ISSN: 2281-1346

Department of Economics and Management

DEM Working Paper Series

ICOs success drivers: a textual and statistical analysis

Paola Cerchiello
(Università di Pavia)

Anca Mirela Toma
(Università di Pavia)

 # 164 (07-18)

Via San Felice, 5
I-27100 Pavia
economiaweb.unipv.it

July 2018

ICOs success drivers: a textual and statistical analysis

Paola Cerchiello, Anca Mirela Toma

University of Pavia, Italy

Draft Version

Abstract

Initial coin offering (aka ICOs) represents one of the several by-product of the cryptocurrencies world. New generation start-up and existing businesses in order to avoid rigid and long money raising protocols imposed by classical channels like banks or venture capitalists, offer the inner value of their business by selling tokens, i.e. units of the chosen cryptocurrency, like a regular firm would do with and IPO. The investors of course hope in a value increasing of the tokens in the near future, provided a solid and valid business idea typically described by the ICO issuers in a white paper, both a descriptive and technical report of the proposed business. However, fraudulent activities perpetrated by unscrupulous start-up happen quite often and it would be crucial to highlight in advance clear signs of illegal money raising. In this paper, we employ a statistical approach to detect which characteristics of an ICO are significantly related to fraudulent behaviours. We leverage a number of different variables like: entrepreneurial skills, number of people chatting on Telegram on the given ICO and relative sentiment, type of business, country issuing, token pre-sale price. Through logistic regression, classification tree we are able to shed a light on the riskiest ICOs.

Keywords: ICOs, cryptocurrencies, fundraising, classification models, text analysis

1. Introduction

Initial coin offerings (aka ICOs) are becoming more and more popular and represent an alternative strategy to raise money thanks to a new technology known as blockchain. New generation start-up and agile existing businesses, in order to avoid rigid and long money raising protocols imposed by classical channels like banks or venture capitalists, can offer the inner value of their business by selling tokens, i.e. units of a chosen cryptocurrency. When we say cryptocurrency, we refer to a digital currency, a new mean for exchange, which most popular examples are Bitocoin and Ethereum. Blockchain (chain of blocks) is the core technology at the basis of a cryptocurrency; it is a Distributed Ledger Technology defined as distributed, shared, encrypted database that serves as an irreversible and incorruptible repository of information (Wright, De Filippi, 2015). The number of cryptocurrencies available worldwide is close to 1600 and constantly growing and if we consider market capitalization, Bitcoin is currently the largest blockchain network, followed by Ethereum, Ripple, Bitcoin Cash, Litecoin, and Stellar (Coinmarketcap.com, June 15, 2018). The success of such decentralized technology lays on the fact that it works without the commitment and the control of a central authority: the blockchain is a Peer-to-Peer technology. A Peer-to-Peer (P2P) is a way of structuring distributed applications such that the individual nodes [] can act as both a client and a server. [] A key concept for P2P systems is therefore to permit any two peers to communicate with one another in such a way that either ought to be able to initiate the contact (Peer-to-Peer Research Group, 2013). The more a P2P network is distributed, scalable, autonomous and secure, the more the P2P network is valuable.

All these precious features are allowing the fast growing of cryptocurrencies not just per se but as a tool for crow-funding purposes, giving birth to the so called Initial Coin Offerings. Moreover, what is further fostering the development of ICOs is the absence of regulation (even if many countries are currently working on it) and, at the moment, there are just few examples of ban acts (namely China, India, South Korea). Investors buy ICO tokens hoping in very high returns, sometimes even before the business is put in place, since the corresponding cryptocurrencies (typically Ethereum) can be immediately traded.

In the first 6 months of 2018, there have been 440 ICOs, with a peak in May (125) raising more than 10 billion US, where Telegram ICO (Pre-sale 1 & 2) is by far the most reworded one with 1.7 billion US (Coinschedule.com, 18 June, 2018). In 2017, the total amount raised by 210 ICOs was about 4 billion US, and overcame venture capital funnelled toward high tech initiatives in the same period The first token sale (also known as an ICO) was held by Mastercoin in July 2013 but one of the most successful and still operative is Ethereum which raised 3,700 BTC in its first 12 hours in 2014, equal to approximately 2.3 million at the time.

