A topic modelling analysis of white papers in security token offerings: Which topic matters for funding?

Journal Item

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

Paola Bongini[1], Francesco Osborne[1,2], Andrea Maurino[1], Monica Rossolini[1]

[1]*Department of Business and Law, University of Milano Bicocca, Milan, Italy*
[2]*Knowledge Media Institute, The Open University, Milton Keynes, United Kingdom*

Abstract

Security token offerings (STOs), based on blockchain technology, are attracting increasing attention as an innovative alternative means of venture financing. Information about specific STOs is generally provided in white papers. This study analyses the content of white papers using a unique sample of 188 STOs from 2017 to 2021 to identify which topic is related to campaign success. We leverage latent Dirichlet allocation (LDA) topic modelling to identify the topics and themes in white papers. Nine topics are identified through LDA: company description, distributed ledger technology components, energy and green issues, financial and legal issues, artificial intelligence, and tech applications in different industries—specifically, healthcare, manufacturing and construction, education, and financial services. We find that energy and green issues represent one of the most prominent topics among all types of projects and that their disclosure is positively related to the probability of campaign success and the amount of funding raised. Another prominent topic that affects campaign outcomes is technology in the healthcare industry, reflecting wider investment trends in this sector. Our results may help entrepreneurs to improve their campaign disclosures and present new issues for policymakers regarding investments in digital tokens.

Keywords: blockchain; disclosure; Latent Dirichlet Allocation; security token offering; topic modelling; white paper

1. Introduction[1]

Distributed ledger technologies (DLT) and blockchain are two of the newest and fastest-growing innovations, relevant across many industries and offering great value for future adoption (Bellavitis et al., 2021; Momtaz, 2019, 2020). Although the initial attention to DLT and its earliest successful applications related mainly to digital currencies, its revolutionary impact is not limited to payment services. Digital assets, such as initial coin offerings (ICOs) and security token offerings (STOs), have the potential to become efficient and inclusive means of raising capital for small and medium-sized enterprises (Block et al., 2018; Fish et al., 2020; Momtaz, 2021; Simonella and Kondova, 2020), and the European Commission (2020) has recognized their importance in advancing the Capital Markets Union. Indeed, ICOs and STOs have increasingly considered alternatives to mainstream debt and equity fundraising. These instruments are based on a direct peer-to-peer mechanism that enables an entrepreneurial project to raise funds in exchange for cryptographically secured tokens, issued on a DLT system, 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 acquire the company's product or service. In contrast, in STOs, investors purchase security tokens that confer a share of future earnings, voting rights, or ownership rights in the issuing company. Security tokens are therefore full-fledged financial assets, potentially subject to regulation under securities law.

The higher investor protection for STOs, compared to ICOs, has increased expectations of continuous growth in this market (Lambert et al., 2021; Mazzorana-Kremer, 2019), which has achieved growth in the total amount raised from $22 million in 2017 to $442 million in 2018 (PwC, 2020). However, research on STOs is nascent and limited to a small number of papers (Ante and Fiedler, 2019; Beinke et al., 2021; Lambert et al., 2021; Miglo, 2021; Myalo, 2019). In contrast, because ICOs emerged before STOs, the literature on ICOs is more mature and has produced significant insights into the functioning of this market (Bellavitis et al., 2021; Huang et al., 2020) and the determinants of venture funding (Adhami et al., 2018; Ante et al., 2018; Fish, 2019; Fish and Momtaz, 2020; Howell et al., 2020; Meoli and Vismara, 2022; Momtaz, 2019; Xu et al., 2021; Zhao et al., 2020). Specifically, this literature has highlighted the importance of transparent and detailed disclosure of information about the issuing firm, its business, and its governance features in boosting the success of ICO campaigns (Samieifar and Baur, 2021, Zhang et al., 2019). White papers are the

[1] Abbreviations
DLT: distributed ledger technology
ICO: initial coin offerings
IPO: initial public offering
LDA: latent Dirichlet allocation
MiCA: Markets in Crypto-Assets regulation
MiFID: Markets in Financial Instruments Directive
STO: security token offerings

typical means of conveying such information and reducing information asymmetry with third parties—primarily investors.

White papers are the only formal technical documents that help potential purchasers in making informed decisions about their investments (Adhami et al., 2018). Like initial public offering (IPO) prospectuses, white papers are an informative means of attracting investors' interest and communicating the issuers' quality. Unlike IPO prospectuses, the content and form of white papers are not dictated by law or regulation and they do not have a common structure (Zhang et al., 2019). The lack of a mandatory, standardized format and specified disclosure requirements[2] leaves room for voluntary disclosure and greater transparency of the proposed project, which could influence the funds collected and the success of the campaign itself, as has been demonstrated in the ICO market (Chen, 2019; Fish, 2019; Meoli and Vismara, 2022). While STO issuers await detailed information disclosure rules (for example, the Markets in Crypto-Assets Regulation -MiCA-, currently at the proposal stage in the European Parliament), we argue that such firms should familiarize themselves with the types of content of white papers that have been proven beneficial to venture financing.

Accordingly, this paper advances the literature on STOs by exploring the topics of white papers that are related to campaign performance. Existing literature on STOs mainly focuses on whether the presence of a white paper influences campaign dynamics or on quantifying its content in terms of informativeness and readability (e.g., Ante and Fiedler, 2019; Blaseg, 2018; Samieifar and Baur, 2021; Zhang et al., 2019). We argue that firms can also choose the topics on which their disclosures provide information and that this choice has implications for investors' decisions. The fact that the company's prospectus content can influence investors' decision-making has been strongly supported by literature in the IPO domain, where specific contents such as those in Form-S1 affect investors' evaluation (Arnold et al. 2008; Beatty and Ritter, 1986; Hanley and Hoberg, 2010; Loughran and McDonald, 2013). No previous studies have examined the relationship with the fundraising success of topical features mined from textual descriptions of STO projects. Therefore, as highlighted by Rui

[2] An increasing number of countries are implementing regulatory frameworks specifically addressing crypto-assets and crypto-asset service providers. In Europe, STOs are not currently specifically regulated; however, following a European Commission consultation on crypto-assets which closed in March 2020, Markets in Crypto-Assets Regulation (MiCA) has been proposed as a framework that imposes an obligation on issuers of crypto-assets to publish a white paper with mandatory disclosure requirements. This regulation would apply only to STOs for which tokens are not covered by EU financial services legislation (unless the tokens also qualify as e-money tokens). As outlined below, many STOs are currently covered by EU financial services legislation applicable to Markets in Financial Instruments Directive (MiFID) financial instruments, and so appear to fall outside MiCA's scope. Nevertheless, given the technological component of STOs, it is still important for the regulator to leverage the provisions of MiCA to understand what specific information that is not traditionally included in prospectuses could be further considered. MiCA suggests that the white paper should include general information on the issuer, the project to be carried out, the offering of crypto-assets or their admission to trading, the rights and obligations attached to the crypto-assets, information on the underlying technology used, and the related risks. The information included in the white paper and the marketing communications related to the offering are required to be fair, clear, and not misleading. For details of US regulations, see Lambert et al. (2021) or Momtaz (2021), and for other countries see Blandin et al. (2019).

Chen and Chen (2020), it is necessary to explore how investors' behaviours are influenced by different types of information content published during STO campaigns. Referring to signalling theory and the use of voluntary disclosure to reduce information asymmetry between investors and entrepreneurs (Diamond and Verrecchia, 1991; Verrecchia, 1990), our study fills the research gap described above by answering the following questions: *Which topics have characterized STO white papers so far? Which of these topics is related to campaign success?*

We adopted a well-known generative probabilistic topic model, Latent Dirichlet Allocation (LDA), to investigate the topics addressed by a sample of 188 STO white papers from 2017 to 2021. The full text of each with paper was retrieved from several STO portals. We used LDA to identify the main topics and characterize each document according to its most significant topics. We then performed a multivariate analysis to test the relationships between white paper topics and STOs' success.

We identified nine topics from the white papers in our sample: i) energy and green issues; ii) company description; iii) artificial intelligence (AI) and machine learning; iv) tech for art and education; v) DLT component; vi) tech for the construction and energy industry; vii) tech for health care; viii) financial and legal issues, and ix) tech for finance and other services. Our results reveal that the share of information dedicated to energy use and green aspects is among the highest in our sample and disclosure of such information is positively related to campaign success. Another important topic for funding is the disclosure of information related to tech for the healthcare field, thus revealing a prominent focus of attention for this industry.

This study makes numerous contributions to the literature and has several practical implications. First, it advances the literature on STOs' determinants of success (Ante and Fiedler, 2019; Lambert et al., 2021) because we are the first, to our knowledge, to analyse the relationship between white paper topics and campaign performance. Our findings suggest that the presence of technical words is related to campaign funding, but so too does the share of information dedicated to specific topics. Second, this study contributes to the literature on the role of disclosure in digital markets (Hornuf et al., 2021; Florysiak and Schandlbauer, 2022; Zhang et al., 2019). We apply LDA and show which themes should be effectively disclosed to increase campaign success. Third, as more companies consider using STOs as a financial instrument, our study suggests that a more extensive discussion of energy use and the underlying technology of the business would benefit both entrepreneurs and STO portals. Finally, the study contributes to the current debate on token offering regulation. It gives essential insights to regulators into how investors are influenced by information published in the campaign, supporting one of the main objectives of the current regulation: instilling appropriate investor protection (European Parliament, 2020).

The remainder of the paper is organized as follows. Section 2 summarizes existing studies that explore white paper content and introduces our hypotheses. Sections 3 and 4 present the research methodology and empirical findings, respectively. Section 5 discusses our findings and concludes the paper.

