



A TOPIC MODELLING ANALYSIS OF WHITE PAPERS IN SECURITY TOKEN OFFERINGS:

Which topic matters for fundings?

Paola Bongini - University of Milano-Bicocca

Francesco Osborne - The Open University

Alessia Pedrazzoli - University of Modena and Reggio Emilia

Monica Rossolini - University of Milano-Bicocca

A QUICK OVERVIEW

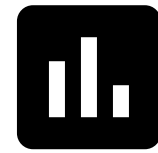
- **Goal:** To investigate the relevance of the various topics and contents in white papers and their impact on campaign performance.

RQs: Which topics characterized STOs' white paper? which topics were able to impact campaign success? what is the role of technology and environmental disclosure?

- **Data:** 188 STOs from 2017 to 2021
- **Method:** latent Dirichlet allocation (**LDA**) topic modelling to identify the topics and themes in white papers and **multivariate analysis** to examine determinants of funding success
- **Results:** Nine different topics emerge in the white papers of our sample, the amount of information dedicated to **energy use and green aspects** is one of the highest in the white papers. Two topics influence the probability of campaign success and the total amount raised: i. the disclosure about the **technology applied in the healthcare** sector and, ii. the **energy use and green aspects**.

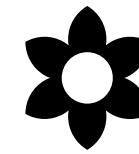
STUDY MOTIVATIONS

STOs growth



- The higher investor protection in STOs compared to ICOs has increased the expectation for the continuous growth of this market
- few studies focused on STOs (Ante and Fiedler, 2019; Myalo, 2019; Beinke et al., 2021 Lambert et al., 2021)

White paper relevance in campaign success



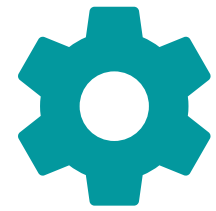
- informative means to attract prospective investors' interest, communicate the issuers' quality, and support the liquidity of the market
- **positive impact on the campaign performance in ICOs**
(Boreiko and Sahdev, 2018; Cerchiello et al., 2019; Giudici et al., 2018)

White paper disclosure



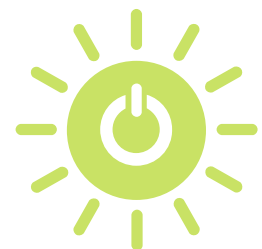
- **STRUCTURE** (readability, length, Amsden and Schweizer, 2019; Zhang et al., 2019)
- **LANGUAGE STYLE** (Momtaz, 2020)
- **QUALITY** (number of “technical” words, the presence of specific section, Chen, 2019; Feng et al., 2019, and Fish, 2019)

- the technological components of the project and their application in the business proposed are crucial factors in the evaluation of the company and to reduce information asymmetry between investors and entrepreneurs (Fish, 2019; Fish et al., 2021).



TECHNOLOGY DISCLOSURE

- The energy consumption of technology differs significantly between different design choices, and subsequently, it could represent an important dimension to evaluate and to communicate to the market during the origination of a blockchain-based IT solution (Kannengießer et al. 2019; Sedlmeir et al., 2020) The environmental disclosure has commonly been viewed as a pre-emptive step to mitigate adverse regulatory or legislative pressures in the future (Brammer and Pavelin, 2008).