Despite the interest arose by ICOs and the constantly growing trends, it is worth mentioning that almost half of ICOs sold in 2017 failed by February 2018 (Hankin, 2018). In fact, what should drive more attention on ICOs is the consistent presence of scam activities only devoted to raise money in a fraudulent way. According to Cointelegraph the Ethereum network (the prevalent blockchain platform for ICOs) has experienced considerable phishing, Ponzi schemes, and other scams events, accounting for about 10% of ICOs (Ethereumscamdb.info, 2017). On the other hand, it is interesting to assess which factors affect the probability of success of an ICO. Adhami et al. in 2018, based on the analysis of 253 ICOs, showed that the following characteristics contribute: the availability of the code source, the organization of a token presale and the possibility for contributors to access a specific service (or to share profits).

Despite the boom of the ICOs world and the raise of interest from the general audience, only few scientific studies have been conducted and published. Besides the aforementioned Adhami et al 2018, we should mention the working paper by Zetzsche et al., that is focused on legal and financial risk aspects of ICOs, but its second section contains a taxonomy, and some data about ICOs that the authors claim are continuously updated. Recently, Subramanian in 2018 quoted the ICOs as an example of decentralized blockchain-based electronic marketplace. The main source of information about blockchains, tokens and ICOs is obviously the Web. Here we can find sites enabling to explore the various blockchains associated to the main cryptocurrencies, including Ethereum's one. We can also find Web sites giving extensive financial information on prices of all the main cryptocurrencies and tokens, and sites specialized in listing the existing ICOs and giving information about them. Often, these sites also evaluate the soundness and likeliness of success of the listed ICOs. One of the most popular among these sites is icobench.com, which evalutes all the listed ICOs, and provides an API to automatically gather information on them.

ICO are usually characterized by the following features: a business idea, typically explained in a white paper, a proposer team, a target sum to be collected, a given number of tokens, that is a new cryptocurrency, to be given to subscribers according to a predetermined exchange rate with one or more existing cryptocurrencies. Nowadays, a high percentage of ICOS is managed through Smart Contracts running on Ethereum blockchain, and in particular through ERC-20 Token Standard Contract. Cloning an ERC-20 contract, it is very easy to create a new token, issue a given number of tokens, and trade these tokens with Ethers, the Ethereum cryptocurrency, which has a monetary value according to a given exchange rate.

On top of all the characteristics explained so far, there is a further and not yet explored point of interest: the Telegram chats. Telegram is a cloud-based instant messaging and voice over IP service developed by Telegram Messenger founded by the Russian entrepreneur Pavel Durov. In March 2018, Telegram stated that it had 200 million monthly active users -'This is an insane number by any standards. If Telegram were a country, it would have been the sixth largest country in the world'(Telegram, 2018). Telegram is completely free and has no ads, users can send any kind of media or documents, and can program messages to self-destruct after a certain period of time. Some characteristics are imposing Telegram among the first social network, in fact it intentionally does not collect data about where its clients live and what they use the platform for. This is one of the main reason why, according to AppAnnie rankings, Telegram is particularly popular in countries like Uzbekistan, Ukraine, and Russia, where Internet access may be limited or closely monitored by the government. As of October 2017, Telegram was by far the most popular official discussion platform for current and upcoming ICOs, with 75%+ of these projects utilizing it. These means that retrieving telegram discussions associated to each ICOs, would produce a huge amount of textual information potentially useful for understanding the chance of success and more interestingly possible signs of scam activities.

In this paper we propose a combined approach based on classical classification models like logistic regression and random forest to highlight significant variables in distinguishing success from scam, and on text analysis. Specifically, we shall elicit from text whether some words/topic and/or a specific sentiment is expressed differently in successful/failure/scam ICOs. We have monitored and collected data (still collecting at the present date) for 121 ICOs from the beginning of 2018.

The paper is organized as follows: in section 2 we present the statistical methodology, in section 3 we describe collected data, in section 4 we illustrate results and in section 5 we report our final comments.

2. Methodology

In this paper we leverage two kinds of information: structured and unstructured ones. Regarding the former, more widely described in Section 3, we take advantage of classical statistical classification models to distinguish successful, failure and scam ICOs. Logistic regression aims to classify the dependent variable in two groups, characterized by a different status [1=scam vs 0=success or 1=success vs 0=failure] in which ICOs are classified by logistic regression, specified by the following model:

$$ln(\frac{p_i}{1-p_i}) = \alpha + \sum_j \beta_j x_{ij}, \tag{1}$$

where p_i is the probability of the event of interest, for ICO i, $x_i = (x_{i1}, \ldots, x_{ij}, \ldots, x_{iJ})$ is a vector of ICOs-specific explanatory variables, and the intercept parameter α, as well as the regression coefficients β_j , for $j = 1, ..., J$, are to be estimated from the available data. It follows that the probability of success (or scam) can be obtained as:

$$p_i = \frac{1}{1 + exp(\alpha + \sum_j \beta_j x_{ij})}, \tag{2}$$

However, in case of highly unbalanced distribution of the target variable, which can typically occur for scam ICOs, logistic regression could be replaced by other generalised linear models, such as the generalized extreme regression scoring model proposed by Calabrese and Giudici (2015) and Calabrese et al. (2016). This means that scam features are better represented by the tail of the response curve for values close to one, which can be modelled using a generalized extreme values (GEV) random variable (Kotz and Nadarajah, 2000; Falk et al., 2010). Because our focus is on scam ICOs, we exploit the quantile function of a GEV random variable and specify the link function

$$\frac{[ln(p_i)]^{-\tau} - 1}{\tau} \tag{3}$$

where $\tau \in \Re$ is the tail parameter. Hence, in (1) we replace $ln(\frac{p_i}{1-p_i})$ with (3). Since a GEV link can be asymmetric, underestimation of the default probability may be overcome.

Because of the still limited size of the sample employed in this paper (120 observations) and the relative large number of variables to be evaluated, we take advantage of a class of non parametric models to understand which are the most significant ones. Specifically, we employ Random forests (RF), which are again classification models with the advantage of no assumptions about the distribution of the data. RF are a powerful approach to improve decision trees, a class of models able to split the space of predictors according to the maximization (or minimization) of a measure of interest (typically Gini or Entropy measure). Decision trees, specifically classification trees in this context, are rather flexible and easy to fit and interpret but suffer of lack of stability and poor predictive performance. Insofar, RF can provide an improvement over a single tree by aggregating a number of trees on bootstrapped training samples. But when building these decision trees, each time a split in a tree is considered, a random sample of m predictors is chosen as split candidates from the full set of p predictors. Thus the split is allowed to use only one of those m predictors. This trick allows to improve the overall performance by considering different subset of variables, avoiding a possible bias given by just one very important predictor that would result in being always the first split of the tree. In this paper, we use RF more as an exploratory and confirmatory tool to highlight the most relevant predictors since we have too many variables compared to the size of the sample. This is an issue from a statistical point of view, since the estimates can result in being not reliable or not available at all (not enough data points, i.e. degrees of freedom, for the estimation phase). Considering the textual analysis of Telegram chat we take advantage of quantitative analysis of human languages to discover common features of written text. In particular the analysis of relatively short text messages like those appearing on micro-blogging platform presents a number of challenges. Some of these are, the informal conversation (e.g. slang words, repeated letters, emoticons) and the level of implied knowledge necessary to understand the the topics of discussion. Moreover, it is important to consider the high level of noise contained in the chats, witnessed by the fact that only a fraction of them with respect to the total number available is employed in our sentiment analysis. We have applied a Bag of Word (BoW) approach, where a text is represented as an unordered collection of words, considering only their counts in each comment of the chat. The word and document vectorization has been carried out by collecting all the word frequencies in a Term Document Matrix (TDM). Afterwards such matrix has been weighted by employing the popular TF-IDF (Term

Frequency Inverse Document Frequency) algorithm. Classical text cleaning procedures have been put in place like stop-words, punctuation, unnecessary symbols and space removal, specific topic words addition. For descriptive purposes we have used wordclouds for each and every Telegram chat according to the general content and to specific subcategories like sentiments and expressed moods. The most critical part of the analysis relies on the sentiment classification. In general, two different approaches can be used:

- Score dictionary based: the sentiment score is based on the number of match between predefined list of positive and negative words and terms contained in each text source (a tweet, a sentence, a whole paragraph);
- Score classifier based: a proper statistical classifier is trained on a large enough dataset of prelabeled examples and then used to predict the sentiment class of a new example.

However, the second option is rarely feasible because in order to fit a good classifier, a huge amount of pre-classified examples is needed and this represents a particularly complicated task when dealing with short and extremely non conventional text like micro-blogging chats. Insofar, we decided to focus on a dictionary based approach, adapting appropriate lists of positive and negative words relevant for ICOs topics in English language. We employ 3 vocabularies from the R package 'tidytext':

- AFINN from Finn rup Nielsen;
- BING from Bing Liu and collaborators;
- NRC from Saif Mohammad and Peter Turney.