2. Literature review and hypotheses development

2.1 Related work

As in crowdfunding platforms, token offering portals publish information about the issuers, and investors evaluate the quality of the offering and decide whether to invest. In this context, a white paper could be a persuasive and informative tool for investors in token offerings (Zhao et al., 2020). Although the content, length and form of white papers are not dictated by law or regulation and they do not have a common structure, they tend to comply with a minimum of information and typically include descriptions of the firm's industry, market status, relevant competitors, finances, risk factors of the project, token evaluation, and team members (Zhang et al., 2019). STOs' companies do not typically have a long history of tangible information (i.e., past positive earning streams or revenue) to facilitate forecasting cash flows, thus the investors infer the expected company's economic/business value from ample amounts of intangible information available in it. The information contained in the white papers impacts the prediction of the success of the token offering (Meoli and Vismara, 2022). In this regard, white papers are the primary form of voluntary disclosures used by businesses to reduce a perceived asymmetry of information between insiders (management, the entrepreneur, or shareholders) and third parties (lenders, customers, funding subjects, etc.).

In the specific case of token offerings, investors are faced with hard decisions (i.e., choosing high-quality businesses via a complex decision-making process) with no or few regulatory requirements regarding the level and depth of information the issuing firm must provide. Compared to more regulated IPOs, the information asymmetry between fundraisers and investors is higher in token offerings because of the large variation between projects and the projects' complicated fundamental technical details (Chen, 2019; Yen, 2021). Information asymmetry in 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) even when the disclosures are viewed as only partially credible (Florysiak and Schandlbauer, 2022).

White papers aim to provide investors with the minimum amount of information needed to facilitate an effective investment decision. Few studies exist on STOs, whereas plenty of literature has been developed on ICOs. Our paper is related to this rapidly growing stream of literature and, in particular, on the role of white papers in token offerings.

Literature on ICOs explored whether the presence of white papers impacts funding success. Except for two isolated studies (Adhami et al, 2018; Boreiko and Vidusso, 2019), most studies identify a positive impact of white paper availability on the amount raised (Boreiko and Sahdev, 2018; Cerchiello et al., 2019; Chen, 2019; Fish, 2019).

Focusing on the white paper structure, the first group of studies investigates the role of white paper readability. Highly readable documents are generally considered to be 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). Different readability measures have been used, such as Fog rate and the lengths of sentences and words (Zhang et al., 2019), or the length of the entire document (Samieifar and Baur, 2021). More readable white papers are associated with a higher initial return for ICO investors (Zhang et al., 2019), and publishing a longer white paper improves the likelihood of completing the offering and increases the amount raised during the campaign (Amsden and Schweizer, 2019). Moreover, the average length of white papers has 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, 2021). Florysiak and Schandlbauer (2022) and Yen et al. (2021) investigate white papers' content information structure. Florysiak and Schandlbauer (2022) distinguish between standard and informative white papers and find that informative content is unrelated to the trading volume generated after the campaign. However, Yen et al. (2021) show that ICOs with unique content raise more funds or are subject to more active trading and higher market values in the post-ICO period.

Momtaz (2020) introduces a second stream in the literature that explores whether the style of language used in white papers influences the funding of new ventures. The study quantifies informational exaggeration in ICO white papers 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.

The third stream of studies focuses on white paper quality, measured through a dichotomous categorization based on the presence of specific information and sections or by quantifying the white paper information content. For example, white papers with a high number of 'technical' words (for instance, 'block,' 'node,' and 'ledger') are considered of higher quality (Bourveau et al., 2022; Lyandres et al., 2019). Chen (2019), Feng et al. (2019), and Fish (2019) provide evidence that offerings' technical details in the white paper can effectively signal the quality of an ICO project, with a positive effect on the amount of funds raised. Another measure of white paper quality is the presence of a detailed project evaluation, shown to be positively related to the duration of the ICO (Blaseg, 2018; Boreiko and Sahdev, 2018) and its secondary market liquidity and trading volume

(Bourveau, et al., 2022; Howell et al., 2020). Finally, other measures of white paper quality are whether a section discussing risk (Konstantinidis et al., 2018) or a roadmap (Blaseg, 2018) are included in the text.

Our study differs from previous studies in the following two ways. First, we focus specifically on STOs. As Lambert et al. (2021) highlight, STOs are not a subset of ICOs but differ in nature and scope. Specific investigation is therefore required to assess the underlying success factors of this nascent and promising market, but such literature is scarce, limited to three studies addressing other aspects of these tokens, namely cheap or fake quality signals (Ante and Fiedler, 2019), or the functioning of the market and the critical success factors, evidenced in the literature on entrepreneurial finance (Beinke et al., 2021; Lambert et al., 2021).

Second, no previous studies on ICOs examine the relationship on the fundraising success of topical features mined from textual descriptions of projects. We argue that it is necessary to explore how investor behaviours are influenced by information content published during the campaign. The importance of the information content disclosed in a white paper is higher for STOs than ICOs due to the financial nature of this kind of digital asset. To the best of our knowledge, we are the first to address this issue.

2.2 Hypotheses development

According to signalling theory (Spence, 1973), voluntary disclosure in corporate reporting is one of the crucial means of signalling and allows companies to disclose more information than is mandatory to signal their quality (Campbell et al., 2000; Verrecchia, 1990). Disclosure reduces the information asymmetry between insiders and investors and the associated cost of capital (Diamond and Verrecchia,1991).

Regarding the voluntary disclosure for high-technological companies, it comprises the underlying technological components in the form of patents, R&D expenditures, or the innovation introduced by the business (Simpson and Tamayo, 2020). Firms may elect to disclose information regarding their technological and innovative characteristic to reduce information asymmetry between the manager and the investors and thus facilitate the financing process. At the same time, as shown by Bhattacharya and Ritter's (1983) model, technological firms are subject to a trade-off between disclosing innovation and secrecy. Revealing information about firms' technology leads to better financing terms but at the same time reduces the firm's technological advantage over competitors. For this reason, investors generally have concerns about the credibility of disclosure by high-tech firms, because to avoid litigation risk and proprietary costs, managers may withhold key information or issue biased disclosure (Gu and Fi, 2007).

Industry conditions are also determinants of firms' decisions on whether to disclose or to keep their innovation secret (Saidi and Žaldokas, 2021). For example, in the fintech industry, Chen et al. (2019) find that the disclosure of specific technologies such as the Internet of things (IoT), Robo-advising, and blockchain generates the highest valuable stock market responses. The disclosure of IT announcements issued by an enterprise can be regarded as a signal that the enterprise is willing to participate in the technology management transformation process and business model innovation (Klockner et al., 2022), which should lead to better future performance for the enterprise (Cahill et al., 2020), an increase in the investors' attention, and a higher stock market value (Cahill et al., 2020; Liu et al., 2022).

In light of this background, voluntary disclosures on relevant technology in the white paper may impact the evaluation of the venture in STO market. In token offerings, the presence of a patent, the presence of a technical description of the project, or a GIThub code often captures the technological capabilities of the venture (Bourveau et al., 2022; Chen, 2019; Lambert et al., 2021; Meoli and Vismara, 2022; Roosenboom et al., 2020). As a matter of fact, in ICOs, the release of information about the underlying blockchain technology positively affects different corporate performance indicators (Chen, 2019; Yen and Wang, 2021), such as short-term market reaction (Akyildirim et al., 2020; Cahill et al., 2020; Cheng et al., 2019) or long-term firm evaluation (Yen and Wang, 2021). Moreover, ICO investors' technological motives are the most important motives guiding their investment decision, followed by financial and ideological motives (Fish et al., 2021). In Fish et al.'s (2021) paper, 'technological motives' comprise items such as 'personal enthusiasm for the technology of the ICO venture' and 'personal enthusiasm for the business model/idea.'

We argue that notwithstanding the differences between the two markets and types of tokens, the signal of the technological component of the project and its application in the business are crucial for STO ventures and investors. Therefore, we claim that the likelihood of an STO's success is related to the prevalence in the project description of technological concepts, especially those that improve investors' understanding of the project's underlying quality and token characteristics.

Hp1: In the white paper, a high share of information on relevant technologies is positively related to the probability of campaign success.

Among voluntary disclosure, the companies' environmental impact has attracted in recent years increasing stakeholder attention. The majority of large businesses volunteer information concerning the effects of their activities on the environment and how these impacts are managed within the firm (Hackston and Milne, 1996; Roberts, 1992). Signalling theory views environmental disclosure as a pre-emptive step to mitigate adverse regulatory or legislative pressures in the future (Brammer and Pavelin, 2008; Lourenco et al., 2014): it signals compliance with societal norms concerning sustainable business conduct, which is assumed to lead to increased legitimacy, reputation, and reduced risks (Reber et al., 2021). Moreover, environmental disclosure helps companies in attracting more investments (Verrecchia, 1990). Studies suggest that investors consider environmental disclosures as a relevant component for measuring a company market value (Cormier and Magnan, 2007; Cormier et al., 2011), and the environmental disclosure increases the company's chances to survive and thrive during an IPO and in the aftermarket (Reber et al., 2021). To sum up: signalling the company's environmental impact serves multiple functions, such as communicating the orientation of the company to sustainable practice, avoiding risk, and attracting sustainable investors.

About the latter, 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. 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. Equity crowdfunding studies have evidenced the benefits of signalling project environmental or social orientation in the campaign presentation with a positive impact on the amount raised and on the number of involved investors (Butticè et al., 2019; Calic and Mosakowski, 2016; Vismara, 2019).

These previous studies lead us to formulate our second hypothesis:

Hp2: In the white paper, a high share of information on a company's environmental impact is positively related to the probability of campaign success.