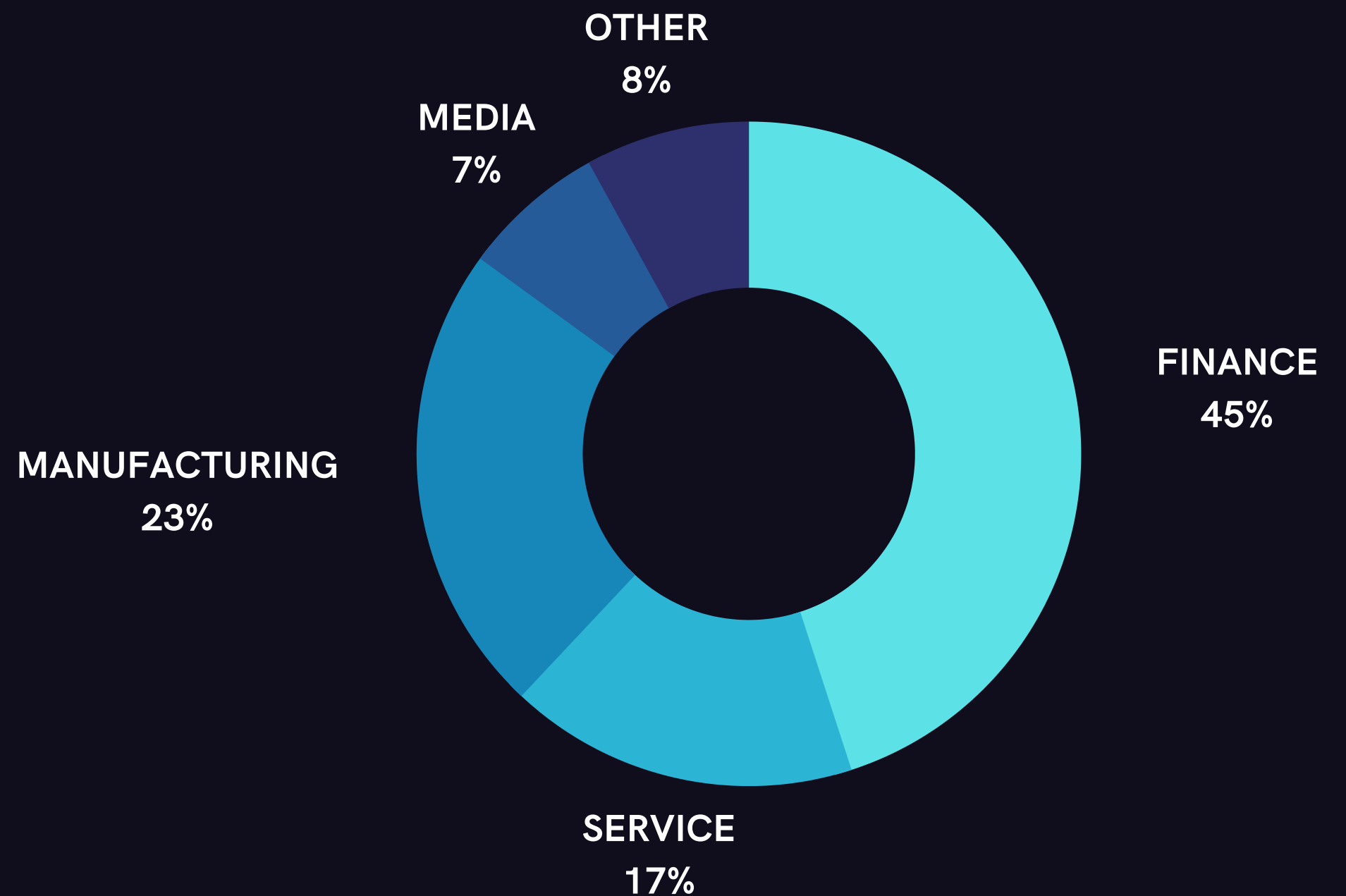
ENERGY AND ENVIRONMENTAL DISCLOSURE

SAMPLE

STOs aggregators:

Coinintelligence.com, Tokenmarket.net, STOsphere.com, STOrating.com, STOWise.com, Tokenmarket.net, Blockdata, and ICObench.com.

