmed by the type of information they look at. We contribute to the literature on which information attracts investors in token offering and contributes to mitigating informational asymmetries. Opposite to crowdfunding, where radical innovation does not find support among small investors (Chang and Parhankangas, 2017), in STOs, technology is a convincing topic that increases campaign success.

About environmental disclosure, preliminary work of Mansouri and Momtaz (2021) finds evidence about ESG disclosure in utility token offering and its positive relationship with the funding amount. Our work focuses on STOs detecting only the environmental component of the project and in particular, the energy issue. Thus, the "E" of ESG. We surmise that ICOs and STOs meet fundamentally different disclosure requirements, such as different investors' expectations. STOs are alternative investments instruments related to an expected return in the future (Cohney et al., 2018). Thus, investors mainly focus their attention on token characteristics and to the company's economic sustainability in the future. A fruitful avenue for future research would be to propose a more granular approach to ESG and sustainability evaluations in token offerings taking in to account a distinction between ICOs and STOs.

Our study also contributes to the nascent literature about disclosure in digital context. Even if innovative entrepreneurs may face a trade-off between disclosing technological components and secrecy (Moser, 2011; Glaeser, 2018), our work evidence that in STOs this disclosure strongly influences campaign success. In decentralized finance, investors face a high level of information asymmetry and they have to support the investment decision using this public information. Therefore, sceptical entrepreneurs may gain other extra-financial benefits if they combine the disclosure of blockchain technology information and interactive space on the campaign page. For example, a forum through which investors can give valuable feedback and ideas to the project or posing questions about token characteristics would increase the company's market awareness and project visibility. Thus, open the company to external ideas and open innovation concepts.

Moreover, the study has implications for innovative entrepreneurs, investors, and policymakers. For entrepreneurs, our results suggest adopting structural aspects in the document to increase its readability and the effectiveness of the funding round. Pictures and a more extended section in the white paper dedicated to blockchain technology increases campaign success. Visual cues are rapidly processed and integrated during the investor evaluation (Khaneman, 2003), and in general, investors tend to process easy-interpretable signals when a lot of information is available (Chan, 2019). As in crowdfunding (Scheaf et al., 2018), STO entrepreneurs may use many visual elements in the campaign presentation to create a rich imaginary for potential investors and support the communication of blockchain technology. Therefore, even if a lengthy white paper harms campaign success, when more information about blockchain technology becomes available in the white paper, it will positively impact. For small investors, the complexity of the technology creates difficulties in assessing the business's growth potential and increases their due diligence costs (Mason and Harrison, 2003). The use of visual elements may support the document's readability and consequently increase small investors' participation. Our findings are also relevant for the policymaker. The supervisory authorities discuss the possibility of categorising the different tokens as financial instruments and consequently making them subject to MIFID2 and its disciplines. In light of the paper's results, we argue that, in such circumstances, one of the relevant aspects to be considered is the revision of standard prospectus' content, giving adequate space to the technological part. If the prospectus topics becomes mandatory in STOs, the supervisory authority should provide a table of content for the technological information and not only for the financial one as well as a supervision of the content published.

The current study also has limitations. First, our results could be affected by the relatively short sample period. Future research could extend the observation period and consequently the sample dimension. Secondly, regulators are developing specific disclosure requirements for STOs. Thus, future research could explore how different requirements could impact company disclosure and investors' participation in different countries.

References

Adams, R., Kewell, B., Parry, G., 2018. Blockchain for Good? Digital Ledger Technology and Sustainable Development Goals, in: World Sustainability Series. pp. 127–140. https://doi.org/10.1007/978-3-319-67122-2_7

Adhami, S., Giudici, G., Martinazzi, S., 2018. Why do businesses go crypto? An empirical analysis of initial coin offerings. J. Econ. Bus. 100, 64–75. https://doi.org/10.1016/j.jeconbus.2018.04.001

Akyildirim, E., Corbet, S., Sensoy, A., Yarovaya, L., 2020. The impact of Blockchain related name changes on corporate performance. J. Corp. Financ. 65. https://doi.org/10.1016/j.jcorpfin.2020.101759

