A Social Network Analysis based approach to investigate user behaviour during a cryptocurrency speculative bubble

Journal of Information Science 2023, Vol. 49(4) 1060–1085

DOI: 10.1177/01655515211047428 journals.sagepub.com/home/jis

A Social Network Analysis based approach to investigate user behavior during a cryptocurrency speculative bubble

Gianluca Bonifazi, Enrico Corradini, Domenico Ursino*, Luca Virgili

Department of Information Engineering, Polytechnic University of Marche
* Corresponding author
g.bonifazi@univpm.it; e.corradini@pm.univpm.it; d.ursino@univpm.it; l.virgili@pm.univpm.it

Abstract

In this paper, we present a Social Network Analysis based approach to investigate user behavior during a cryptocurrency speculative bubble in order to extract knowledge patterns about it. Our approach is general and can be applied to any past, present and future cryptocurrency speculative bubble. To verify its potential, we apply it to investigate the Ethereum speculative bubble happened in the years 2017 and 2018. We also describe several interesting knowledge patterns about the behavior of specific categories of users that we obtained from this investigation. Furthermore, we describe how our approach can support the construction of an identikit of the speculators who maneauvered behind the Ethereum bubble analyzed. Finally, we show that this capability of supporting the hunting for speculators is intrinsic of our approach and can cover past, present and future bubbles.

Keywords: Social Network Analysis; Blockchain; Cryptocurrency; Ethereum; Speculative Bubble; Speculator Identikit

1 Introduction

The popularity of blockchains has been growing continuously from the appearance of Bitcoin in 2008 [32], and the interest on cryptocurrencies followed the same growth. In the past years, cryptocurrencies were the subject of a speculative bubble, similar to the tulipans' and stock market ones [38]. For instance, the price of Bitcoin surged almost 2,800% in four years and has fallen by 80% in just few weeks, between the end of 2017 and the beginning of 2018, leading to a huge gain for a few people and a big loss for the majority of the investors. These events are interesting to investigate from a data science perspective, because they allow the extraction of knowledge patterns to prevent other similar cases. As a matter of fact, several studies investigate the whole speculative cryptocurrency bubble and its consequences for economy and technology [39, 19]. However, a very limited number of studies take the intrinsic nature of blockchain as a social network into account. Actually, the relationships between blockchain users are extremely relevant in the extraction of unknown patterns and in the disclosure of new viewpoints for analyzing this speculative bubble. For this reason, Social Network

Analysis notions [17, 24] can provide a big help to study the relationships in the blockchain network. In this activity, it is reasonable to think of a social network in which each node indicates a user, represented through her/his blockchain address, whereas each arc denotes a transaction between two users. Starting from this idea, in this paper we propose a Social Network Analysis based approach to extract knowledge patterns concerning user behavior during a cryptocurrency speculative bubble. Our approach is general and can be used to investigate any past, current and future cryptocurrency bubble. In order to understand its potential, we applied it to the Ethereum bubble happened in the years 2017 and 2018 and considered the pre-bubble, bubble and post-bubble phases.

In order to reach its goals, our approach focuses on certain categories of users, namely: *(1) The power addresses*, i.e., the most active users on the blockchain of interest, who were responsible for most of the transactions of this network. More specifically, we consider the power addresses for each of the periods of interest (i.e., the pre-bubble, bubble and post-bubble). *(2) The Survivors*, i.e., those users who were power addresses in all the three periods of interest. *(3) The Missings*, i.e., those users who were power addresses in the pre-bubble period and stopped being power addresses in the bubble and post-bubble periods. *(4) The Entrants*, i.e., those users who were not power addresses in the pre-bubble period and became power addresses in the bubble and post-bubble periods.

Then, for each user category, our approach employs Social Network Analysis based techniques to identify the main characteristics that distinguish the corresponding users from the others. In this task, the concept of ego network [13] plays an important role. Afterwards, our approach checks if and when there are backbones linking the users of a certain category. The presence of such backbones can be hypothesized on the basis of the principle of homophily [30], characterizing many social networks. However, only a set of experimental analyses can show whether this hypothesis is true or not. Also in this case, ego networks play a key role to support analytical investigations. They are flanked by k-cores [14], which help in giving a graphical idea of the analytical results. Finally, the last tasks of our approach aim at predicting, given a certain period (i.e., pre-bubble, bubble), who will be the main actors in the next ones (i.e., bubble, post-bubble), based on some parameters. The experiments we performed on Ethereum to test our approach and its potential allowed us to extract knowledge patterns concerning user behavior during the speculative bubble happened in the years 2017 and 2018.