3. Sample and method

3.1 Sample

Following previous research (Adhami et al., 2018; Lambert et al., 2021; Lyandres, 2019), we hand-collected information on token offerings from multiple well-known STOs aggregators: Coinitelligence.com, Tokenmarket.net, Blockdata, STOscope.com, STOrating.com, STOwise.com, STOcheck.com, STOAnalytics, and ICObench.com. We manually matched data across various sources to generate a sample of uniquely identified offerings, yielding a sample of 430 STOs issued from December 2017 to February 2021. The white papers related to these offerings were then downloaded from aggregators and from the issuing company's website when present. Because some aggregators delete some campaign information affecting the availability of such information, we used

Google to search for white papers not available from aggregators or issuing companies. We selected only white papers written in the English language. In total, we obtained 193 STO white papers[3]. Following Lambert et al. (2021), we considered only 'true' STOs and excluded ICOs registered as STOs and stable coins (five in total). Therefore, our final sample consisted of 188 STO white papers, with 77% of campaigns closed successfully and an average amount raised of $21.5 million. Table 1 shows descriptive statistics for our sample.

Regarding geographical area, 54% of the issues are from non-European countries[4], 46% are from European countries. Looking at the sample by industry group, referring to the one-digit Standard Industrial Classification (SIC) code categories, most companies that issue STOs are in the finance, insurance, and real estate (45% of the sample), manufacturing (23%), or services (17%) sectors. Smaller numbers are in the mining and construction (2%), retail and wholesale trade (6%), and media and communication (6%) sectors. The majority of STOs issues occurred in 2018 and 2019 (90% of our sample).

Table 1 also shows other characteristics of the white papers in our sample. The average white paper has 2,996 words (approximately five pages) and a size of 5,432 KB. Readability, measured by fog rate,[5] is about 15, which is higher than the fog rate found for ICO white papers (13.70, Samieifar and Baur, 2021). Only 22% of issuers present additional documents, e.g., a summary version of the white paper. On average, white papers include six pictures and about two tables in terms of visual elements.

Finally, in terms of campaign duration and other project characteristics, the average period of each offering is 101 days, only 36% of the issuers provide a pre-sale offering, and 63% indicate a soft cap (the minimum amount of funding required for the offering to be considered successful).

[3] The sample is larger and more up-to-date than many comparable recent studies, such Ante and Fiedler (2019), who collected 151 STOs, and Lambert et al. (2021), who considered 124 STOs up to 2018.

[4] Non-European countries include: North and South America; Mena (Middle East-North Africa) countries, Russia, Asia, Australia, fiscal paradise countries (Panama, Liechstenstein, Cayman Islands)

[5] The fog rate is calculated as [(average number of words per sentence+number of words of 3 syllables or more) ∗ 0.4]. A fog rate of 18 suggests the text is unreadable, and the ideal value is between 10 and 12.

Table 1
Descriptive statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
Success (d)	188	0.77	0.42		
Amount (ml$)	188	21.5	44	0	350
Extra_Europe (d)	188	0.54	0.50		
Europe (d)	188	0.46	0.45		
Agriculture, forestry and fishing (d)	188	0.01	0.14		
Finance, insurance, real estate (d)	188	0.45	0.50		
Manufacturing (d)	188	0.23	0.42		
Mining and Construction (d)	188	0.02	0.13		
Services (d)	188	0.17	0.38		
Media and Communication (d)	188	0.06	0.25		
Retail and Wholesale trade (d)	188	0.06	0.24		
Year 2017 (d)	188	0.02	0.22		
Year 2018 (d)	188	0.34	0.48		
Year 2019 (d)	188	0.56	0.50		
Year 2020 (d)	188	0.07	0.26		
Year 2021 (d)	188	0.01	0.07		
Words	188	2,996.88	1,271.36	110	4,608
Size (Kb)	188	5,432.69	9,360.04	75	100,558
Fog rate	188	13.31	2.58	7.83	31.36
Extra doc. (d)	188	0.22	0.42		
Pictures	188	6.4	5.67	0	26
Tables	188	1.91	2.61	0	20
Soft cap use(d)	188	0.63	0.48		
Offering days	188	101.65	81.48	6	548
Pre-sale offering (d)	187	0.36	0.48		

d): dummy variable

3.2 Method

The analysis presented in this paper is structured in two stages. First, we apply LDA (described in Section 3.3) to answer our first research question, '*Which topics have characterised STO white papers so far?*' We then apply probit and Tobit regressions (described in Section 3.4) to answer our second research question "*Which of these topics influenced campaign success?*" and to test Hp1 and Hp2.

3.3 Identifying topics with LDA

We leverage LDA to identify the main topics discussed in white papers and characterise each document according to its most significant topic. LDA is a generative probabilistic topic model that fully automates the production of a probability distribution of topics in a corpus (Hannigan et al., 2019, Savin et al., 2020). Like factor or cluster analysis, LDA uses statistical associations of words to generate a latent thematic structure within a large amount of text (Blei et al., 2007, 2010; Chen, 2011; Savin et al., 2021a, 2021b). Each latent topic is a cluster of a bag-of-words that jointly represent a theme in the text and emerges without the aid of pre-defined, explicit dictionaries or interpretive rules (Hannigan et al., 2019). The method relies on the fact that words frequently appearing together tend to be semantically related (Storopoli, 2019). This process reduces researcher bias because 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 (Li et al., 2021; Wang et al., 2020; Yen and Wang, 2021; Savin et al., 2022a, 2022b). The application of LDA in the categorisation of ICO white papers is a very recent development, and it has been used in only two recently published papers. Chuanjie et al. (2019) apply data science techniques to categorize topics in ICO white papers, and Bian et al. (2018) use LDA to categorize ICOs into ten different topics based on a manual labelling process. The main novelty of the present study is to analyse the relationship between topics identified by LDA and STO funding success. LDA is particularly suitable for this task because it does not impose dictionaries, a predefined content structure, or interpretative restrictions on the researcher.

In order to apply LDA to the STO white papers, we first pre-processed the corpus by 1) converting words to lowercase, 2) removing standard English stop words and punctuation, and 3) lemmatizing all the words by means of the Natural Language Toolkit[6] lemmatiser. We then analysed the distribution of terms with domain experts and filtered out generic terms that appeared in more than 60% of the white papers (Zhang et al., 2021). Following Chuanjie et al. (2019), we also discarded a set of very common terms in this domain, such as 'platform,' 'user,' 'service,' 'system,' 'network,' 'contract,' 'exchange,' and 'business'.

LDA takes as its inputs a set of documents, represented as a bag of words, and the desired number of topics *K*, and returns a set of *K* topics. It is typically advisable to run the model multiple times with different values of *K* and then use a measure of topic quality, such as topic coherence, and a manual inspection by domain experts to assess the best number of topics.

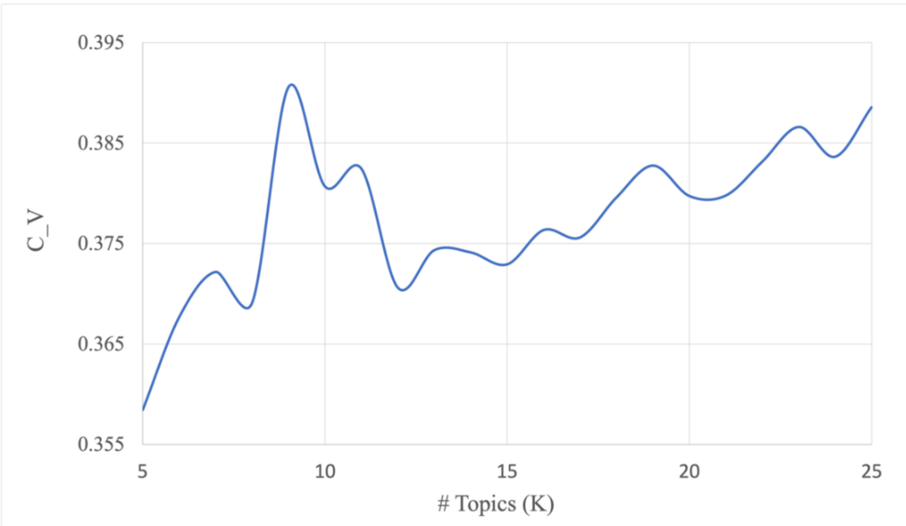
To assess the number of topics, we followed this procedure and ran LDA 20 times for each value of K for $5 \leq K \leq 25$. We then analysed the resulting models according to two coherence metrics[7]: *C_V* (Röder et al., 2015) and *C_NPMI* (Bouma, G., 2009), based on normalized pointwise mutual information. These metrics measure the degree of semantic similarity between high-scoring terms in each topic and are typically used for assessing the quality of LDA topics. In contrast with *perplexity* (Blei et al., 2007), that is sometimes used for the same purpose, they proved to be fairly aligned with

[6] NLTK - https://www.nltk.org/
[7] Computed using the CoherenceModule of the Gensim library with default settings.

the human assessment (Röder et al., 2015). In particular, C_V is used as default by the well-known Python library Gensim[8].