Amsden, R., Schweizer, D., 2019. Are Blockchain Crowdsales the New "Gold Rush"? Success Determinants of Initial Coin Offerings. SSRN Electron. J. https://doi.org/10.2139/ssrn.3163856

Ante, L., Fiedler, I., 2019. Cheap Signals in Security Token Offerings. SSRN Electron. J. https://doi.org/10.2139/ssrn.3356303

Ante, L., Sandner, P., Fiedler, I., 2018. Blockchain-Based ICOs: Pure Hype or the Dawn of a New Era of Startup Financing? J. Risk Financ. Manag. 11, 80. https://doi.org/10.3390/jrfm11040080

Arora, J., Chakraborty, M., 2021. Does the ease of reading of financial disclosures influence investment decision? Econ. Lett. 204. https://doi.org/10.1016/j.econlet.2021.109883

Bai, C., Sarkis, J., 2020. A supply chain transparency and sustainability technology appraisal model for blockchain technology. Int. J. Prod. Res. 58, 2142–2162. https://doi.org/10.1080/00207543.2019.1708989

Bellavitis, C., Fisch, C., Wiklund, J., 2021. A comprehensive review of the global development of initial coin offerings (ICOs) and their regulation. J. Bus. Ventur. Insights 15. https://doi.org/10.1016/j.jbvi.2020.e00213

Bhattacharya, S., Ritter, J.R., 1983. Innovation and Communication: Signalling with Partial Disclosure. Rev. Econ. Stud. 50, 331. https://doi.org/10.2307/2297419

Bian, S., Deng, Z., Li, F., Monroe, W., Shi, P., Sun, Z., & Li, J. 2018. Icorating: A deep-learning system for scam ico identification. arXiv preprint arXiv:1803.03670.

Blaseg, D., 2018. Dynamics of Voluntary Disclosure in the Unregulated Market for Initial Coin Offerings. SSRN Electron. J. https://doi.org/10.2139/ssrn.3207641

Blei, D., Carin, L., Dunson, D., 2010. Probabilistic topic models. IEEE Signal Process. Mag. 27, 55–65. https://doi.org/10.1109/MSP.2010.938079

Blei, D.M., Lafferty, J.D., 2007. A correlated topic model of Science. Ann. Appl. Stat. 1. https://doi.org/10.1214/07-aoas114

Block, J.H., Colombo, M.G., Cumming, D.J., Vismara, S., 2018. New players in entrepreneurial finance and why they are there. Small Bus. Econ. 50, 239–250. https://doi.org/10.1007/s11187-016-9826-6

Boreiko, D., Sahdev, N.K., 2018. To ICO or not to ICO – Empirical analysis of Initial Coin Offerings and Token Sales. SSRN Electron. J. https://doi.org/10.2139/ssrn.3209180

Boreiko, D., Vidusso, G., 2019. New blockchain intermediaries: Do ico rating websites do their job well? J. Altern. Investments 21, 67–79. https://doi.org/10.3905/jai.2019.21.4.067

Bourveau, T., De George, E.T., Ellahie, A., Macciocchi, D., 2018. Initial Coin Offerings: Early Evidence on the Role of Disclosure in the Unregulated Crypto Market. SSRN Electron. J. https://doi.org/10.2139/ssrn.3193392

Brammer, S., Pavelin, S., 2008. Factors influencing the quality of corporate environmental disclosure. Bus. Strateg. Environ. 17, 120–136. https://doi.org/10.1002/bse.506

Butticè, V., Colombo, M.G., Fumagalli, E., Orsenigo, C., 2019. Green oriented crowdfunding campaigns: Their characteristics and diffusion in different institutional settings. Technol. Forecast. Soc. Change 141, 85–97. https://doi.org/10.1016/j.techfore.2018.07.047

Cahill, D., G. Baur, D., (Frank) Liu, Z., W. Yang, J., 2020. I am a blockchain too: How does the market respond to companies' interest in blockchain? J. Bank. Financ. 113. https://doi.org/10.1016/j.jbankfin.2020.105740