Finally, we show how our approach can allow the definition of the identikit of cryptocurrency bubble speculators and, therefore, how it can be applied to hunt past, present and future cryptocurrency bubble speculators.

The outline of this paper is as follows: In Section 2, we present related literature and highlight the novelties of our approach with respect to it. In Section 3, we illustrate the proposed approach. In Section 4, we apply it to investigate the speculative bubble involving Ethereum in the years 2017 and 2018. In Section 5, we show the application of our approach to hunt cryptocurrency bubble speculators. Finally, in Section 6, we draw our conclusions and highlight some possible future developments of our research.

2 Related literature

Since the introduction of Bitcoin in 2008 [32], thousands of cryptocurrencies have been created [11], and the interest about them has increased significantly. At the same time, the scientific literature

about blockchain and digital currencies has progressively grown [25, 15, 1, 36, 37]. The spread of this new technology has also created a lively discussion in the economic field on the possibility of speculations around these assets [5, 2, 7, 4].

Indeed, at the end of 2017, the price of Bitcoin (as well as the ones of the other cryptocurrencies, like Ethereum or Litecoin) increased by almost 600% (reaching an all-time high value of $19,475.80) before falling by 80% in few weeks, until January 2018 [39, 27, 3, 16]. This is the biggest bubble in the cryptocurrencies history so far. Researchers have strived to analyze every detail of this particular event to understand the corresponding dynamics in order to prevent other speculations in the future. For instance, in [40], the authors investigate market efficiency and volatility persistence in 12 highly priced and capitalized cryptocurrencies, based on daily data from August 7^{th}, 2015 to November 28^{th}, 2018. They observe a random walk pattern in returns of most cryptocurrencies, including Bitcoin and Ethereum, making the price of these assets unpredictable.

In [12], the authors examine the existence and the time intervals of pricing bubbles in Bitcoin and Ethereum. Specifically, they adopt three measures to best represent the key theoretical components of cryptocurrency pricing structures, namely: *(i)* the mining difficulty, which reflects how difficult it is computing the next block of the blockchain; *(ii)* the hashrate, which represents the speed at which a computer is completing an operation in the blockchain code; *(iii)* the relationship between cryptocurrency returns, volatility and liquidity. This study highlights that there are periods characterized by a clear bubble behavior. The period between 2017 and 2018 could be identified as one of them.

Another interesting research field in digital currencies regards the definition of approaches to predict speculative bubbles [9, 18, 34]. For instance, in [19], the authors introduce an automatic peak detection method that classifies price time series into periods of uninterrupted market growth (i.e., drawups) and periods of uninterrupted market decrease (i.e., drawdowns). In [31], the authors investigate a new approach to predict speculative bubbles involving four cryptocurrencies (Bitcoin, Litecoin, Ethereum, and Monero) based on the behavior of new online social media indicators. For this purpose, they leverage a Hidden Markov Model for detecting epidemic outbreaks in the blockchain setting. In [8], the authors propose another possible way to detect speculative bubbles in cryptocurrencies through an approach based on a social microblogging platform for investors and traders. Specifically, they evaluate the sentiment of users on StockTwits[1] and, then, exploit it as a transition variable in a smooth transition autoregression.

A further approach to investigate the cryptocurrencies market is based on the analysis of the corresponding blockchain. It starts from the consideration that a blockchain represents a public ledger in which all committed transactions are stored in a chain of blocks [42, 41]. This chain can be represented and analyzed like a graph with nodes and edges [23, 35, 6, 22]. This reasoning leads the authors of [26] to examine the transaction network of Bitcoin during the first four years of its existence. The results obtained outline the business distribution by countries and their evolution over time. The authors also show that there is a gambling network that features many small transactions. In [28], the authors present a set of analyses on the user graph obtained by performing a heuristic clustering of the Bitcoin blockchain graph. They figure out a set of interesting properties of the network, including the "rich get richer" property and the existence of central nodes acting as privileged bridges between

[1] `https://stocktwits.com/`

different parts of the network. Finally, in [35], the authors exploit network analysis techniques to investigate the trading dynamics of ERC20 Blockchain. They model ERC20 as a social network, which nodes represent all trading wallets and which edges stand for the buy-sell trades. This social network is inline with the current network theory expectations and presents strong power law properties.