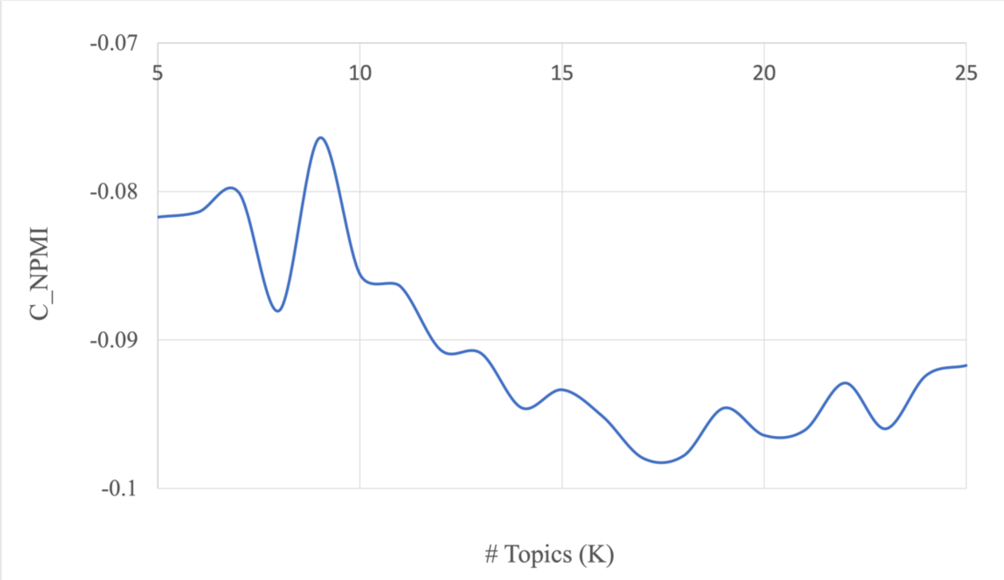
Figure 1 and Figure 2 report respectively the average C_V and C_NPMI for each value of *K*. C_V increases with the number of topics until *K=9*, which is the highest peak, it then lowers down to *K=12* and raises slowly again. C_NPMI has a similar trend, peaking at *K=9,* lowering to *K=17*, and then raising again. While C_V and C_NPMI are generally trusted metrics, in this phase it is also crucial to consider the interpretability of the topics (Zhang et al., 2021). We thus carefully inspected the results manually and concluded that *K=9* was indeed the best solution since the resulting topics were easily interpretable and several of them are aligned with well-known themes in this space, already identified by previous works, such as 'Energy and green issues' (Zhang et al., 2021) and 'AI and machine learning' (Bian et al. 2018; Chuanjie et al., 2019).

Fig. 1. Average C_V coherence for number of topics (*K*).



[8] The Gensim library - https://github.com/RaRe-Technologies/gensim

Fig. 2. Average C_NPMI coherence for number of topics (*K*).



The topics produced by LDA are characterized only by a set of weighted terms, so it is normal practice to assign them simple human-readable labels. To do this, we analysed each topic's distribution of the most significant terms and produced a relevant label, following the methodology of Zhang et al. (2021).

Table 2 reports the labels and the most significant terms of the nine topics. Two of the topics are related to the presentation of the issuing company and financial and legal matters: *company description (T2)* and *financial and legal issues (T8)*. Six topics represent various technological aspects and can thus be used to test our first hypothesis (Hp1, discussed in Section 2.2). Specifically, two topics address general technologies that are often mentioned in white papers: *AI and machine learning (T3)* and *DLT components (T5)*. Four topics refer instead to the use of technology in specific domains relevant to the STOs: *art and education (T4)*, *construction and energy (T6)*, *healthcare (T7)*, and *finance and other services (T9)*. Finally, the topic of *energy and green issues (T1)* addresses energy use and environmental implications of DLT, as well as specific projects with an environmental focus, and can thus be used to verify our second hypothesis (Hp2).

For the sake of reproducibility, we make available the LDA model, the distribution of the topics in the STOs, and the results of the empirical analysis (discussed in Section 4) in a FigShare repository[9].

[9] For the sake of the double anonymized review process, we temporarily set up a private anonymous link (https://figshare.com/s/e1c357b9f88d638c2c5c). We will make the repository publicly available after publication.

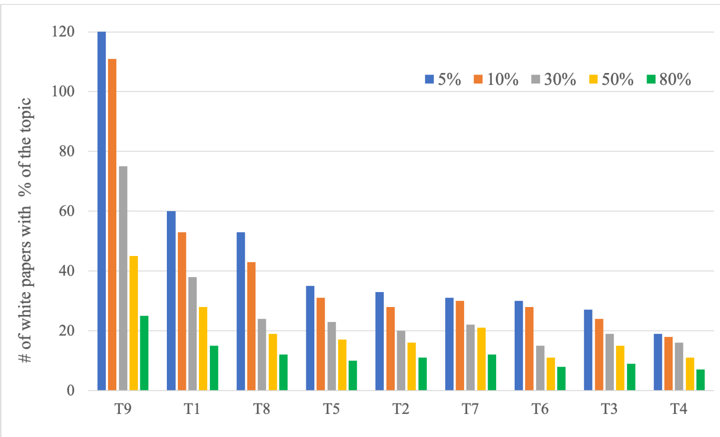
Table 2.
The 9 topics obtained from the STO white papers.

Topic	Label	Significant terms
T1	Energy and green issues	Energy, green, mining, electricity, production, capacity, wind
T2	Company description	Start-up, peer, brand, production, series, venture, exit, consumer
T3	AI and machine learning	Compute, film, equipment, cloud, computing, learn, ai, device
T4	Tech for art and education	Algor, art, learn, ai, chain, education, artwork
T5	DLT components	Chain, quantum, node, block, Algor, hash, protocol
T6	Tech for construction and energy	Plant, gas, material, construction, oil, turbine, engineer
T7	Tech for healthcare	Patient, healthcare, medical, data, care, compute, app
T8	Financial and legal issues	Issuer, bond, tax, portfolio, prospectus, purchaser, income
T9	Tech for finance & other services	Game, player, vote, app, banking, credit, easy, web

For each white paper, we calculated the posterior probabilities of the nine topics to assign a distribution of the topics where the probabilities sum up to one. For example, the white paper for the 'Minthealth' project, a healthcare platform for creation of a patient health identity, is prevalently associated with T7, 'tech for healthcare,' whereas the white paper for 'Windmine', a project that proposes a wind park, is mainly associated with T1, 'energy and green issues.'

Figure 3 shows the number of white papers that are associated with each topic with a percentage of at least 5%, 10%, 30%, 50%, and 80%. The most common topics are tech for finance and other services (T9), energy and green issues (T1), and financial and legal issues (T8).

Fig. 3. Topic coverage – number of white papers associated with each topic (considering a 5%, 10%, 30%, 50%, and 80% threshold).



One of the most interesting results of our descriptive analysis concerns the content distribution in the white papers. Table 3 shows the coverage of the nine topics. 'Tech for finance and other services' is the most prevalent topic (present in 28% of the total text), followed by 'energy and green issues' (15%) and 'financial and legal issues' (11%). Other topics related to the application technology in specific industries have low percentages (between 5% and 9%), as does 'company description' (8%).

Table 3
Topic analysis

Variable	Obs.	Mean	Std. Dev.	Min	Max
T1 Energy and green issues	188	0.15	0.28	0	1
T2 Company description	188	0.08	0.22	0	1
T3 AI and machine learning	188	0.07	0.21	0	1
T4 Tech for art and education	188	0.06	0.18	0	1
T5 DLT components	188	0.09	0.24	0	1
T6 Tech for construction and manufacturing	188	0.07	0.20	0	1
T7 Tech for Healthcare	188	0.09	0.25	0	1
T8 Financial and legal issues	188	0.11	0.24	0	1
T9 Tech for finance and other services	188	0.28	0.32	0	1

3.4 Econometric models

To measure the impact of white paper topics on the likelihood of STO success, we used a probit model (eq. 1), a nonlinear model where the dependent variable (Y) is a dichotomous variable. The likelihood of STO success was modelled by a dummy variable that equals 1 if the total funds raised in the STO are greater than the minimum cap, if available (Moro and Wang, 2019), or if the STO has raised any capital where no minimum cap was specified in the campaign, and 0 otherwise. Following Lambert et al. (2021) and Ante and Fiedler (2019), we identified a second measure of success as the logarithm of the total amount raised[10] and performed a Tobit regression, a censored regression model designed to estimate linear relationships between variables when there is either left- or right-censoring in the dependent variable (eq. 2).

In the probit model, we estimate the probability that firm *i* achieves a successful issue using the following equation:

[10] As in Lambert et al. (2021), the variable *Amount raised* is left-censored because a failing STO generally has a value of zero in our sample, which justifies the use of the Tobit model.

$$Prob\text{-}Y_i = 1 = \alpha + \sum_i^N \beta_i X_i + \sum_i^N \gamma_i Z_i + J_i + W_i + U_i + \varepsilon_i \ . \qquad \text{(eq. 1)}$$

In the Tobit model, we estimate the effects of the covariates on the amount raised by firm *i* using the following equation:

$$Amount\ raised_i = \alpha + \sum_i^N \beta_i X_i + \sum_i^N \gamma_i Z_i + J_i + W_i + U_i + \varepsilon_i \ , \qquad \text{(eq.2)}$$

where the dependent variable is either *successfully closing* the STO campaign for firm *i* (probit model eq. 1) or *Amount raised*$_i$ by the STO campaign, expressed as a natural logarithm (Tobit model eq. 2), α is a constant term, *Xi* is a vector of variables describing the white paper content, and *Zi* is a vector of variables describing the offering characteristics.

Specifically, vector *Xi* includes: i) measures of 'readability'—following previous literature (Samieifar and Baur, 2021; Zhang et al., 2019), represented by the logarithm of the number of words in the document (*Words*), and the logarithm of the document's size in KB (*Size*); ii) the presence of additional documents published with the white paper e.g., a short version of the white paper that briefly synthesizes the key offering points (*Extra doc.*); iii) proxies for the presence of visual elements in the white paper, including the number of pictures (*Pictures*), and tables (*Tables*); iv) indicators for the nine topics identified by LDA—we created nine dummy variables, *T1* to *T9*, where each variable equals one if the percentage of words in the document related to the specific topic is greater than 5%, and 0 otherwise. This threshold of 5% was calculated after eliminating generic words. We included all nine indicator variables simultaneously in our analysis.

We consider T1 'energy and green issues' to test Hp2 while multiple topics are related to Hp1 - technological components, and the adoption of the technology implemented in specific sectors- such as those from T3 to T7 and T9.

Vector *Zi* includes soft cap use; offering days, and presale dummy (Ante and Fiedler, 2019; Hornuf et al., 2021). According to Lambert et al. (2021) we selected these variables to account for important characteristics of STO while maximizing sample size. Following Fish (2019), we also controlled for sectorial distribution: Agriculture, Forestry, and Fishing; Finance, Insurance, Real Estate; Manufacturing; Mining and Construction; Services; Media and Communication; Retail and Wholesale trade (*Ji*), geographical: European and Non-European area, where the locational status denotes where the token is issued; (*Wi*), and time-fixed effects from 2017 to 2021 (*Ui*).