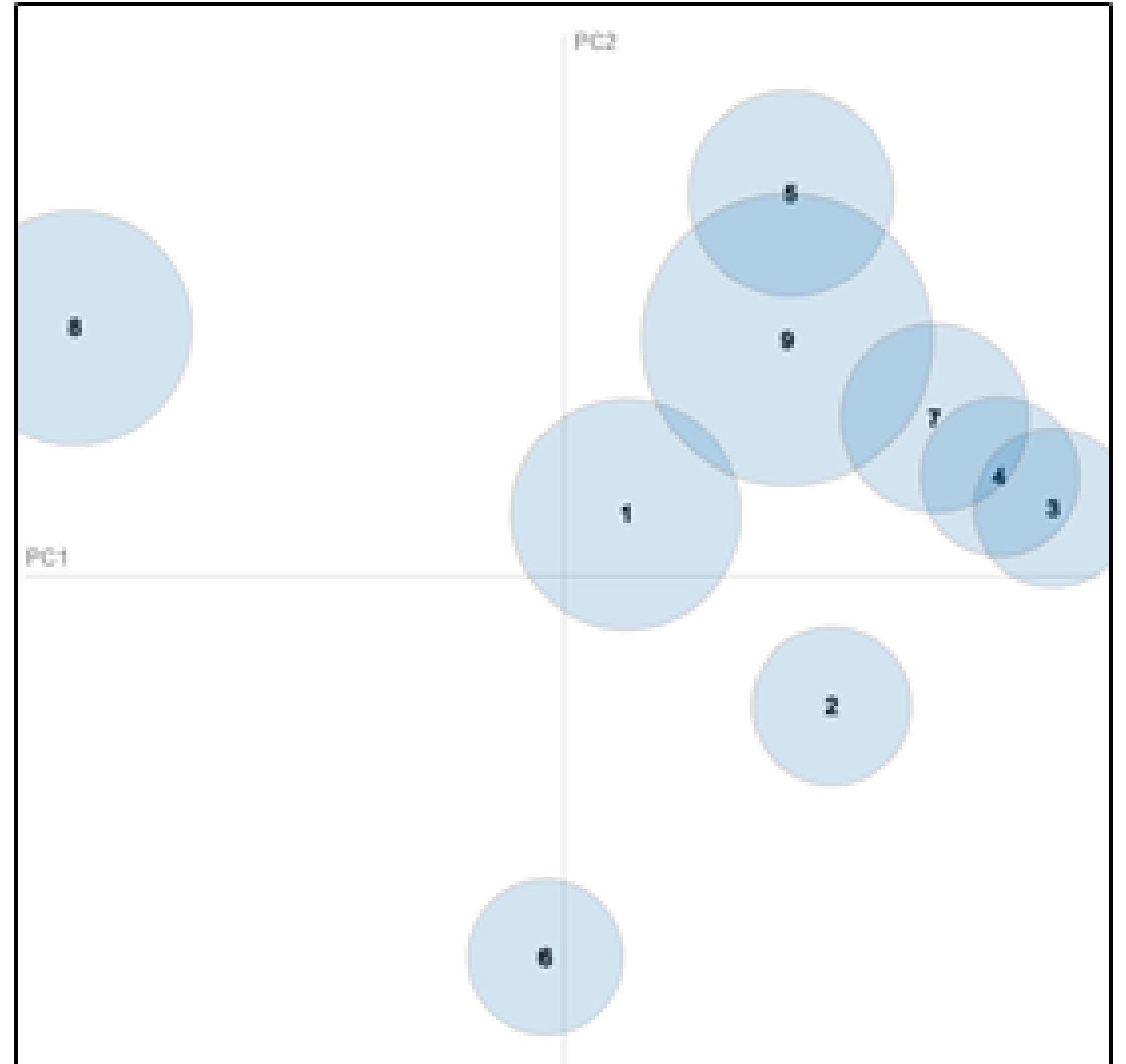
- **188 STOs, 2017-2021**
53% extra European countries



WHICH TOPICS? LDA METHOD

- LDA does not impose on the researcher dictionaries or interpretative restrictions.
- Words frequently appearing together tend to be semantically related (Stroropoli, 2019).
- To identify the number of topics we refer to topic coherence a standard metric that measures the degree of semantic similarity between high scoring terms in each topic.