Our approach is in line with the latest ones mentioned above, because it uses Data Science [21, 29], and in particular Social Network Analysis [20, 33], to investigate a blockchain. However, it presents some novelties with respect to them. Indeed, it introduces several categories of users, based on their behavior in the pre-bubble, bubble and post-bubble periods. Moreover, it leverages ego networks [13] and k-cores [14] to identify the characteristics of the various categories of users. Although ego networks and k-cores are classical tools of Social Network Analysis, to the best of our knowledge, they have never been employed to study the behavior of users during a cryptocurrency bubble. Furthermore, it detects the existence of backbones linking users of certain categories in the pre-bubble, bubble and post-bubble periods, which is a knowledge not found in past literature on the cryptocurrency bubbles. Finally, similarly to other papers mentioned above, it also presents a prediction task. However, it differs from the previous approaches for the target of the prediction, which, in this case, concerns the discovery, in a certain period (pre-bubble, bubble), of the most relevant features of the users who will be the main actors in the next period.

3 Proposed approach

In order to reach its goals, our approach simply needs a set of records, each corresponding to a transaction carried out in the blockchain of interest. Each record stores only four fields related to the associated transaction, namely: *(i)* the blockchain address starting it; *(ii)* the blockchain address receiving it; *(iii)* its timestamp; *(iv)* the amount of money transferred. These fields are very general and available for any cryptocurrency blockchain. A pseudocode formalization of our approach is shown in Algorithms 1 and 2.

Our approach receives the cryptocurrency blockchain $\mathcal{B}$ of interest and the time interval I during which there was a speculative bubble involving $\mathcal{B}$.

It first calls the function `Extract_Dataset()`, which returns the dataset D of the transactions of $\mathcal{B}$ during I. Next, it activates the function `Determine_Intervals()` to partition I into three sub-intervals I_{Pre}, I_B and I_{Post}, related to the pre-bubble, bubble and post-bubble periods, respectively. After that, it calls the functions `Detect_From_Power_Addresses()`, `Detect_To_Power_Addresses()` and `Detect_Super_Power_Addresses()` to determine the power addresses with the largest number of incoming arcs, outgoing arcs and both. Finally, it activates the functions `Detect_Multi_Interval_From_Power_Addresses()` and `Detect_Multi_Interval_To_Power_Addresses()` to determine the addresses that continue to be *From_Power_Addresses* and *To_Power_Addresses* when passing from the pre-bubble period to the bubble one and from the bubble period to the post-bubble one.

At this point, our algorithm has collected all the data necessary to activate `Detect_Survivors()`, `Detect_Missings()` and `Detect_Entrants()`, which aim at determining the Survivors $\mathcal{S}^F$ and $\mathcal{S}^T$, the Missings $\mathcal{M}^F$ and $\mathcal{M}^T$ and the Entrants $\mathcal{E}^F$ and $\mathcal{E}^T$. Next, it activates the function `Construct_Social_Networks()`, which returns the social networks $\mathcal{N}_{Pre}$, $\mathcal{N}_B$ and $\mathcal{N}_{Post}$ relative to the pre-bubble, bubble and post-bubble period. After that, it calls the functions `Construct_Survivors_`

Algorithm 1 Investigating user behavior during a cryptocurrency speculative bubble (first part)

Input

- $\mathcal{B}$: the cryptocurrency blockchain of interest
- I: the time interval to investigate