To check the robustness of our results we ran the models defined in equations 1 and 2 using different measures of topics. We measured the prevalence of the topics in the various models as: i) the percentage of words related to the topic in the total words of the white paper after deleting generic words; ii) the logarithm of that percentage, calculated as ln(1+topic percentage) following Yen and Wang (2021); iii) one of nine dummy variables, representing *T1* to *T9*, where each variable equals 1 if the percentage of words in the document related to the specific topic is greater than 10%, and 0 otherwise.

Table A.1 in the Appendix provides detailed variable definitions.

4. Empirical results

4.1 Main results

This section reports the results of our analysis. Table 4 shows the main differences between successful and unsuccessful STOs. White papers of successful STOs have smaller size (KB), present additional documents, and are more likely to indicate a soft cap, compared with white papers of unsuccessful campaigns. In terms of themes, successful STOs most frequently concern technology in the healthcare field, according to the content of white papers (T5).

Table 4
Successful and Unsuccessful STOs

		Success					Unsuccess					
Variable	**Obs.**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Obs.**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Difference in means**	
Amount (000$)	145	27,800	48,400	19,800	350,000	43	186.05	859.286	0	4,500	27,614	***
Words	145	3,060	1,252	275	4,608	43	2,782	1,324	110	4,565	278	
Size (KB)	145	4,540	5809	75	39,155	43	8,443	16,196	158	100,558	-3,903	**
Fog rate	145	15.37	2.84	9.21	33.43	43	15.16	2.10	9.39	19.82	-0.23	
Extra doc. (d)	145	0.26	0.44	0	1	43	0.09	0.29	0 1		0.17	**
Pictures	145	7	6	0	26	43	6	5	0	25	1.00	
Table	145	2	2	0	12	43	2	4	0	20	0.00	
Soft cap use (d)	145	0.7	0.5	0	1	43	0.3	0.5	0	1	0.40	***
Offering days	145	100	80	6	366	43	115	96	17	548	-15.00	
Pre-sale offering (d)	144	0	1	0	1	43	0	0	0	1	0.00	
Extra_Europe (d)	145	0.51	0.5	0	1	43	0.67	0.5	0	1	-0.10	
Europe (d)	145	0.49	0.4	0	1	43	0.33	0.3	0	1	0.16	*
T1 Energy and green issues	145	0.16	0.28	0	1	43	0.10	0.3	0	1	0.06	
T2 Company description	145	0.07	0.20	0	1	43	0.12	0.3	0	0.9	-0.05	
T3 AI and machine learning	145	0.06	0.19	0	1	43	0.10	0.3	0	1	-0.04	
T4 Tech for art and education	145	0.06	0.18	0	1	43	0.06	0.2	0	1	0.00	
T5 DLT components	145	0.10	0.25	0	1	43	0.07	0.2	0	0.8	0.03	
T6 Tech for construction and energy industry	145	0.06	0.18	0	1	43	0.10	0.3	0	1	-0.04	
T7 Tech for Healthcare	145	0.11	0.27	0	1	43	0.02	0.1	0	0.8	0.09	**
T8 Financial and legal issues	145	0.11	0.24	0	1	43	0.11	0.2	0	1	0.00	
T9 Tech for finance and other services	145	0.27	0.32	0	1	43	0.32	0.3	0	1	-0.05	

The Table represents the descriptive statistics of a sample of 188 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 shows the results of the probit and Tobit regression models.

Model 1 includes the control variables "readability, extra documents, and visual elements". The analysis shows that heavy documents (measured in KB) have less probability of success, but additional documents (e.g., a white paper summary) presented with the white paper increase campaign success. The provision of additional documents increases the probability of reaching the minimum cap by 16%. This suggests that voluntary disclosure of an additional short version of the document is perceived as a positive signal by STO investors and enhances the white paper comprehension. Other control variables do not show statistically significant coefficients.

Model 2 further includes an analysis of topics' effects. Topic prevalence is measured using dummy variables that equal 1 if the percentage of words in the document related to the specific topic is greater than 5%, and 0 otherwise. Our initial results regarding document size and the provision of additional documents are confirmed; moreover, also the length of the document (words number) increases campaign success.

The model further identifies the effects of the different topics. The topics that is related with the STO success are the energy use and green issues (T1) and technology applied in the healthcare sector (T7); the presence of these topics increases rates of success by 13% and 18%, respectively. Models 3 confirms the influence of T1 and T7 when measuring STO success by the total amount raised. When the topic of energy and green issues is identified in more than 5% of total words, the total amount raised increases by \$10, and the same prevalence of the topic of tech components in the healthcare sector increases the amount raised by \$30.

Models 4 and 5 also feature three control variables: the presence of a soft cap, the duration of the campaign (offering days), and the presence of a pre-sale offering. Both models confirm the relationship of T1 and T7 with the probability of success and the total amount raised. Indicating a soft cap in the campaign increases the probability of success by 21% and the amount raised by \$96. We do not find statistically significant effects for campaign duration or the presence of a pre-sale offering.[11]

[11] We also run the models also with a control for fog rate and find the effect of this variable is not statistically significant. These results are available in Appendix, Tab A.2.

Table 5
Main Analysis

VARIABLES	**Model 1 Success**	**Model 1 Margins**	**Model 2 T5% Success**	**Model 2 T5% Margins**	**Model 3 T5% Amount Raised**	**Model 4 T5% Success**	**Model 4 T5% Margins**	**Model 5 T5% Amount Raised**
Words	0.146	0.037	0.323*	0.073*	1.445*	0.349*	0.0751*	1.523*
	-0.166	-0.042	(0.183)	(0.041)	(0.852)	(0.198)	(0.044f)	(0.840)
Size	-0.245**	-0.061**	-0.217*	-0.049*	-0.687	-0.212*	-0.046*	-0.502
	-0.111	-0.028	(0.115)	(0.026)	(0.453)	(0.124)	(0.0273)	(0.449)
Pictures	0.032	0.008	0.035	0.008	0.141	0.034	0.00730	0.122
	-0.023	-0.006	(0.023)	(0.005)	(0.180)	(0.022)	(0.00495)	(0.0747)
Tables	-0.063	-0.016	-0.050	-0.0115	-0.146	-0.048	-0.0104	-0.123
	-0.042	-0.011	(0.042)	(0.009)	(0.210)	(0.049)	(0.0106)	(0.184)
Extra Doc.	0.779***	0.156***	0.891***	0.154***	2.457**	0.665**	0.116**	1.523*
	-0.297	-0.047	(0.302)	(0.043)	(0.954)	(0.331)	(0.0498)	(0.907)
T1 Energy and green issues			0.662**	0.130***	2.430**	0.684**	0.127***	2.363**
			(0.276)	(0.045)	(0.941)	(0.291)	(0.0468)	(0.930)
T2 Company description			-0.047	-0.011	-0.453	0.228	0.0451	0.957
			(0.315)	(0.074)	(1.371)	(0.358)	(0.0653)	(1.333)
T3 AI and machine learning			0.225	0.047	0.564	0.371	0.0683	1.303
			(0.345)	(0.065)	(1.447)	(0.371)	(0.0594)	(1.428)
T4 Tech for art and education			0.014	0.003	-0.255	-0.366	-0.0916	-1.496
			(0.405)	(0.091)	(1.572)	(0.407)	(0.115)	(1.586)
T5 DLT components			0.299	0.061	0.829	0.361	0.0680	0.991
			(0.351)	(0.063)	(1.130)	(0.363)	(0.0601)	(1.214)
T6 Tech for construction and manufacturing			-0.154	-0.037	-0.317	-0.385	-0.0949	-1.455
			(0.298)	(0.074)	(1.321)	(0.322)	(0.0883)	(1.231)
T7 Tech for Healthcare			1.225***	0.176***	3.425***	1.212***	0.163***	3.635***
			(0.411)	(0.038)	(1.024)	(0.428)	(0.0379)	(1.077)
T8 Financial and legal issues			0.112	0.025	1.444	0.134	0.0278	1.371
			(0.280)	(0.060)	(1.065)	(0.292)	(0.0589)	(0.991)
T9 Tech for finance and other services			0.125	0.029	-0.136	0.045	0.0107	-0.380
			(0.274)	(0.064)	(1.083)	(0.291)	(0.0633)	(1.002)
Soft cap use						0.859***	0.207***	4.564***
						(0.296)	(0.0796)	(1.093)
Offering days						-0.003	-0.001	0.344

				(0.001)	(0.000)	(1.003)
Pre-sale offering				0.162	0.034	-0.011
				(0.269)	(0.055)	(0.007)
Constant	10.670***	7.830***	11.150	8.400***		12.110
	(1.680)	(1.900)	(8.210)	(2.094)		(7.770)
Geo FE	Yes	Yes	Yes	Yes		Yes
Sector FE	Yes	Yes	Yes	Yes		Yes
Year FE	Yes	Yes	Yes	Yes		Yes
Pseudo R2	0.16	0.23	0.03	0.31		0.05
ROC AUC	0.77	0.82	.	0.86		.
Prob > chi2	0.36	0.22	.	0.65		.
Obs.	187	187	188	187		187

Table 5 shows the results of the probit and Tobit regression models with with '***', '**' and '*' representing the 1%, 5% and 10% levels, respectively. The dependent variables are: a dummy variable, which equals 1 when the STO closes with success, and the total amount raised in the campaign. Time, regional and sector fixed effects are three vectors that capture differences at chronological, geographical, and sectorial levels. The 'Success' columns report the coefficient of the probit regression; the 'Margins' columns report the marginal effects of the probit regression. Model 1 introduces the control variables about documents' readability, Model 2 and 3 present the effect of topics on the dependent variables. From Model 2-5 topics are dummy variables with value 1 if the topic is dealt at least 5% in the document. Model 4 and 5 introduce three control variables: soft cap use, offering days and pre-sale offering.