All three of these lexicons are based on unigrams, i.e., single words. These lexicons contain many English words and the words are assigned scores for positive/negative sentiment, and also possibly emotions like joy, anger, sadness, and so forth. The nrc lexicon categorizes words in a binary fashion (yes/no) into categories of positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The bing lexicon categorizes words in a binary fashion into positive and negative categories. The AFINN lexicon assigns words with a score that runs between -5 and 5, with negative scores indicating negative sentiment and positive scores indicating positive sentiment.

3. Data

In this preliminary work we examine 120 ICOs occurred in 2017 and 2018. For each project we gather information from web-based sources, mainly rating platforms such as: icobench.com, TokenData.io, ICO Drops.com, CoinDesk.com and project' s websites.

The process of building up the ICOs data set reflects the main phases that an ICOs follows to be launched: from the birth of the business idea, the team building, the purpose of the tokens, the technical requirements (white paper), the promotion and the execution phase.

3.1. A. Retrieving ICO data from Internet.

The first step in collecting data about each project is to collect information from the most used Internet sources as icobench, TokenData or similar. In this step we look for general characteristics such as the name, the token symbol, start and end dates of the crowdfunding, the country of origin, financial data such as the total number of issued token, the initial price of the token, the platform used, data on the team proposing the ICO, data on the advisory board, data on the availability of the website, availability of white paper and social channels.

Some of these data, such as short and long description, and milestones are textual descriptions. Others are categorical variables, such as the country, the platform, the category (which can assume many values), and variables related to the team members (name, role, group). The remaining variables are numeric, with different degrees of discretization. Unfortunately, not all ICOs record all variables, so there are several missing data. The ICO web databases that we use are fully checked in order to minimize the missing values of one of the platforms, therefore we validate the information checking for the details on the website and on the white paper. As a result, the complete set of reliable information comes from the matching between the website and the white paper.

Figure 1 about here

The variables set, continuous and categorical data, show us that the main area of origin of the projects is Europe with the highest percentage in Switzerland and Germany. As shown in in Fig.1 Europe world region is followed by the USA and by Asian countries as Singapore and Hong Kong. The Switzerland peak is due to the national regulator approach - FINMA (Financial

Market Supervisory Authority)- that in 2015 proposed to equate legal status of cryptocurrency in the country to foreign currencies. By doing so, transactions with cryptocurrencies are not subject to VAT, which is in line with the existing practice in the EU. Most recently in July 2017, the Swiss Federal Council in its official press release, announced the creation of a "normative sandbox" aimed at creating an enabling environment for start-ups in the field of financial technologies. In the fragmented regulatory framework this is one of the so called "crypto-friendly" countries, that attract worldwide investors.

3.2. B. Unstructured data

Social channels are more personal than every database, rating platform or websites, so they are a way to reach a wide range of users, to update them constantly about the evolution of the project and in the end to create a trusty environment that can finalize in a successful crowdfunding activity. In order to conduct the textual analysis, we enrich our database with the social channels data, such as the presence of a channel, the numbers of users as a proxy of the community engagement and as mentioned in the introduction the textual chat, retrieved backward till the creation of the chat. The most used social channels are Telegram, Twitter, Facebook, Bitcointlak, Medium, while Linkedin, Reddit and Slack are not frequently used.

In crowdfunding projects the entrepreneur and the community in which is embedded works as a strong control for the attractiveness of a business. Some studies have investigated the social network community and the entrepreneurial activity finding out that the amount of capital collected in crowdfunding is heavily dependent on the range of social networks the entrepreneurs belong to (E. Mollick, 2014).

With regards to the entrepreneurial dimension, we investigate the team components, pointing out that the members checked until now are almost 1000, with a median size of 7 for project. For each team member we checked general informations related to the social engagement, looking for the Linkedin channel activity (48 % of them do not have an individual page), the numbers of connections, the job position in the project and the academic background.

Moreover the presence of advisors can play a crucial role in ensuring the reliability of an ICO, provided a wise choice of such advisors. The same applies to institutional investors doing due diligence on a potential venture. In collecting our data, we focused on the academic background and the current area of expertise of the declared advisors.

As it concerns the unstructured data, insightful information can be derived by the white papers in terms of quality of the technical report and specific content. A white paper is a summary report that provides detailed information about the project, its originality and the benefits it can give to investors and users, about the technological features, team behind the project, project's background and future plans.

In Table 1 we report some characteristics of our current sample, consisting of 120 ICOs. In Table 2 we report the complete list of collected and employed variables.

Table 1 about here

Table 2 about here

4. Empirical Evidence

In this section we report our main results obtained from classification analysis and textual analysis.