Output

- PA^F_{Pre}, PA^F_B, PA^F_{Post}, PA^T_{Pre}, PA^T_B, PA^T_{Post}: power addresses of the dataset
- SPA_{Pre}, SPA_B, SPA_{Post}, PA^F_{Pre-B}, PA^F_{B-Post}, PA^T_{Pre-B}, PA^T_{B-Post}: power addresses of the dataset
- $\mathcal{S}^F$, $\mathcal{S}^T$: the Survivors; $\mathcal{M}^F$, $\mathcal{M}^T$: the Missings; $\mathcal{E}^F$, $\mathcal{E}^T$: the Entrants
- $EgoKPSet$: a set of knowledge patterns derived from the ego network analyses
- $BackboneKPSet$: a set of knowledge patterns on the possible presence of backbones
- $BSurvivorsSet$: a set of potential Survivors
- $PBSurvivorsSet$: a set of potential Survivors
- $PBEntrantsSet$: a set of potential Entrants

Require:

- D: a dataset of transactions;
- I_{Pre}, I_B, I_{Post}: time intervals;
- $\mathcal{N}_{Pre}$, $\mathcal{N}_B$, $\mathcal{N}_{Post}$: social networks;
- $ENSet^{\mathcal{S},F}_{Pre}$, $ENSet^{\mathcal{S},T}_{Pre}$, $ENSet^{\mathcal{M},F}_{Pre}$, $ENSet^{\mathcal{M},T}_{Pre}$, $ENSet^{\mathcal{E},F}_{Pre}$, $ENSet^{\mathcal{E},T}_{Pre}$: a set of ego networks;
- $ENSet^{\mathcal{S},F}_{B}$, $ENSet^{\mathcal{S},T}_{B}$, $ENSet^{\mathcal{M},F}_{B}$, $ENSet^{\mathcal{M},T}_{B}$, $ENSet^{\mathcal{E},F}_{B}$, $ENSet^{\mathcal{E},T}_{B}$: a set of ego networks;
- $ENSet^{\mathcal{S},F}_{Post}$, $ENSet^{\mathcal{S},T}_{Post}$, $ENSet^{\mathcal{M},F}_{Post}$, $ENSet^{\mathcal{M},T}_{Post}$, $ENSet^{\mathcal{E},F}_{Post}$, $ENSet^{\mathcal{E},T}_{Post}$: a set of ego networks;
- T^F_{Pre}, T^T_{Pre}, T^F_B, T^T_B, T^F_{Post}, T^T_{Post}: top power addresses of the dataset;

D = Extract_Dataset($\mathcal{B}, I$)

$\langle I_{Pre}, I_B, I_{Post}\rangle$ = Determine_Intervals(D)

$\langle PA^F_{Pre}, PA^F_B, PA^F_{Post}\rangle$ = Detect_From_Power_Addresses($I_{Pre}, I_B, I_{Post}, D$)

$\langle PA^T_{Pre}, PA^T_B, PA^T_{Post}\rangle$ = Detect_To_Power_Addresses($I_{Pre}, I_B, I_{Post}, D$)

$\langle SPA_{Pre}, SPA_B, SPA_{Post}\rangle$ = Detect_Super_Power_Addresses($PA^F_{Pre}, PA^F_B, PA^F_{Post}, PA^T_{Pre}, PA^T_B, PA^T_{Post}$)

$\langle SPA^F_{Pre-B}, SPA^F_{B-Post}\rangle$ = Detect_Multi_Interval_From_Power_Addresses($PA^F_{Pre}, PA^F_B, PA^F_{Post}$)

$\langle SPA^T_{Pre-B}, SPA^T_{B-Post}\rangle$ = Detect_Multi_Interval_To_Power_Addresses($PA^T_{Pre}, PA^T_B, PA^T_{Post}$)

$\langle \mathcal{S}^F, \mathcal{S}^T\rangle$ = Detect_Survivors($PA^F_{Pre}, PA^F_B, PA^F_{Post}, PA^T_{Pre}, PA^T_B, PA^T_{Post}$)

$\langle \mathcal{M}^F, \mathcal{M}^T\rangle$ = Detect_Missings($PA^F_{Pre}, PA^F_B, PA^F_{Post}, PA^T_{Pre}, PA^T_B, PA^T_{Post}$)

$\langle \mathcal{E}^F, \mathcal{E}^T\rangle$ = Detect_Entrants($PA^F_{Pre}, PA^F_B, PA^F_{Post}, PA^T_{Pre}, PA^T_B, PA^T_{Post}$)