4.2 Additional checks

To verify the stability of our results we performed several robustness checks adopting alternative methods for representing the topics. Our results are reported in Table 6. Models 6 and 7 measure the amount of focus on each topic as a percentage, whereas Models 8 and 9 use the logarithm of the percentage, calculated as ln (1+ topic percentage), as suggested by Yen and Wang (2021). Models 10–11 measure topics using dummy variables with 10% thresholds instead of the 5% threshold used in the main model. This approach allows us to consider how prominent the topic is in the white paper.

In Models 6–9, only T7, technology in the healthcare sector, is statistically significant. An increase of 1% in the proportion of words related to technology in the healthcare sector compared to the proportion of words related to the company's description (baseline variable) increases the probability of campaign success by 44% (Model 6) and the amount raised by $1,459 (Model 7). The significance of the topic is confirmed also when it is measured in logarithm form (Models 8 and 9). These results indicate that all topics found to be relevant in the main analysis (e.g., T1, 'energy and green issues') must comprise a percentage of total words of at least 5% to be an effective influencing factor; if a topic is mentioned but not sufficiently discussed in the text, it does not impact campaign success.

Models 10–11 confirm all results from our main analysis. The more extensive discussion of T1 and T7 is not associated with an increase in the probability of success, which remains the same as in our main models (compare with Model 2), nevertheless, we detect a slight positive effect on the amount raised. If the topic 'energy and green issues' represents at least 10%, the total amount raised increases from $10 to $13 (compared to the main Model 3). T7 'technology for healthcare' has a higher economic impact than T1. If words related to technology in the healthcare industry represent at least 10% of the document, the amount raised increases by $35.

Table 6
Additional tests

Variables	Model 6 T% Success	Model 6 T% Margins	Model 7 T% Amount raised	Model 8 Tln Success	Model 8 Tln Margins	Model 9 Tln Amount raised	Model 10 T10% Success	Model 10 T10% Margins	Model 11 T10% Amount raised
Words	0.220	0.0501	1.234	0.239	0.0541	1.293	0.363*	0.0814**	1.563*
	(0.181)	(0.0411)	(0.794)	(0.181)	(0.0410)	(0.783)	(0.189)	(0.0415)	(0.815)
Size	-0.214*	-0.0486*	-0.534	-0.211*	-0.0478*	-0.508	-0.224*	-0.0502*	-0.724
	(0.114)	(0.026)	(0.504)	(0.115)	(0.0261)	(0.500)	(0.115)	(0.0259)	(0.458)
Pictures	0.0301	0.00684	0.130	0.0301	0.00681	0.133	0.0311	0.00698	0.119
	(0.022)	(0.00529)	(0.0851)	(0.0232)	(0.00534)	(0.0869)	(0.024)	(0.00545)	(0.0833)
Tables	-0.057	-0.0131	-0.182	-0.0551	-0.0125	-0.172	-0.0523	-0.0117	-0.146
	(0.042)	(0.009)	(0.227)	(0.0420)	(0.009)	(0.228)	(0.0417)	(0.00940)	(0.201)
Extra Doc.	0.872***	0.152***	2.526**	0.892***	0.154***	2.573**	0.912***	0.155***	2.415**
	(0.307)	(0.0444)	(1.003)	(0.305)	(0.0433)	(1.001)	(0.314)	(0.0425)	(0.957)
T1 Energy and green issues	0.737	0.168	3.378	1.286	0.291	5.403	0.680**	0.128***	2.568**
	(0.686)	(0.154)	(2.596)	(0.876)	(0.195)	(3.409)	(0.305)	(0.0460)	(1.088)
T2 Company description							0.194	0.0404	0.521
							(0.344)	(0.0659)	(1.469)
T3 AI and machine learning	0.0318	0.00724	-0.0884	0.326	0.0739	0.899	0.124	0.0265	0.241
	(0.727)	(0.165)	(3.753)	(0.969)	(0.220)	(4.947)	(0.353)	(0.0716)	(1.529)
T4 Tech for art and education	-0.080	-0.0183	0.499	0.133	0.0300	1.446	-0.0616	-0.0142	-0.130
	(0.755)	(0.172)	(3.476)	(0.992)	(0.225)	(4.529)	(0.440)	(0.104)	(1.724)
T5 DLT components	0.692	0.157	2.325	1.120	0.254	3.839	0.526	0.0961*	0.925
	(0.676)	(0.155)	(2.772)	(0.867)	(0.199)	(3.664)	(0.328)	(0.0491)	(1.127)
T6 Tech for construction and manufacturing	-0.201	-0.0458	-3.788	-0.0494	-0.0112	-4.203	-0.185	-0.0445	-0.652
	(0.717)	(0.163)	(3.415)	(0.954)	(0.216)	(4.781)	(0.305)	(0.0769)	(1.340)
T7 Tech for Healthcare	1.945**	0.442**	7.286***	2.923**	0.663**	10.77***	1.309***	0.179***	3.578***
	(0.884)	(0.200)	(2.686)	(1.156)	(0.260)	(3.672)	(0.426)	(0.0382)	(1.150)
T8 Financial and legal issues	0.178	0.0406	2.203	0.444	0.101	3.681	0.0287	0.00638	0.742
	(0.737)	(0.168)	(3.350)	(0.943)	(0.215)	(4.439)	(0.320)	(0.0706)	(1.276)
T9 Tech for finance and other services	0.275	0.0624	1.752	0.554	0.126	2.916	-0.0967	-0.0215	-0.615
	(0.630)	(0.144)	(2.894)	(0.841)	(0.192)	(3.897)	(0.295)	(0.0649)	(1.032)
Constant	8.991***		10.62	8.549***		8.719	7.511***		11.26
	(1.881)		(7.823)	(1.909)		(8.104)	(1.961)		(8.031)

Country FE	Yes		Yes	Yes		Yes	Yes		Yes
Sector FE	Yes		Yes	Yes		Yes	Yes		Yes
Year FE	Yes		Yes	Yes		Yes	Yes		Yes
Pseudo R2	0.2		0.03	0.2109		0.03	0.24		0.03
ROC AUC	0.81		.	0.81		.	0.83		.
Prob > chi2	0.18		.	0.17		.	0.18		.
Obs.	187		188	187		188	187		188

Table 6 shows the results of the probit and Tobit regression models with with '***', '**' and '*' representing the 1%, 5% and 10% levels, respectively. The dependent variables are: a dummy variable, which equals 1 when the STO closes with success, and the total amount raised in the campaign. Time, regional and sector fixed effects are three vectors that capture differences at chronological, geographical, and sectorial levels. The ‘Success' columns report the coefficient of the probit regression; the 'Margins' columns report the marginal effects of the probit regression. Model 6 and 7 introduce topics in percentage terms. Model 8 and 9 topics are in logarithm terms. In Model 10 and 11 topics are dummy variables with value 1 if the topic is dealt at least 10% in the document.

5. Discussion and conclusions

The purpose of this study is twofold: 1) to explore which topics characterize STO white papers and 2) to verify whether the disclosure of relevant technologies (*Hp1*) and the company's environmental impact (*Hp2*) is related to the probability of campaign success. We identified nine topics from the white papers in our sample. As expected, some of these topics regard company descriptions and the financial and legal issues of the offering; others are related to the underlying technology adopted and the specific sector of the project under the campaign. Our experiments show that the topic of technology components in the healthcare field (T7) is consistently associated with STO funding, even when it has a relatively low coverage in the document (Figure 3), suggesting a strong interest from investors in this field. This seems to at least partially confirm Hp1. Other technological components instead do not seem to affect the likelihood of success. Therefore, Hp1 may require more investigation in future works. We also find that Energy and green issues (T1), which communicate the environmental company's features, significantly increases the probability of campaign success and the total amount raised, confirming *Hp2*.

In addition, our analysis evidence that the length of the document in terms of the number of words used to present the project and the offering positively impacts the campaign success confirming previous literature on ICO (Amsden and Schweizer, 2018; Fisch, 2019; Samieifar and Baur, 2021). On the contrary, a white paper that is too heavy (with a lot of pictures and tables) negatively influences campaign success. This could be ascribable at first to the fact that heavy documents are less accessible and downloadable, especially from mobile devices. Secondly, pictures and images may not always convey information about the campaign. For instance, in our sample of STOs we found images of evocative landscapes, buildings, and team photos. These kinds of figures increase the document size without conveying any additional information about the project. This reading justifies the non-significant relationship between pictures/tables and campaign success.

As a practical implication, it would be preferable to produce relatively lightweight documents with few (and possibly informative) images. Furthermore, providing additional documents (e.g., a summary of the white paper) positively influences both the probability of campaign success and the total amount raised. It thus may be advisable for the issuer to complement the white paper with a short textual summary as this happens with traditional financial instrument offerings such as mutual funds or exchange traded funds (ETFs), which are required by the regulations with the aim to facilitate readability in addition to the completeness of the information. This result is supported by previous literature that pinpoints the limited incentives for small investors to perform due diligence in the context of digital finance (Block et al., 2021).

All these results bring new insights to the nascent literature about determinants of STOs success (Ante and Fiedler, 2019; Beinke et al., 202; Lambert et al., 2021; Myalo, 2019) focusing on white paper content.