9 TOPICS

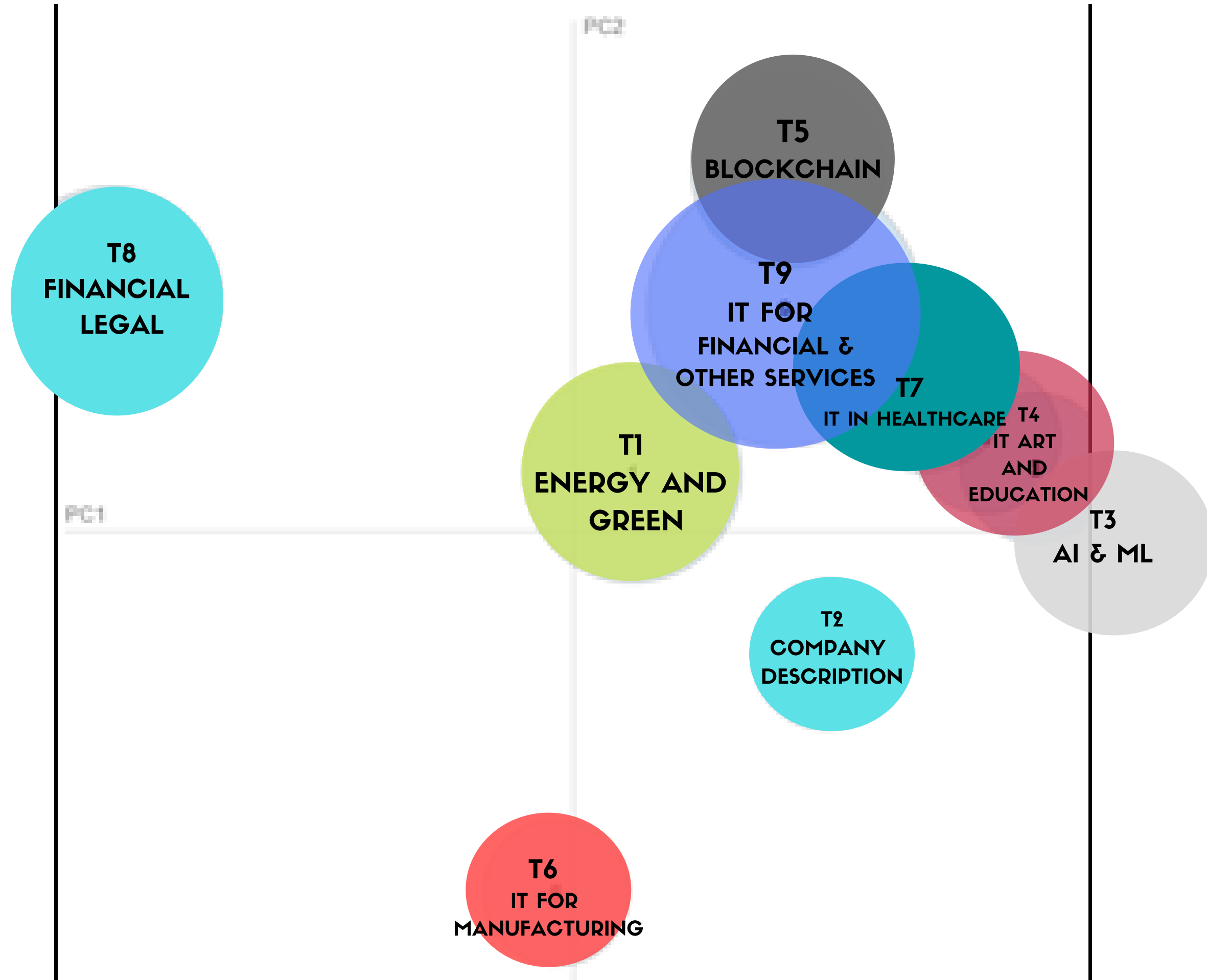


- LABEL IDENTIFICATION: We refer to Chuanjie et al. (2019) and Zhang et al.,(2021) to support the generation of interpretable label.