Calic, G., Mosakowski, E., 2016. Kicking Off Social Entrepreneurship: How A Sustainability Orientation Influences Crowdfunding Success. J. Manag. Stud. 53, 738–767. https://doi.org/10.1111/joms.12201

Campbell, D., Shrives, P., Bohmbach-Saager, H., 2001. Voluntary Disclosure of Mission Statements in Corporate Annual Reports: Signaling What and To Whom? Bus. Soc. Rev. 106, 65–87. https://doi.org/10.1111/0045-3609.00102

Cerchiello, P., Tasca, P., Toma, A.M., 2019. ICO success drivers: A textual and statistical analysis. J. Altern. Investments 21, 13–25. https://doi.org/10.3905/jai.2019.21.4.013

Chan, C.S.R., Parhankangas, A., 2017. Crowdfunding Innovative Ideas: How Incremental and Radical Innovativeness Influence Funding Outcomes. Entrep. Theory Pract. 41, 237–263. https://doi.org/10.1111/etap.12268

Chen, K., 2019. Information asymmetry in initial coin offerings (ICOs): Investigating the effects of multiple channel signals. Electron. Commer. Res. Appl. 36. https://doi.org/10.1016/j.elerap.2019.100858

Chuang, J., Manning, C.D., Heer, J., 2012. Termite: Visualization techniques for assessing textual topic models, in: Proceedings of the Workshop on Advanced Visual Interfaces AVI. https://doi.org/10.1145/2254556.2254572

Chuanjie, F., Koh, A., Griffin, P., 2019. Automated Theme Search in ICO Whitepapers. J. Financ. Data Sci. 1, 140–158. https://doi.org/10.3905/jfds.2019.1.011

Rui Chen, R., Chen, K., 2020. A 2020 perspective on "Information asymmetry in initial coin offerings (ICOs): Investigating the effects of multiple channel signals." Electron. Commer. Res. Appl. 40. https://doi.org/10.1016/j.elerap.2020.100936

Chen, E., 2011. Introduction to Latent Dirichlet Allocation. Blog.Echen.Me 1.

Cheng, S.F., De Franco, G., Jiang, H., Lin, P., 2019. Riding the blockchain mania: Public firms' speculative 8-k disclosures. Manage. Sci. 65, 5901–5913. https://doi.org/10.1287/mnsc.2019.3357

Chuanjie, F., Koh, A., Griffin, P., 2019. Automated Theme Search in ICO Whitepapers. J. Financ. Data Sci. 1, 140–158. https://doi.org/10.3905/jfds.2019.1.011

Cohney, S., Hoffman, D., Sklaroff, J., Wishnick, D., 2019. Coin-operated capitalism. Columbia Law Rev. 119, 591–676. https://doi.org/10.2139/ssrn.3215345

European Parliament (2020), COMMISSION DELEGATED REGULATION (EU) 2020/2034 October 6, accessible at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32020R2034&from=en

Fisch, C., 2019. Initial coin offerings (ICOs) to finance new ventures. J. Bus. Ventur. 34, 1–22. https://doi.org/10.1016/j.jbusvent.2018.09.007

Fisch, C., Masiak, C., Vismara, S., Block, J., 2021. Motives and profiles of ICO investors. J. Bus. Res. 125, 564–576. https://doi.org/10.1016/j.jbusres.2019.07.036

Fisch, C., Meoli, M., Vismara, S., 2020. Does blockchain technology democratize entrepreneurial finance? An empirical comparison of ICOs, venture capital, and REITs. Econ. Innov. New Technol. https://doi.org/10.1080/10438599.2020.1843991

Fisch, C., Momtaz, P.P., 2020. Institutional investors and post-ICO performance: an empirical analysis of investor returns in initial coin offerings (ICOs). J. Corp. Financ. 64. https://doi.org/10.1016/j.jcorpfin.2020.101679

Feng, C., Li, N., Wong, M.H.F., Zhang, M., 2019. Initial Coin Offerings, Blockchain Technology, and White Paper Disclosures. SSRN Electron. J. https://doi.org/10.2139/ssrn.3256289