Our first contribution to the literature is related to the disclosure of technological components in token offerings. Previous studies of ICOs have provided evidence on specific technological signals, such as patents and the number of technical words in white papers on the funds collected (e.g., Florysiak and Schandlbauer, 2022; Howell et al., 2020; Samieifar and Baur, 2021). In STOs, technology disclosure is also an important subject for investors, and more disclosure on this theme does not refrain investors' interest; on the contrary, spending more space within the white paper to describe the underlying proposed technology seems to boost investors' participation, especially in the healthcare sector, that appears to be a sector with high potential for future growth. Investment in blockchain technology in the healthcare industry is expected to exceed \$500 million by 2022 (Hasselgren et al., 2020). In the current initial phase, the STO market is mainly composed of accredited or experienced investors (Lambert et al., 2021) who are particularly interested in supporting businesses with high potential for development and that combine environmentally and socially valuable activities. Similarly, within investment megatrends, health and technology have become two of the most compelling investment themes (Aruna et al., 2021). In this case, information about the technology adopted by health companies may be perceived by the market as a quality signal that reduces information asymmetry between parties. As suggested by Bhattacharya and Ritter (1983), disclosure of firms' technological progress impacts their financing activities if it provides a credible signal about their innovation processes.

Our second contribution refers to the long-standing issue of the environmental impact of these new technologies. On the one side, DLT and blockchain have offered considerable opportunities to deliver 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 on activities responsible for pollution and environmental degradation (Hou et al., 2020; Saberi et al., 2018). However, on the opposite side, energy-intensive design of many algorithms, processes, and computations within the blockchain poses challenges for electricity consumption and greenhouse gas emissions (Goodkind et al., 2020; Truby, 2018). Additionally, the energy consumption of DLT differs significantly between different design options, and so it is understandable that this topic could represent an important signal to communicate to the market in the origination of a blockchain-based IT solutions (Kannengießer et al. 2019; Sedlmeir et al., 2020), such as a STO. Environmental disclosure is commonly viewed as a signal to mitigate adverse regulatory or legislative pressure in the future (Brammer and Pavelin, 2008); yet, in the case of ICOs (Mansouri and Momtaz, 2021), and according to studies in crowdfunding domain (Butticè et al., 2019; Calic and Mosakowski, 2016; Vismara, 2019), signalling a project's environmental orientation in campaign information seems to be also beneficial in attracting environmentally concerned investors, and therefore promoting campaign success. Preliminary work by Mansouri and Momtaz (2021) finds evidence that environmental social, and governance (ESG) disclosure in utility token offerings has a positive relationship with the funding amount. Our work focuses on STOs and evidences the impact of the environmental component of disclosure, in particular the issue of energy use (the 'E' of ESG). We, therefore, surmise that adequate disclosure of energy and green topics is a credible signal to investors for inferring the quality of the company and a 'trigger' to attract investors' interest. Therefore, our findings strengthen findings of leading roles of technological and ideological motives for investors in token offerings (Fish et al., 2021), confirming these roles by the types of information investors value.

Our study finally contributes to the nascent literature about disclosure in digital contexts. Even if innovative entrepreneurs may face a trade-off between disclosing technological information and maintaining secrecy (Glaeser, 2018; Moser, 2011), our work shows that in STOs, disclosure of this information is strongly related to the campaign success. In decentralized finance, investors face a high level of information asymmetry, and they must support their investment decisions using public information. Sceptical entrepreneurs may also gain extra-financial benefits if they combine the disclosure of technology information with interactive space related to the campaign; for example, an interactive forum could enable investors to give valuable feedback and ideas to the project team or pose questions about company characteristics, which could increase the company's market awareness and the project's visibility, exposing the company to external ideas and open innovation concepts.

This study has implications for both innovative entrepreneurs and policymakers. For entrepreneurs, our results discourage publication of heavy white papers and show that it is preferable to provide light main documents with additional supporting summary document. In addition, addressing environmental issues represents a success factor regardless of the sector to which the issuer belongs to and we believe this topic will maintain its relevance in the future, given the present hype for climate and transition risks.

Regarding policymakers, we are aware that governments around the world have settled on different approaches to regulating the crypto industry. Many legislatures and regulators proactively drafted new laws, regulations, guidance, and frameworks for the crypto industry, while others use existing frameworks and enforcement actions to regulate the industry (see Lambert al al., 2021 for a review). For instance, the European Commission adopts a proactive approach on how to categorize crypto-assets and when tokens should be treated as financial transferable securities. This issue is crucial because if STOs are 'transferable securities' they fall under financial regulation; in Europe, this is MiFID II. This means that the requirements of prospectus regulations will apply, unless certain exemptions will be applicable (e.g., if the size of the offer is below a minimum threshold or if there is no public offering). In this case, STOs would be subject to traditional prospectus discipline and not to any emerging discipline for crypto assets (such as MiCA). Even in such case, given the technological component of STOs, it remains important for the regulator to leverage the provisions of specific crypto assets regulations to understand what specific information that is not traditionally included in a prospectus should be further considered. For example, MiCA suggests that the white paper should include, among other topics, information on the underlying technology used and the related risks. Our findings support this view. We argue that regulators should consider revising standard prospectus contents to include information on technological and energy issues. Because prospectus topics will become mandatory in STOs, supervisory authorities should provide guidance on the requirements for disclosure of technological and energy information and not only for financial information, as well as a supervision of the content published.

The current study's limitations should be noted and addressed in future research. First, our results might be affected by the relatively short sample period used. Future research will extend the observation period and consequently the sample's dimensions. Second, regulators around the world are currently developing specific disclosure requirements for STOs. Future research could explore how different countries' regulatory requirements could impact companies' disclosures and investors' participation. Third, while LDA is the standard technique for this kind of analysis, it detects topics only according to the distribution of terms in the documents and does not consider their meaning. Therefore, we plan to explore other kind of topic models that address these limitations, such as the structural topic model (Roberts et al 2019), which can incorporate additional metadata, and the embedded topic model (Dieng et al., 2020), which also takes advantage of word embeddings. We also plan to utilize entity linking techniques (Mendes et al., 2011) in order to detect specific entities such as organizations, persons, and technologies by exploiting large-scale knowledge graphs of relevant concepts, such as DBpedia (Auer et al., 2007) and Wikidata (Vrandečić and Krötzsch, 2014).

****This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.*

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

Arnold, T., Fishe, R.P.H., North, D., 2010. The Effects of Ambiguous Information on Initial and Subsequent IPO Returns. Financ. Manag. 39, 1497–1519. https://doi.org/10.1111/j.1755-053X.2010.01120.x

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

Aruna Sri, P.S.G., Lalitha Bhaskari, D., 2021. A blockchain framework for healthcare data management using consensus based selective mirror test approach, in: Intelligent Systems Reference Library. pp. 77–95. https://doi.org/10.1007/978-3-030-69395-4_5

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. In The semantic web (pp. 722-735). https://doi.org/10.1007/978-3-540-76298-0_52

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

Beatty, R.P., Ritter, J.R., 1986. Investment banking, reputation, and the underpricing of initial public offerings. J. financ. econ. 15. https://doi.org/10.1016/0304-405X(86)90055-3

Beinke, J.H., Rohde, K., Pohl, F., Teuteberg, F., 2021. Exploring the Success Factors of Security Token Offerings: An Empirical Approach. Int. J. Inf. Technol. Decis. Mak. 20. https://doi.org/10.1142/S0219622021500358

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.

Blandin, A., Cloots, A.S., Hussain, H., Rauchs, M., Saleuddin, R., Allen, J.G., Zhang, B.Z., Cloud, K., 2019. Global Cryptoasset Regulatory Landscape Study. SSRN Electron. J. https://doi.org/10.2139/ssrn.3379219

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

Block, J.H., Groh, A., Hornuf, L., Vanacker, T., Vismara, S., 2021. The entrepreneurial finance markets of the future: a comparison of crowdfunding and initial coin offerings. Small Bus. Econ. 57, 865–882. https://doi.org/10.1007/s11187-020-00330-2

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

Bouma, G., 2009. Normalized (pointwise) mutual information in collocation extraction. Proceedings of GSCL, 30, pp.31-40.

Bourveau, T., De George, E.T., Ellahie, A., Macciocchi, D., 2022. The Role of Disclosure and Information Intermediaries in an Unregulated Capital Market: Evidence from Initial Coin Offerings. J. Account. Res. 60, 129–167. https://doi.org/10.1111/1475-679X.12404

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

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

Chen, M.A., Wu, Q., Yang, B., 2019. How Valuable Is FinTech Innovation? Rev. Financ. Stud. https://doi.org/10.1093/rfs/hhy130

Chuanjie, F., Koh, A., Griffin, P., 2019. Automated Theme Search in ICO White papers. 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

Cormier, D., Magnan, M., 2007. The revisited contribution of environmental reporting to investors' valuation of a firm's earnings: An international perspective. Ecol. Econ. 62, 613–626. https://doi.org/10.1016/j.ecolecon.2006.07.030

Cormier, D., Ledoux, M.J., Magnan, M., 2011. The informational contribution of social and environmental disclosures for investors. Manag. Decis. 49, 1276–1304. https://doi.org/10.1108/00251741111163124

Diamond, D. W., Verrecchia, R. E., 1991. Disclosure, liquidity, and the cost of capital. The journal of Finance, 46(4), 1325-1359.

Dieng, A.B., Ruiz, F.J.R., Blei, D.M., 2020. Topic modeling in embedding spaces. Trans. Assoc. Comput. Linguist. 8, 439–453. https://doi.org/10.1162/tacl_a_00325

European Commission (2020), Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions An SME Strategy for a sustainable and digital Europe. Available at: https://ec.europa.eu/info/sites/default/files/communication-sme-strategy-march-2020_en.pdf

European Commission (2020), Proposal for a regulation of the European parliament and of the council on Markets in Crypto-assets, and amending Directive (EU) 2019/1937. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0593

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., 2022. Experts or charlatans? ICO analysts and white paper informativeness, J. Bank. Financ. 139. https://doi.org/10.1016/j.jbankfin.2022.106476.