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

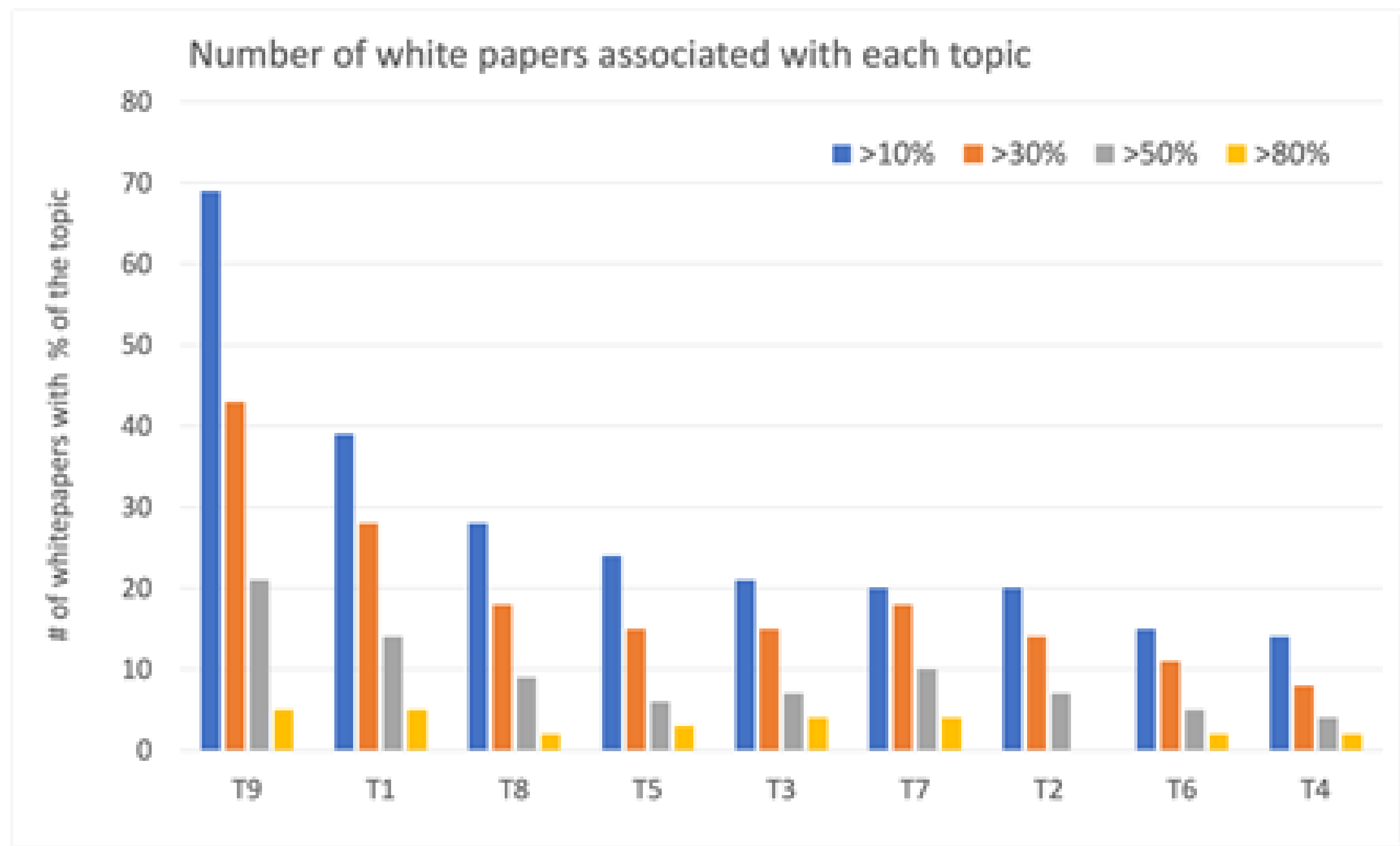
WHICH TOPICS? LDA METHOD

A VISUALIZATION OF THE 9 TOPICS FROM STO WHITE PAPERS





WHICH TOPICS? LDA METHOD



Where: T9 IT Financial & other services; T1 Energy and green Issues; T8 Financial and legal issues; T5 DLT components; T3 AI and machine learning; T7 IT for healthcare; T2 Company description; T6 IT for construction and manufacturing; T4 IT for art and education

VARIABLES



DEPENDENT

Success:

- dummy variable equals 1 if the total fundraised is greater than the minimum target or the STO has raised any capital in the case no minimum target requirement is specified, and 0 otherwise
- total amount raised (ln)



INDIPENDENT VARIABLES

Topic_1 to Topic_9: the percentage of words in the document related to a specific topic

Topic_1 to Topic_9 (d): when the topic is present at least 30% in the text



CONTROL

Length: number of Words
Size: size of the document in KB

Extra Document: number of files attached to the white paper

Pictures number

Tables number

Offering days number

Pre-sale(dummy)

Soft cap use (dummy)