Florysiak, D., Schandlbauer, A., 2018. The Information Content of ICO White Papers. SSRN Electron. J. https://doi.org/10.2139/ssrn.3265007

Gerber, E.M., Hui, J., 2013. Crowdfunding: Motivations and deterrents for participation. ACM Trans. Comput. Interact. 20, 32. https://doi.org/10.1145/2530540

Glaeser, S., 2018. The effects of proprietary information on corporate disclosure and transparency: Evidence from trade secrets. J. Account. Econ. 66, 163–193. https://doi.org/10.1016/j.jacceco.2018.04.002

Hackston, D., Milne, M.J., 1996. Some determinants of social and environmental disclosures in New Zealand companies. Accounting, Audit. Account. J. 9, 77–108. https://doi.org/10.1108/09513579610109987

Hannigan, T.R., Haan, R.F.J., Vakili, K., Tchalian, H., Glaser, V.L., Wang, M.S., Kaplan, S., Jennings, P.D., 2019. Topic modeling in management research: Rendering new theory from textual data. Acad. Manag. Ann. 13, 586–632. https://doi.org/10.5465/annals.2017.0099

Harrison, R.T., Mason, C.M., 2019. Venture Capital 20 years on: reflections on the evolution of a field. Ventur. Cap. https://doi.org/10.1080/13691066.2019.1562627

Hornuf, L., Kück, T., Schwienbacher, A., 2021. Initial coin offerings, information disclosure, and fraud. Small Bus. Econ. https://doi.org/10.1007/s11187-021-00471-y

Hou, J., Wang, C., Luo, S., 2020. How to improve the competiveness of distributed energy resources in China with blockchain technology. Technol. Forecast. Soc. Change 151. https://doi.org/10.1016/j.techfore.2019.119744

Howell, S.T., Niessner, M., Yermack, D., 2020. Initial coin offerings: Financing growth with cryptocurrency token sales. Rev. Financ. Stud. https://doi.org/10.1093/rfs/hhz131

Huang, W., Meoli, M., Vismara, S., 2020. The geography of initial coin offerings. Small Bus. Econ. 55, 77–102. https://doi.org/10.1007/s11187-019-00135-y

Kahneman, D., 2003. Maps of bounded rationality: Psychology for behavioral economics. Am. Econ. Rev. https://doi.org/10.1257/000282803322655392

Kannengießer, N., Lins, S., Dehling, T., Sunyaev, A., 2019. What does not fit can be made to fit! Trade-offs in distributed ledger technology designs, in: Proceedings of the Annual Hawaii International Conference on System Sciences. pp. 7069–7078. https://doi.org/10.24251/hicss.2019.848

Kher, R., Terjesen, S., Liu, C., 2020. Blockchain, Bitcoin, and ICOs: a review and research agenda. Small Bus. Econ. https://doi.org/10.1007/s11187-019-00286-y

Konstantinidis, I., Siaminos, G., Timplalexis, C., Zervas, P., Peristeras, V., Decker, S., 2018. Blockchain for business applications: A systematic literature review, in: Lecture Notes in Business Information Processing. pp. 384–399. https://doi.org/10.1007/978-3-319-93931-5_28

Lambert, T., Liebau, D., Roosenboom, P., 2021. Security Token Offerings Available at SSRN: https://ssrn.com/abstract=3634626 or http://dx.doi.org/10.2139/ssrn.3634626 (June 25, 2021)

Lee, T. (David), Yun, T.W., Haley, E., 2017. What You Think You Know: The Effects of Prior Financial Education and Readability on Financial Disclosure Processing. J. Behav. Financ. 18, 125–142. https://doi.org/10.1080/15427560.2017.1276064

Lyandres, E., Palazzo, B., Rabetti, D., 2019. Do Tokens Behave like Securities? An Anatomy of Initial Coin Offerings. An Anat. Initial Coin Offer. (April 20, 2019)