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

Goodkind, A.L., Jones, B.A., Berrens, R.P., 2020. Cryptodamages: Monetary value estimates of the air pollution and human health impacts of cryptocurrency mining. Energy Res. Soc. Sci. 59. https://doi.org/10.1016/j.erss.2019.101281

Gu, F., Li, J.Q., 2007. The credibility of voluntary disclosure and insider stock transactions. J. Account. Res. 45, 771–810. https://doi.org/10.1111/j.1475-679X.2007.00250.x

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

Hanley, K.W., Hoberg, G., 2010. The information content of IPO prospectuses. Rev. Financ. Stud. 23, 2821–2864. https://doi.org/10.1093/rfs/hhq024

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

Hasselgren, A., Kralevska, K., Gligoroski, D., Pedersen, S.A., Faxvaag, A., 2020. Blockchain in healthcare and health sciences—A scoping review. Int. J. Med. Inform. https://doi.org/10.1016/j.ijmedinf.2019.104040

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

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

Klöckner, M., Schmidt, C.G., Wagner, S.M., 2022. When Blockchain Creates Shareholder Value: Empirical Evidence from International Firm Announcements. Prod. Oper. Manag. 31, 46–64. https://doi.org/10.1111/poms.13609

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)

Li, M., Wang, W., Zhou, K., 2021. Exploring the technology emergence related to artificial intelligence: A perspective of coupling analyses. Technol. Forecast. Soc. Change 172. https://doi.org/10.1016/j.techfore.2021.121064

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

Liu, W., Wang, J., Jia, F., Choi, T. M., 2022. Blockchain announcements and stock value: a technology management perspective. Int. J. Oper. Prod. Manage. 1–12. https://doi.org/10.1108/IJOPM-08-2021-0534.

Loughran, T., McDonald, B., 2013. IPO first-day returns, offer price revisions, volatility, and form S-1 language. J. financ. econ. 109, 307–326. https://doi.org/10.1016/j.jfineco.2013.02.017

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)

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

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

Mendes, P.N., Jakob, M., García-Silva, A. and Bizer, C., 2011. DBpedia spotlight: shedding light on the web of documents. In Proceedings of the 7th international conference on semantic systems (pp. 1-8). https://doi.org/10.1145/2063518.20635198).

Meoli, M., Vismara, S., 2022. Machine-learning forecasting of successful ICOs. J. Econ. Bus. 106071. https://doi.org/10.1016/j.jeconbus.2022.106071

Miglo, A., 2021. STO vs. ICO: A Theory of Token Issues under Moral Hazard and Demand Uncertainty. J. Risk Financ. Manag. 14. https://doi.org/10.3390/jrfm14060232

Miller, B.P., 2010. The effects of reporting complexity on small and large investor trading. Account. Rev. https://doi.org/10.2308/accr.00000001

Momtaz, P.P., 2021. Security Tokens. SSRN Electron. J. https://doi.org/10.2139/ssrn.3865233

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

Reber, B., Gold, A., Gold, S., 2021. ESG Disclosure and Idiosyncratic Risk in Initial Public Offerings. J. Bus. Ethics. https://doi.org/10.1007/s10551-021-04847-8

Roberts, M. E., Stewart, B. M., Tingley, D., 2019. STM: An R Package for Structural Topic Models. Journal of Statistical Software, 91(2), 1–40. https://doi.org/10.18637/jss.v091.i02

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

Röder, M., Both, A., Hinneburg, A., 2015. Exploring the space of topic coherence measures. In Proceedings of the eighth ACM international conference on Web search and data mining (pp. 399-408). https://doi.org/10.1145/2684822.2685324

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

Saidi F., Žaldokas A. 2021. How Does Firms' Innovation Disclosure Affect Their Banking Relationships?. Management Science 67(2):742-768. https://doi.org/10.1287/mnsc.2019.3498

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

Savin I., Chukavina K., Pushkarev A. 2022a. Topic-based classification and identification of global trends for startup companies, Small Business Economics, forthcoming

Savin, I., Drews, S., Maestre-Andrés, S., van den Bergh, J., 2020. Public views on carbon taxation and its fairness: a computational-linguistics analysis. Clim. Change 162, 2107–2138. https://doi.org/10.1007/s10584-020-02842-y

Savin, I., Drews, S., van den Bergh, J., 2021a. Free associations of citizens and scientists with economic and green growth: A computational-linguistics analysis. Ecol. Econ. 180. https://doi.org/10.1016/j.ecolecon.2020.106878

Savin, I., Ott, I., Konop, C., 2022b. Tracing the evolution of service robotics: Insights from a topic modeling approach. Technol. Forecast. Soc. Change 174. https://doi.org/10.1016/j.techfore.2021.121280

Savin, I., van den Bergh, J., 2021. Main topics in EIST during its first decade: A computational-linguistic analysis. Environ. Innov. Soc. Transitions. https://doi.org/10.1016/j.eist.2021.06.006

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

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

Spence, M., 1973. Job market signaling. Q. J. Econ. 87, 355–374. https://doi.org/10.2307/1882010

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

Vrandečić, D. and Krötzsch, M., 2014. Wikidata: a free collab. orative knowledgebase. Communications of the ACM, 57(10), pp.78-85. https://doi.org/10.1145/2629489

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 white paper 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

Xu, W., Wang, T., Chen, R., Zhao, J.L., 2021. Prediction of initial coin offering success based on team knowledge and expert evaluation. Decis. Support Syst. 147. https://doi.org/10.1016/j.dss.2021.113574

Appendix

Tab. A.1
Variable definitions

Variable	Description
Depended variables	
Success	Dummy = 1 if the STO reaches the minimum target amount or if it presents the total amount raised in the campaign
Amount Raised	Logarithm of the total amount raised in the campaign
Readability variables	
Words	Logarithm of the number of words contained in each document
Size	Logarithm of the white paper's weight in KB
Fog rate	Readability rate that considers the number of complex words in the text
Visual elements variables	
Pictures	Number of pictures in the white paper
Tables	Number of tables in the white paper
Extra doc.	Dummy = 1 if there is a short version of the white paper published apart with the key information (i.e product presentation, financial forecast)
Topics	
T1	Dummy = 1 if the percentage of words in the document related to energy and green issue is greater than 5%
T2	Dummy = 1 if the percentage of words in the document related to company description is greater than 5%
T3	Dummy = 1 if the percentage of words in the document related to AI and machine learning is greater than 5%
T4	Dummy = 1 if the percentage of words in the document related to tech for art and education is greater than 5%
T5	Dummy = 1 if the percentage of words in the document related to DLT components is greater than 5%
T6	Dummy = 1 if the percentage of words in the document related to tech for construction and manufacturing is greater than 5%
T7	Dummy = 1 if the percentage of words in the document related to tech for healthcare is greater than 5%
T8	Dummy = 1 if the percentage of words in the document related to financial and legal issue is greater than 5%
T9	Dummy = 1 if the percentage of words in the document related to tech applied in finance and other service is greater than 5%
Control variables	
Softcap use	Dummy equals 1 if the campaign has set a soft cap
Offering days	Number of days of the offering
Pre-sale offering	Dummy equals 1 if the offering provides for a pre-sale
Extra_Europe (d)	Dummy equals 1 if the issue refers to extra-European countries
Europe (d)	Dummy equals 1 if the issue refers to European countries
Agriculture, forestry and fishing (d)	Dummy equals 1 if the company's sector is agriculture, forestry and fishing

Finance, insurance, real estate (d)	Dummy equals 1 if the company's sector is finance, insurance and real estate
Manufacturing (d)	Dummy equals 1 if the company's sector is manufacturing
Mining and Construction (d)	Dummy equals 1 if the company's sector is mining and construction
Services (d)	Dummy equals 1 if the company's sector is services
Media and Communication (d)	Dummy equals 1 if the company's sector is media and communication
Retail and Wholesale trade (d)	Dummy equals 1 if the company's sector is retail and wholesale trade
Year 2017 (d)	Dummy equals 1 if the offering started in 2017
Year 2018 (d)	Dummy equals 1 if the offering started in 2018
Year 2019 (d)	Dummy equals 1 if the offering started in 2019
Year 2020 (d)	Dummy equals 1 if the offering started in 2020
Year 2021 (d)	Dummy equals 1 if the offering started in 2021

Tab. A.2

Results with Fog Rate

VARIABLES	Model 12 T5% Success	Model 13 T5% Margins
Words	0.341*	0.0735*
	(0.197)	(0.0431)
Size	-0.212*	-0.0457*
	(0.123)	(0.0270)
Pictures	0.0354	0.00764
	(0.0219)	(0.00481)
Tables	-0.0479	-0.0103
	(0.0488)	(0.0107)
Extra Doc.	0.675**	0.118**
	(0.334)	(0.0485)
T1 Energy and green issues	0.685**	0.127***
	(0.292)	(0.0455)
T2 Company description	0.223	0.0442
	(0.356)	(0.0652)
T3 AI and machine learning	0.358	0.0664
	(0.370)	(0.0597)
T4 Tech for art and education	-0.341	-0.0845
	(0.433)	(0.121)
T5 DLT components	0.356	0.0672
	(0.365)	(0.0603)
T6 Tech for construction and manufacturing	-0.384	-0.0947
	(0.322)	(0.0880)
T7 Tech for Healthcare	1.205***	0.163***
	(0.427)	(0.0339)
T8 Financial and legal issues	0.152	0.0315
	(0.295)	(0.0590)
T9 Tech for finance and other services	0.0531	0.0115
	(0.290)	(0.0634)
Fog Rate	-0.0142	-0.00305
	(0.0417)	(0.00898)
Soft cap use	0.859***	0.208***
	(0.296)	(0.0785)
Offering days	-0.00256	-0.000552
	(0.00192)	(0.000423)
Pre-sale offering	0.153	0.0323
	(0.268)	(0.0556)
Constant	8.687***	
	(2.188)	
Country FE	Yes	
Sector FE	Yes	
Year FE	Yes	
Pseudo R2	0.31	
ROC AUC	0.82	
Prob > chi2	0.7	
Obs.	186	

Table A.2 shows the results of the probit regression models adding Fog rate as control variable 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 'Success' columns report the coefficient of the probit regression; the 'Margins' columns report the marginal effects of the probit regression.