Table 3 about here

Table 4 about here

Regarding the former table 3 and table 4 report results respectively for logistic regression on Success/Failure (class 1 variable) and for GEV logistic regression on Scam/non Scam (class 2 variable). The reader can see in table 3 that the only two relevant dummy variables are: the presence of a white paper and of a Telegram chat. Both present positive coefficients showing their impact on the increasing of the probability of success of an ICO. Regarding the two continuous variables, number of elements of the team and number of advisors (both appropriately standardize), are highly significant and positive suggesting that increasing people and advisors has a positive impact. In table 4 we can see results for scam ICOs, on the basis of a logistic regression modified for highly rare events as it occurs in our analysis (only 8 scam ICO out of 120 monitored). Reminding that the target variable 'class1' is labeled with 0 for scam and 1 otherwise, we can infer that both the presence of a website and of the Twitter account have a positive impact in not being a scam. In other words, the absence of these two characteristics is a driver of scam activity suspects. Considering the two continuous variables, number of components of the team and number of advisors, they have been evaluated with a non linear effect (smooth component) and similarly to previous results, we notice a positive impact. Thus, the increasing in the number of people engaged within an ICO, impacts positively on the probability of not being a scam even if not in a linear way.

Further analysis has been conducted on the textual part based on Telegram chats. For sake of simplicity and readability we report an example of ICO for each category of interest: Zenome ICO as a failure example, Fitrova as a scam example and Bancor as a successful one. Figure 2, 3 and 4 report respectively wordclouds of negative words for the 3 ICOs.

Figure 2 about here

Figure 3 about here

Figure 4 about here

It appears the difference in the most relevant words, in particular the failed ICO present words like 'hard', 'difficult', 'worries', 'fake' partially in common with the scam ICO. The latter indeed presents also specific words like 'crazy', 'loss', 'regret'.

From Figure 5 through 7 we compare the most frequent positive and negative words for the 3 ICOs.

Figure 5 about here

Figure 6 about here

Figure 7 about here

It appears that the successful ICO (Bancor) in Figure 7 has more positive than negative words and considering the negative ones they are mainly related to possible 'issues' while 'scam' word has a very low frequency. Instead the two other ICOs have higher frequency for the negative words with a clear pick for words like 'scam', 'bad','hard'.

If we consider sentiment based on AFINN vocabulary and we produce a sentiment score for the 3 considered ICOs, we obtain results reported in table 5. The score is standardized according to the length of the chat, in order to avoid any bias in the comparison. It clearly appears that the successful ICO shows an higher score compared to the other two, so suggesting to further investigate the hypothesis of relevance in the information reported by Telegram users.

5. Conclusions

In this paper we address the issue of discovering the success drivers of an ICO. Initial coin offering (aka ICO) represents one of the several by-product of the cryptocurrencies world. New generation start-up and existing businesses in order to avoid rigid and long money raising protocols imposed by classical channels like banks or venture capitalists, offer the inner value of their business by selling tokens, i.e. units of the chosen cryptocurrency, like a regular firm would do with and IPO. The investors of course hope in a value increasing of the tokens in the near future, provided a solid and valid business idea typically described by the ICO issuers in a white paper, both a descriptive and technical report of the proposed business. However, fraudulent activities perpetrated by unscrupulous start-up can happen and it would be crucial to highlight in advance clear signs of illegal money raising.

While analyzing success vs failure dynamic with a classification model is relatively easy since the incidence of the two classes is almost equal (50-50), it is much more complicated to highlight the key aspects that could witness a fraudulent activity since, in the last 3 years, only few scam events have been reported. In our sample made of 120 ICOs (data collection still active) we have 8 scam ICOs and by fitting a logistic regression model for highly unbalance data, our preliminary results tell that both the presence of a website and of the Twitter account have a positive impact in not being a scam. In other words, the absence of these two characteristics is a driver of scam activity suspect. Considering the two continuous variables, number of components of the team and number of advisors, they have been evaluated with a non linear effect (smooth component) and we notice a positive impact. Thus, the increasing in the number of people engaged within an ICO, impacts positively on the probability of not being a scam even if not in a linear way.

First text analysis on the Telegram chats highlights a difference in the sentiment expressed for the 3 considered ICOs. More negative and specific words appear with high frequency if we consider a scam or failed ICO compared to a successful one. The same trend is confirmed by a sentiment score calculated on the basis of AFFIN vocabulary.

This paper represents a preliminary work and more detailed analysis is needed. Specifically, apart from increasing the size of the sample, we are improving the textual analysis with specific attention to sentiment analysis. We aim at producing a sentiment score for each ICO to be included in the classification models, as a possible driver of success and/or scam activity.