Lukkarinen, A., Wallenius, J., Seppälä, T., 2018. Investor Motivations and Decision Criteria in Equity Crowdfunding. SSRN Electron. J. https://doi.org/10.2139/ssrn.3263434ù

Mansouri, S., Momtaz, P.P., 2021. Financing Sustainable Entrepreneurship: ESG Measurement, Valuation, and Performance in Token Offerings. SSRN Electron. J. https://doi.org/10.2139/ssrn.3844259

Mason, C.M., Harrison, R.T., 2003. “Auditioning for Money.” J. Priv. Equity 6, 29–42. https://doi.org/10.3905/jpe.2003.320037

Mazzorana-Kremer, F., 2019. Blockchain-Based Equity and STOs: Towards a Liquid Market for SME Financing? Theor. Econ. Lett. 09, 1534–1552. https://doi.org/10.4236/tel.2019.95099

Momtaz, P.P., 2020. Entrepreneurial Finance and Moral Hazard: Evidence from Token Offerings. J. Bus. Ventur. https://doi.org/10.1016/j.jbusvent.2020.106001

Momtaz, P.P., 2019. Token sales and initial coin offerings: Introduction. J. Altern. Investments. https://doi.org/10.3905/jai.2019.21.4.007

Moro, A., Wang, D., 2019. FinTech Projects and Initial Coin Offerings: A Research Note. J. Entrep. Financ. 21.

Moser, P., 2011. Innovation without patents: Evidence from world’s fairs. J. Law Econ. 55, 43–74. https://doi.org/10.1086/663631

Myalo, A.S., 2019. Comparative analysis of ICO, DAOICO, IEO and STO. Case study. Financ. Theory Pract. 23, 6–25. https://doi.org/10.26794/2587-5671-2019-23-6-6-25

PWC, 2020. 6 th ICO / STO Report. Available at: https://www.pwc.com/ee/et/publications/pub/Strategy&_ICO_STO_Study_Version_Spring_2020.pdf

Roberts, R.W., 1992. Determinants of corporate social responsibility disclosure: An application of stakeholder theory. Accounting, Organ. Soc. 17, 595–612. https://doi.org/10.1016/0361-3682(92)90015-K

Roosenboom, P., van der Kolk, T., de Jong, A., 2020. What determines success in initial coin offerings? Ventur. Cap. 22, 161–183. https://doi.org/10.1080/13691066.2020.1741127

Rui Chen, R., Chen, K., 2020. A 2020 perspective on “Information asymmetry in initial coin offerings (ICOs): Investigating the effects of multiple channel signals.” Electron. Commer. Res. Appl. 40. https://doi.org/10.1016/j.elerap.2020.100936

Saberi, S., Kouhizadeh, M., Sarkis, J., 2018. Blockchain technology: A panacea or pariah for resources conservation and recycling? Resour. Conserv. Recycl. https://doi.org/10.1016/j.resconrec.2017.11.020

Samieifar, S., Baur, D.G., 2021. Read me if you can! An analysis of ICO white papers. Financ. Res. Lett. 38. https://doi.org/10.1016/j.frl.2020.101427

Scheaf, D.J., Davis, B.C., Webb, J.W., Coombs, J.E., Borns, J., Holloway, G., 2018. Signals’ flexibility and interaction with visual cues: Insights from crowdfunding. J. Bus. Ventur. 33, 720–741. https://doi

Sedlmeir, J., Buhl, H.U., Fridgen, G., Keller, R., 2020. The Energy Consumption of Blockchain Technology: Beyond Myth. Bus. Inf. Syst. Eng. 62, 599–608. https://doi.org/10.1007/s12599-020-00656-x

Sievert, C., Shirley, K., 2015. LDAvis: A method for visualizing and interpreting topics. pp. 63–70. https://doi.org/10.3115/v1/w14-3110

Simonella, G., Kondova, G., 2020. Blockchain in Startup Financing: ICOs and STOs in Switzerland. J. Strateg. Innov. Sustain. 14. https://doi.org/10.33423/jsis.v14i6.2607