$\langle \mathcal{N}_{Pre}, \mathcal{N}_B, \mathcal{N}_{Post}\rangle$ = Construct_Social_Networks($I_{Pre}, I_B, I_{Post}, D$)

$\langle ENSet^{\mathcal{S},F}_{Pre}, ENSet^{\mathcal{S},T}_{Pre},\rangle$ = Construct_Survivors_Ego_Networks_Pre($I_{Pre}, \mathcal{N}_{Pre}, \mathcal{S}^F, \mathcal{S}^T$)

$\langle ENSet^{\mathcal{S},F}_{B}, ENSet^{\mathcal{S},T}_{B},\rangle$ = Construct_Survivors_Ego_Networks_Bubble($I_B, \mathcal{N}_B, \mathcal{S}^F, \mathcal{S}^T$)

$\langle ENSet^{\mathcal{S},F}_{Post}, ENSet^{\mathcal{S},T}_{Post},\rangle$ = Construct_Survivors_Ego_Networks_Post($I_{Post}, \mathcal{N}_{Post}, \mathcal{S}^F, \mathcal{S}^T$)

Algorithm 2 Investigating user behavior during a cryptocurrency speculative bubble (second part)

Require:

$\langle ENSet_{Pre}^{\mathcal{M},F}, ENSet_{Pre}^{\mathcal{M},T}, \rangle$ = Construct_Missings_Ego_Networks_Pre($I_{Pre}, \mathcal{N}_{Pre}, \mathcal{M}^F, \mathcal{M}^T$)

$\langle ENSet_{B}^{\mathcal{M},F}, ENSet_{B}^{\mathcal{M},T}, \rangle$ = Construct_Missings_Ego_Networks_Bubble($I_{B}, \mathcal{N}_{B}, \mathcal{M}^F, \mathcal{M}^T$)

$\langle ENSet_{Post}^{\mathcal{M},F}, ENSet_{Post}^{\mathcal{M},T}, \rangle$ = Construct_Missings_Ego_Networks_Post($I_{Post}, \mathcal{N}_{Post}, \mathcal{M}^F, \mathcal{M}^T$)

$\langle ENSet_{Pre}^{\mathcal{E},F}, ENSet_{Pre}^{\mathcal{E},T}, \rangle$ = Construct_Entrants_Ego_Networks_Pre($I_{Pre}, \mathcal{N}_{Pre}, \mathcal{E}^F, \mathcal{E}^T$)

$\langle ENSet_{B}^{\mathcal{E},F}, ENSet_{B}^{\mathcal{E},T}, \rangle$ = Construct_Entrants_Ego_Networks_Bubble($I_{B}, \mathcal{N}_{B}, \mathcal{E}^F, \mathcal{E}^T$)

$\langle ENSet_{Post}^{\mathcal{E},F}, ENSet_{Post}^{\mathcal{E},T}, \rangle$ = Construct_Entrants_Ego_Networks_Post($I_{Post}, \mathcal{N}_{Post}, \mathcal{E}^F, \mathcal{E}^T$)

$EgoKPSet$ = Analyze_Ego_Pre($ENSet_{Pre}^{\mathcal{S},F}, ENSet_{Pre}^{\mathcal{S},T}, ENSet_{Pre}^{\mathcal{M},F}, ENSet_{Pre}^{\mathcal{M},T}, ENSet_{Pre}^{\mathcal{E},F}, ENSet_{Pre}^{\mathcal{E},T}$)

$EgoKPSet = EgoKPset \cup$ Analyze_Ego_Bubble($ENSet_{B}^{\mathcal{S},F}, ENSet_{B}^{\mathcal{S},T}, ENSet_{B}^{\mathcal{M},F}, ENSet_{B}^{\mathcal{M},T}, ENSet_{B}^{\mathcal{E},F}, ENSet_{B}^{\mathcal{E},T}$)

$EgoKPSet = EgoKPset \cup$ Analyze_Ego_Post($ENSet_{Post}^{\mathcal{S},F}, ENSet_{Post}^{\mathcal{S},T}, ENSet_{Post}^{\mathcal{M},F}, ENSet_{Post}^{\mathcal{M},T}, ENSet_{Post}^{\mathcal{E},F}, ENSet_{Post}^{\mathcal{E},T}$)