6. References

S. Adhami, G. Giudici, S. Martinazzi, Why Do Businesses Go Crypto? An Empirical Analysis of Initial Coin Offerings, October 20 2017, available at SSRN: https://ssrn.com/abstract=3046209

Calabrese R., Marra G., Osmetti S.A. (2016), Bankruptcy Prediction of Small and Medium Enterprises Using a Flexible Binary Generalized Extreme Value Model. Journal of the Operational Research Society, 67(4), 604-615

Coinschedule, (2018). "Cryptocurrency ICO Stats 2018" available at: https://www.coinschedule.com/stats.html [Accessed June 18, 2018] Coinmarketcap, (2018). "All coins" available at:
https://coinmarketcap.com/coins/views/all/ [Accessed June 18, 2018]

Ethereumscamdb, "Ethereum Scam Database" available at:
https://etherscamdb.info/ [Accessed June 18, 2018] Falk, M., Haler, J., and Reiss, R. (2010). Laws of Small Numbers: Extremes and Rare Events. Springer.

A. Hankin, "Nearly half of all 2017 ICOs have failed". Fortune, February 26, 2018 [Retrieved May 4, 2018].

Kotz, S. and Nadarajah, S. (2000). Extreme Value Distributions. Theory and Applications. Imperial College Press, London.

Peer-to-Peer Research Group, (2013).Peer-to-Peer Research Group available at:
https://irtf.org/concluded/p2prg [Accessed 18 June 2018]

E. Mollick, "The dynamics of crowdfunding: an exploratory study" J. Bus. Venturing, 29 (1) (2014), pp. 1-16

H. Subramanian, Decentralized Blockchain-Based Electronic Marketplaces, Comm. ACM, vol. 61, pp. 78-84, January 2018.

D.A. Zetzsche, R.P. Buckley, D.W. Arner, L. Fhr, The ICO Gold Rush: It's a Scam, It's a Bubble, It's a Super Challenge for Regulators, Univ.

Luxembourg Law Working Paper No. 11/2017, available at SSRN: https://ssrn.com/abstract=3072298, November 2017.

A. Wright, P. De Filippi, ”Decentralized Blockchain Technology and the Rise of Lex Cryptographia”, March 10, 2015, available at SSRN: https://ssrn.com/abstract=2580664 or http://dx.doi.org/10.2139/ssrn.2580664