Storopoli, J.E., 2019. Topic Modeling: How and Why to Use in Management Research. Rev. Ibero-Americana Estratégia 18, 316–338. https://doi.org/10.5585/ijsm.v18i3.14561

Truby, J., 2018. Decarbonizing Bitcoin: Law and policy choices for reducing the energy consumption of Blockchain technologies and digital currencies. Energy Res. Soc. Sci. https://doi.org/10.1016/j.erss.2018.06.009

Verrecchia, R.E., 1990. Information quality and discretionary disclosure. J. Account. Econ. 12, 365–380. https://doi.org/10.1016/0165-4101(90)90021-U

Vismara, S., 2019. Sustainability in equity crowdfunding. Technol. Forecast. Soc. Change 141, 98–106. https://doi.org/10.1016/j.techfore.2018.07.014

Wang, M.Y., Lo, H.C., Liao, Y.Y., 2015. Knowledge flow determinants of patent value: Evidence from Taiwan and South Korea biotechnology patents. Int. J. Innov. Technol. Manag. 12. https://doi.org/10.1142/S0219877015400040

Wang, Xuefeng, Yang, X., Wang, Xinglu, Xia, M., Wang, J., 2020. Evaluating the competitiveness of enterprise's technology based on LDA topic model. Technol. Anal. Strateg. Manag. 32, 208–222. https://doi.org/10.1080/09537325.2019.1648789

Yen, J.C., Wang, T., 2021. Stock price relevance of voluntary disclosures about blockchain technology and cryptocurrencies. Int. J. Account. Inf. Syst. 40. https://doi.org/10.1016/j.accinf.2021.100499

Yen, J.C., Wang, T., Chen, Y.H., 2021. Different is better: how unique initial coin offering language in white papers enhances success. Account. Financ. https://doi.org/10.1111/acfi.12760

Zhang, H., Daim, T., Zhang, Y. (Peggy), 2021. Integrating patent analysis into technology roadmapping: A latent dirichlet allocation based technology assessment and roadmapping in the field of Blockchain. Technol. Forecast. Soc. Change 167. https://doi.org/10.1016/j.techfore.2021.120729

Zhang, S., Aerts, W., Lu, L., Pan, H., 2019. Readability of token whitepaper and ICO first-day return. Econ. Lett. 180, 58–61. https://doi.org/10.1016/j.econlet.2019.04.010

Zhao, X., Hou, W., An, J., Liu, X., Zhang, Y., 2020. Initial Coin Offerings: What Rights Do Investors Have (If Any)? SSRN Electron. J. https://doi.org/10.2139/ssrn.3576553

Zhu, H., Zhou, Z.Z., 2016. Analysis and outlook of applications of blockchain technology to equity crowdfunding in China. Financ. Innov. https://doi.org/10.1186/s40854-016-0044-7

Appendix

Tab. A.1 Variable definitions

Variable	Description
Dependent variables	
Success	Dummy = 1 if the STO reaches the minimum target amount
Independent variables	
Pictures	Number of pictures in the white paper
Charts	Number of charts in the White paper
Tables	Number of Tables in the White paper
Table of content	Dummy = 1 if the white paper presents a table of contents
Extra documents	Dummy = 1 if there are other documents published apart from White paper (i.e product presentation, financial forecast)
Lenght	Number of pages
File	Number of files attached to the White paper
Size	White paper's weight in KB
Fog rate	Readability rate that considers the number of complex words in the text
T1	Percentage of words in the White paper related to energy and green issue
T2	Percentage of words in the White paper related to company description
T3	Percentage of words in the White paper related to AI and machine learning
T4	Percentage of words in the White paper related to IT for art and education
T5	Percentage of words in the White paper related to blockchain and technology components
T6	Percentage of words in the White paper related to IT for construction and manufacturing
T7	Percentage of words in the White paper related to IT for healthcare
T8	Percentage of words in the White paper related to financial and legal issue
T9	Percentage of words in the White paper related to IT for gaming, voting and financial services
Offering days	Number of days of the offering
Pre-sale offering	Dummy equals 1 of the offering provide for a pre-sale