$BackboneKPSet$ = Detect_Backbones_Survivor_Pre($ENSet_{Pre}^{\mathcal{S},F}, ENSet_{Pre}^{\mathcal{S},T}, \mathcal{S}^F, \mathcal{S}^T, \mathcal{M}^F, \mathcal{M}^T, \mathcal{E}^F, \mathcal{E}^T$)

$BackboneKPSet = BackboneKPSet \cup$ Detect_Backbones_Missing_Pre($ENSet_{Pre}^{\mathcal{M},F}, ENSet_{Pre}^{\mathcal{M},T}, \mathcal{S}^F, \mathcal{S}^T, \mathcal{M}^F, \mathcal{M}^T, \mathcal{E}^F, \mathcal{E}^T$)

$BackboneKPSet = BackboneKPSet \cup$ Detect_Backbones_Entrants_Pre($ENSet_{Pre}^{\mathcal{E},F}, ENSet_{Pre}^{\mathcal{E},T}, \mathcal{S}^F, \mathcal{S}^T, \mathcal{M}^F, \mathcal{M}^T, \mathcal{E}^F, \mathcal{E}^T$)

$BackboneKPSet = BackboneKPSet \cup$ Detect_Backbones_Survivor_Bubble($ENSet_{B}^{\mathcal{S},F}, ENSet_{B}^{\mathcal{S},T}, \mathcal{S}^F, \mathcal{S}^T, \mathcal{M}^F, \mathcal{M}^T, \mathcal{E}^F, \mathcal{E}^T$)

$BackboneKPSet = BackboneKPSet \cup$ Detect_Backbones_Missing_Bubble($ENSet_{B}^{\mathcal{M},F}, ENSet_{B}^{\mathcal{M},T}, \mathcal{S}^F, \mathcal{S}^T, \mathcal{M}^F, \mathcal{M}^T, \mathcal{E}^F, \mathcal{E}^T$)

$BackboneKPSet = BackboneKPSet \cup$ Detect_Backbones_Entrants_Bubble($ENSet_{B}^{\mathcal{E},F}, ENSet_{B}^{\mathcal{E},T}, \mathcal{S}^F, \mathcal{S}^T, \mathcal{M}^F, \mathcal{M}^T, \mathcal{E}^F, \mathcal{E}^T$)

$BackboneKPSet = BackboneKPSet \cup$ Detect_Backbones_Survivor_Post($ENSet_{Post}^{\mathcal{S},F}, ENSet_{Post}^{\mathcal{S},T}, \mathcal{S}^F, \mathcal{S}^T, \mathcal{M}^F, \mathcal{M}^T, \mathcal{E}^F, \mathcal{E}^T$)

$BackboneKPSet = BackboneKPSet \cup$ Detect_Backbones_Missing_Post($ENSet_{Post}^{\mathcal{M},F}, ENSet_{Post}^{\mathcal{M},T}, \mathcal{S}^F, \mathcal{S}^T, \mathcal{M}^F, \mathcal{M}^T, \mathcal{E}^F, \mathcal{E}^T$)

$BackboneKPSet = BackboneKPSet \cup$ Detect_Backbones_Entrants_Post($ENSet_{Post}^{\mathcal{E},F}, ENSet_{Post}^{\mathcal{E},T}, \mathcal{S}^F, \mathcal{S}^T, \mathcal{M}^F, \mathcal{M}^T, \mathcal{E}^F, \mathcal{E}^T$)

$\langle T_{Pre}^F, T_{Pre}^T, T_B^F, T_B^T, T_{Post}^F, T_{Post}^T, \rangle$ = Detect_Top_Power_Addresses($I_{Pre}, I_B, I_{Post}, D$)

$BSurvivorsSet$ = Predict_Bubble_Survivors($T_{Pre}^F, T_{Pre}^T, T_B^F, T_B^T, \mathcal{S}^F, \mathcal{S}^T, \mathcal{M}^F, \mathcal{M}^T, \mathcal{E}^F, \mathcal{E}^T, I_{Pre}, I_B, D$)