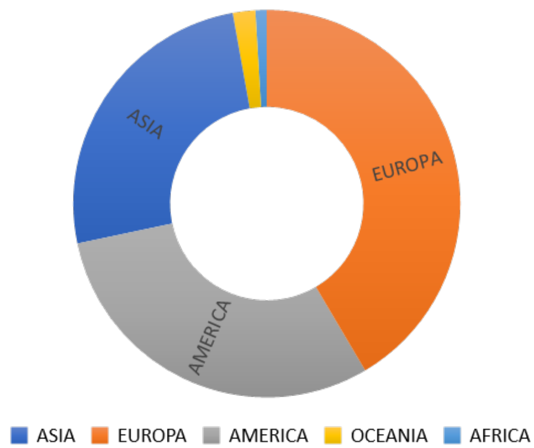

Figure 1: Geographical area of origin

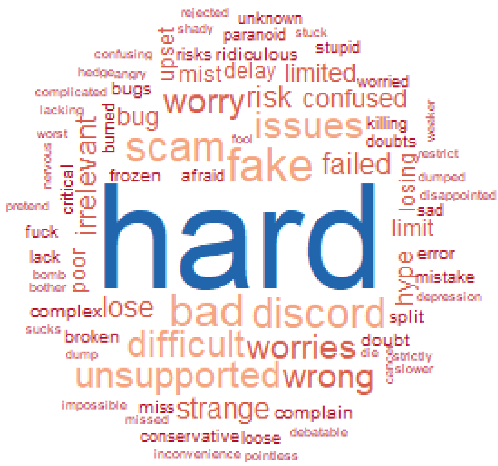

Figure 2: Failed ICO: negative words

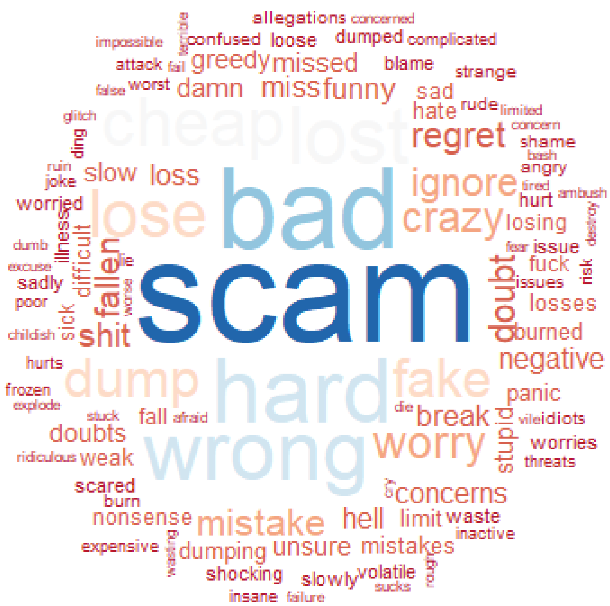

Figure 3: Scam ICO: negative words

Table 1: Sample characteristics

Status		
status	**nr**	**%**
success	77	64%
failed	36	30%
scam	8	7%
Purpose of projects		
purpose	**nr**	**%**
financial service	39	29%
market places and exchanges	26	19%
high tech services	20	15%
others	20	15%
smart contract	12	9%
media and entertainment	7	5%
gambling platfoms	5	4%
gaming	4	3%
adult entertainment	1	1%

Figure 4: Success ICO: negative words

Table 2: Esplicative variables

class0	f=failed, sc=scam su=success
class1	0=scam, 1=failed+success
class2	0=failed, 1= success
w_site	Website (dummy)
tm	Telegram (dummy)
w_paper	White paper (dummy)
usd	presale price in USD
tw	Twitter (dummy)
fb	Facebook (dummy)
ln	Linkedin (dummy)
yt	Youtube (dummy)
gith	Github (dummy)
slack	Slack (dummy)
reddit	Reddit (dummy)
btalk	Bitcointalk (dummy)
mm	Medium (dummy)
nr_team	Number of Team members
adv	Existence of advisors (dummy)
nr_adv	Number of advisors
project	Official name of the ICO
nr_tm	Number of users in Telegram
tot_token	Number of Total Tokens

Table 3: Results from Logistic regression on Success/Failure

	Dependent variable:
	class2
nr_team	4.522***
	(1.494)
nr_adv	1.686***
	(0.634)
w_paper	3.113***
	(1.147)
tm	1.917**
	(0.955)
Constant	−2.189
	(1.458)
Observations	120
Log Likelihood	−28.308
Akaike Inf. Crit.	66.616
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: Results from Gev logistic regression on Scam/non scam

	Dependent variable:
	class1
w_site	2.0115^{***}
	(0.490)
tw	1.230^{*}
	(0.597)
s(nr_team_st)	3.973^{***}
	(smooth components)
s(nr_adv_st)	2.057^{***}
	(smooth components)
Constant	0.9894
	(0.927)
Observations	120
tau	−0.25
total edf	9.03
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5: Sentiment AFINN based for the 3 ICOs