Time, industry, Geo FE

MODELS

PROBIT MODEL

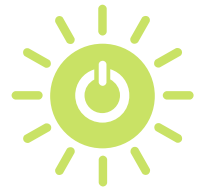
$$\begin{aligned} \text{Prob}(\text{Success}_i = 1) &= F(\alpha + \beta_1(\text{Readability})_i + \beta_2(\text{Structure})_i + \beta_3(\text{Topics})_i + \beta_4(\text{Industry}) \\ &+ \beta_5(\text{Geo}_{\text{area}}) + \beta_6(\text{Year}) + \varepsilon_{i,k,t}) \end{aligned}$$

TOBIT MODEL

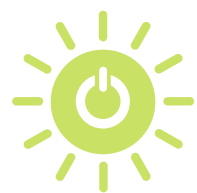
$$\begin{aligned} \text{Amount Raised} &= (\alpha + \beta_1(\text{Readability})_i + \beta_2(\text{Structure})_i + \beta_3(\text{Topics})_i + \beta_4(\text{Industry}) \\ &+ \beta_5(\text{Geo}_{\text{area}}) + \beta_6(\text{Year}) + \varepsilon_{i,k,t}) \end{aligned}$$



Models and results main analysis



VARIABLES	Model 1	Model 1	Model 2	Model 2	Model 3	Model 3
	Success	Margins	Success	Margins	Success (T>30%)	Margins
Words	0.176 (0.161)	0.046 (0.042)	0.22 (0.179)	0.053 (0.042)	0.300* (0.178)	0.075* (0.043)
Size	-0.270** (0.125)	-0.071** (0.031)	-0.214* (0.114)	-0.051** (0.026)	-0.226* (0.126)	-0.056* (0.03)
Pictures	0.0295 (0.0251)	0.007 (0.006)	0.029 (0.0242)	0.007 (0.005)	0.0222 (0.0254)	0.005 (0.006)
Tables	-0.0704* (0.041)	-0.018* (0.01)	-0.0583 (0.0409)	-0.014 (0.009)	-0.0619 (0.0436)	-0.015 (0.01)
Extra Doc	0.797** (0.31)	0.211** (0.081)	0.841*** (0.312)	0.203*** (0.074)	0.884*** (0.322)	0.222*** (0.079)
T1 Energy and green issue			0.417 (0.683)	0.1 (0.163)	0.576* (0.345)	0.144* (0.084)
T3 AI and machine learning			-0.201 (0.72)	-0.048 (0.173)	0.165 (0.424)	0.041 (0.106)
T4 IT for art and education			-0.495 (0.789)	-0.119 (0.189)	-0.179 (0.458)	-0.044 (0.114)
T5 DLT technology			0.396 (0.707)	0.095 (0.17)	0.354 (0.386)	0.089 (0.097)
T6 IT for construction and manufacturing			-0.446 (0.726)	-0.107 (0.174)	-0.325 (0.461)	-0.081 (0.114)
T7 IT for Healthcare			1.520* (0.825)	0.367* (0.202)	1.289** (0.521)	0.323** (0.134)
T8 Financial and legal issue			0.0653 (0.747)	0.015 (0.18)	0.0882 (0.354)	0.022 (0.088)
T9 IT for financial & other services			-0.0945 (0.639)	-0.022 (0.154)	-0.0601 (-0.291)	-0.015 (0.073)
Constant	11.04*** (1.730)		9.454*** (1.892)		7.914*** (-1.965)	
Sector FE	Yes		Yes		Yes	
Year FE	Yes		Yes		Yes	
Country FE	Yes		Yes		Yes	
Observations	188		188		188	
Pseudo R2	0.156		0.202		0.200	