$PBSurvivorsSet$ = Predict_Post_Survivors($T_B^F, T_B^T, T_{Post}^F, T_{Post}^T, \mathcal{S}^F, \mathcal{S}^T, \mathcal{M}^F, \mathcal{M}^T, \mathcal{E}^F, \mathcal{E}^T, I_B, I_{Post}, D$)

$PBEntrantsSet$ = Predict_Post_Entrants($T_B^F, T_B^T, T_{Post}^F, T_{Post}^T, \mathcal{S}^F, \mathcal{S}^T, \mathcal{M}^F, \mathcal{M}^T, \mathcal{E}^F, \mathcal{E}^T, I_B, I_{Post}, D$)

return all outputs

`Ego_Network_Pre()`, `Construct_Ego_Network_Bubble()` and `Construct_Survivors_Ego_Network_Post()` to build the ego networks of the Survivors of the social networks $\mathcal{N}_{Pre}$, $\mathcal{N}_B$ and $\mathcal{N}_{Post}$. Similarly, it proceeds to call the suitable functions for constructing the ego networks of the Missings and the Entrants for the same social networks mentioned above.

The ego networks thus constructed represent the basis for the next analyses aimed at extracting a set $EgoKPSet$ of knowledge patterns on the characteristics of the Survivors, the Missings and the Entrants in the pre-bubble, bubble and post-bubble periods. Our algorithm performs this extraction by calling the functions `Analyze_Ego_Pre()`, `Analyze_Ego_Bubble()` and `Analyze_Ego_Post()`. The next analysis performed concerns the possible existence of backbones linking the Survivors, the Missings or the Entrants in the pre-bubble, bubble and post-bubble periods. To this end, our algorithm calls some functions having the objective of extracting the set $BackboneKPSet$ of knowledge patterns concerning the possible existence of backbones among the various kinds of address of interest.

Once the backbone investigation terminates, our algorithm proceeds with the last, predictive,

analysis. In fact, it aims at predicting, during a certain period, the nodes that will become protagonists in the next one. To this end, it calls the functions `Predict_Bubble_Survivors()`, `Predict_Post_Survivors()` and `Predict_Post_Entrants()`. The first examines nodes during the pre-bubble period and predicts which of them will form the set $BSurvivorsSet$ of potential Survivors during the bubble period. The second and the third examine the nodes during the bubble period and predict which of them will form the set $PBSurvivorsSet$ and $PBEntrantsSet$ of potential Survivors and Entrants during the post-bubble period.

The algorithm terminates returning all the information extracted through the calls of the functions mentioned above.

A more abstract and simplified graphical representation of our approach can be found in Figure 1.