ICOs	AFINN
zenome (scam)	0.09
fitrova (failed)	0.26
bancor (success)	0.45

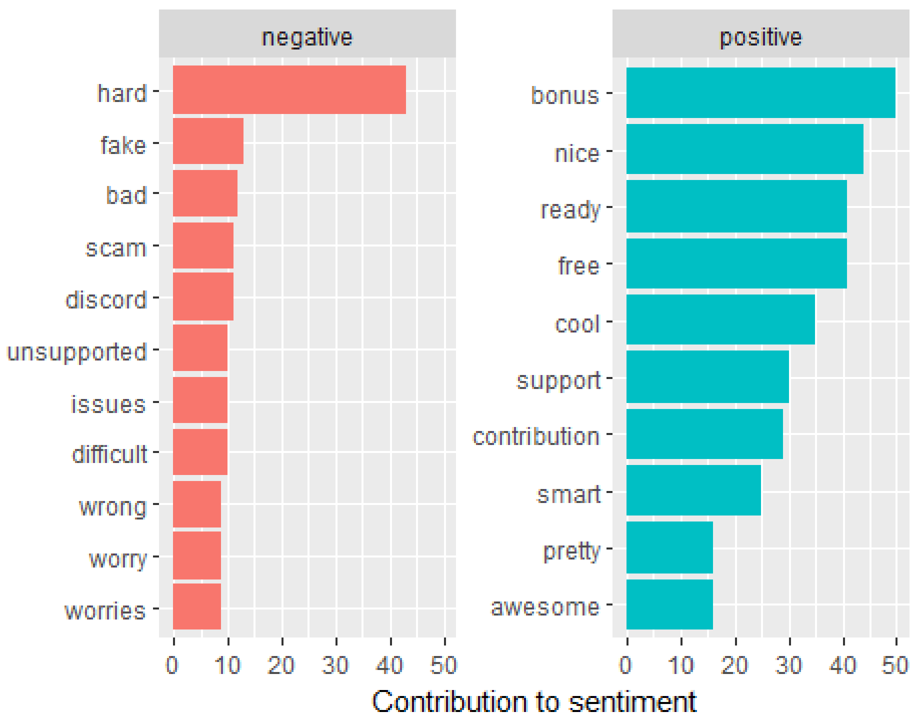

Figure 5: Zenome ICO (failed), most negative and positive words

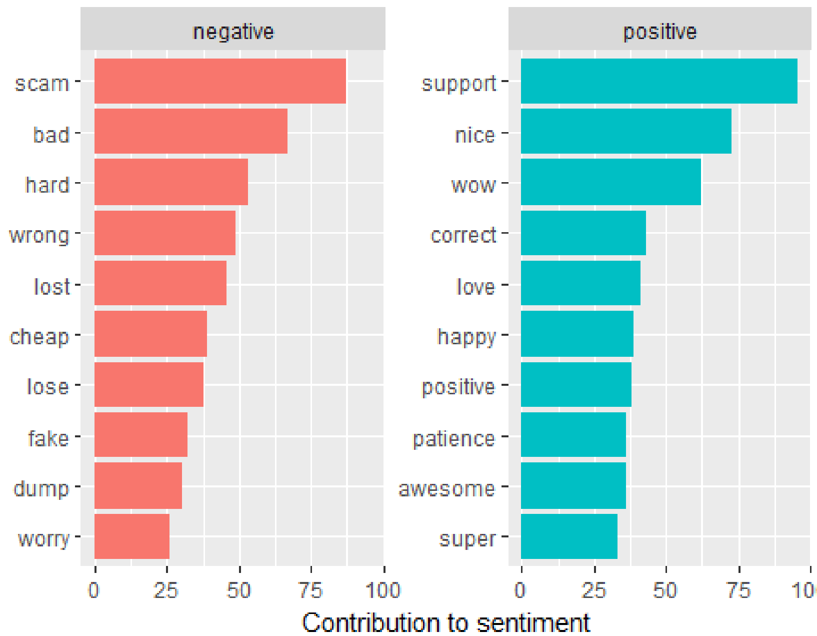

Figure 6: Fitrova ICO (scam), most negative and positive words

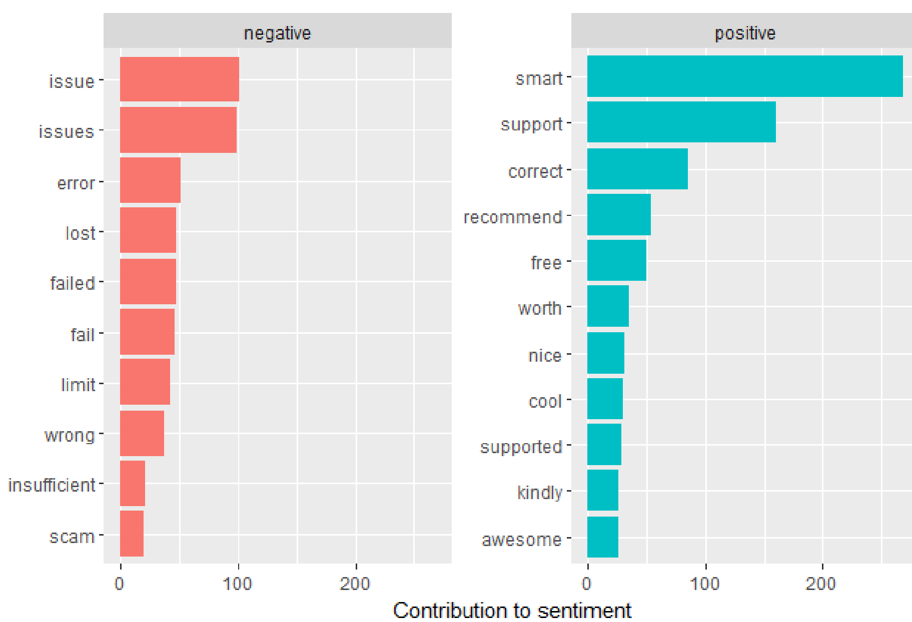

Figure 7: Bancor ICO (success), most negative and positive words