VARIABLES	(4)	(4)	(5)	(5)	(5)	(6)
	Succes	Margins	Succes	Margins	Amount In	Amount In (T>30%)
Word	0.189 (0.196)	0.040 (0.042)	0.189 (0.202)	0.040 (0.043)	1.680 (1.083)	1.866* (1.067)
Size	-0.256** (0.120)	-0.055** (0.025)	-0.234* (0.125)	-0.050* (0.027)	-0.825 (0.691)	-0.853 (0.685)
Pictures	0.042* (-0.023)	0.009* (0.005)	0.0364 (0.023)	0.0078 (0.005)	0.138 (0.115)	0.119 (0.117)
Tables	-0.035 (-0.046)	-0.007 (0.009)	-0.0541 (0.0486)	-0.012 (0.011)	-0.246 (0.300)	-0.223 (0.311)
Extra Doc	0.536 (0.343)	0.116 (0.074)	0.563 (0.354)	0.101* (0.0531)	3.184** (1.279)	3.394** (1.312)
T1 Energy and green issue	0.539 (0.742)	0.116 (0.159)	0.230 (0.760)	0.049 (0.163)	3.338 (3.366)	3.118* (1.682)
T3 AI and machine learning	0.001 (0.780)	0.000 (0.168)	-0.266 (0.819)	-0.057 (0.174)	-0.768 (4.809)	0.418 (2.602)
T4 IT for art and education	-0.982 (0.802)	-0.212 (0.169)	-1.340 (0.847)	-0.287 (0.178)	-1.679 (4.645)	-0.402 (2.546)
T5 DLT technology	0.369 (0.761)	0.079 (0.165)	0.260 (0.774)	0.055 (0.166)	1.949 (3.752)	1.440 (2.213)
T6 IT for construction and manufacturing	-0.667 (0.775)	-0.144 (0.164)	-0.993 (0.780)	-0.212 (0.163)	-5.386 (4.717)	-1.120 (2.954)
T7 IT for Healthcare	1.837** (0.932)	0.397** (0.204)	1.545* (0.931)	0.330* (0.198)	7.607** (3.492)	5.752** (2.242)
T8 Financial and legal issue	-0.253 (0.805)	-0.054 (0.173)	-0.344 (0.815)	-0.074 (0.173)	1.259 (4.253)	0.352 (2.180)
T9 IT financial and other services	0.0572 (0.706)	0.012 (0.152)	-0.229 (0.717)	-0.049 (0.153)	0.556 (3.720)	0.124 (1.714)
softcapuse	1.033*** (0.290)	0.223*** (0.055)	0.944*** (0.289)	0.229*** (0.076)		
Presale dummy			0.273 (0.256)	0.056 (0.051)		
Offering days			-0.001 (0.001)	-0.000 (0.000)		
Constant	10.24*** (2.268)		10.28*** (2.251)		12.34 (10.77)	10.42 (11.20)
Country FE	Yes		Yes		Yes	Yes
Sector FE	Yes		Yes		Yes	Yes
Year FE	Yes		Yes		Yes	Yes
Observations	187		186		188	188
Pseudo R2	0.278		0.286		0.0367	0.0347

Key points and conclusion

- **Energy and green issues** is one of the most represented topic in white papers
- **The technological components of the healthcare sector and disclosure about energy and green issues are hot topics for campaign success.**
- **Readability and Structure:** heavy documents negatively affect the success, whereas light main documents with extra files seem to be more effective

- **IN STO MARKET MEGATRENDS (AGEING , HEALTH AND CLIMATE CHANGE) SEEM TO BE RELEVANT**
- **THE LEADING ROLE OF TECHNOLOGICAL AND IDEOLOGICAL MOTIVES FOR INVESTORS IN TOKEN OFFERINGS (FISH ET AL., 2021) ARE CONFIRMED BY THE TYPE OF INFORMATION THEY LOOK AT.**
- **AN INSIGHT INTO DISCLOSURE IN THE DIGITAL CONTEXT: REVISION OF STANDARD PROSPECTUS' CONTENT, GIVING ADEQUATE SPACE TO THE TECHNOLOGICAL AND TO THE ENVIRONMENTAL PARTS.**

Limits and future developments

- **relatively short sample period: future research could extend the observation period and consequently the sample dimension.**
- **regulators are developing specific disclosure requirements for STOs: how does regulation change disclosure and white paper structure in STOs?**

Thank you!



Paola Bongini
paola.bongini@unimib.it

Francesco Osborne
francesco.osborne@open.ac.uk

Alessia Pedrazzoli
alessia.pedrazzoli@unimore.it

Monica Rossolini
monica.rossolini@unimib.it