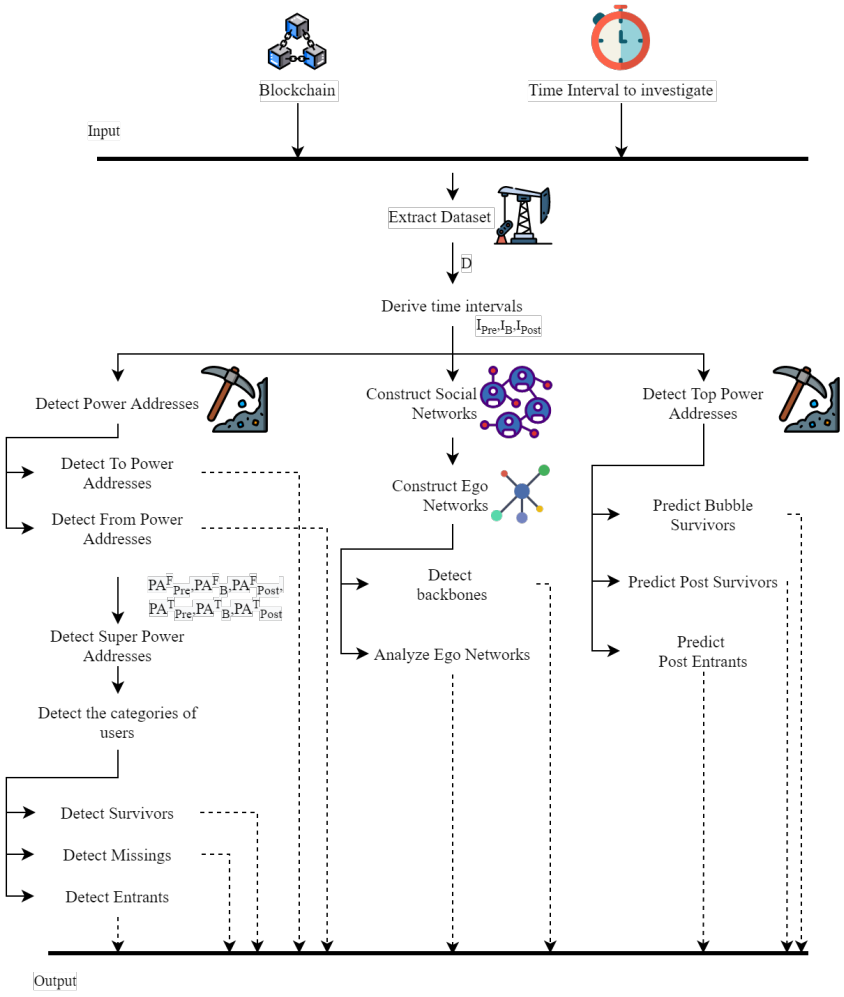

Figure 1: A graphical abstract representation of our algorithm

After describing our approach in detail, in the next section, we apply it on data concerning the speculative bubble that involved Ethereum in the years 2017 and 2018. The ultimate goal of this

activity is testing our approach and, at the same time, giving an idea of its potential regarding the extraction of knowledge patterns related to user behavior during the speculative bubble.

4 Application of our approach to the Ethereum speculative bubble of the years 2017 and 2018

In this section, we illustrate the application of our approach to a real-world case, namely the Ethereum speculative bubble happened in the years 2017 and 2018. In particular, we show how the various functions of the algorithm implementing our approach can be triggered on this data and illustrate the knowledge patterns they return. We start with a description of the dataset we used.

4.1 Dataset extraction

In this subsection, we illustrate the dataset we employed for our analyses. Its extraction corresponds to what is performed by the function `Extract_Dataset()` of our algorithm.

The dataset we used for our analysis is based on the Ethereum blockchain. As stated on the platform official website[2] "Ethereum is a technology that lets you send cryptocurrency to anyone for a small fee. It also powers applications that everyone can use and no one can take down". Ethereum is a programmable blockchain and represents the technological framework behind the cryptocurrency Ether (ETH).

Our dataset was downloaded from Google BigQuery[3]. It contains all the transactions made on Ethereum from January 1^{st}, 2017 to December 31^{st}, 2018. After some data cleaning operations, a row of the dataset, which represents a transaction, contains four columns, namely: *(i)* `from_address`, the blockchain address starting the transaction; *(ii)* `to_address`, the blockchain address receiving the transaction; *(iii)* `timestamp`, the transaction timestamp; *(iv)* `value`, the amount of Weis[4] transferred during the transactions.

The dataset is made of 354,107,563 transactions; the total number of user addresses is 43,537,168. We computed some statistics on it, which are reported in Table 1.

Parameter	*Value*
Number of transactions	354,107,563
Total number of `from_addresses`	38,881,752
Total number of `to_addresses`	42,457,991
Cardinality of the intersection between `from_addresses` and `to_addresses`	37,802,576
Number of null `from_addresses`	2,104,863
Number of null `to_addresses`	0

Table 1: Some preliminary statistics performed on our dataset

[2] `https://ethereum.org/en/what-is-ethereum/`

[3] `https://www.kaggle.com/bigquery/ethereum-blockchain`

[4] Wei is the smallest denomination of Ether; it corresponds to 10^{-18} Ethers.

4.2 Defining the user categories of interest

In this section, we present some preliminary analyses "depicting" the pre-bubble, bubble and post-bubble periods, as well as the general behavior of users during the two years covered by our dataset and, especially, during the three periods of our interest. At the end of these analyses, we will be able to define the user categories of interest.

A first analysis concerns the distributions of the number of transactions against `from_addresses` and `to_addresses`. They are reported in Figure 2. This figure shows that the two distributions follow a power law. We computed some parameters for them; they are reported in Table 2.

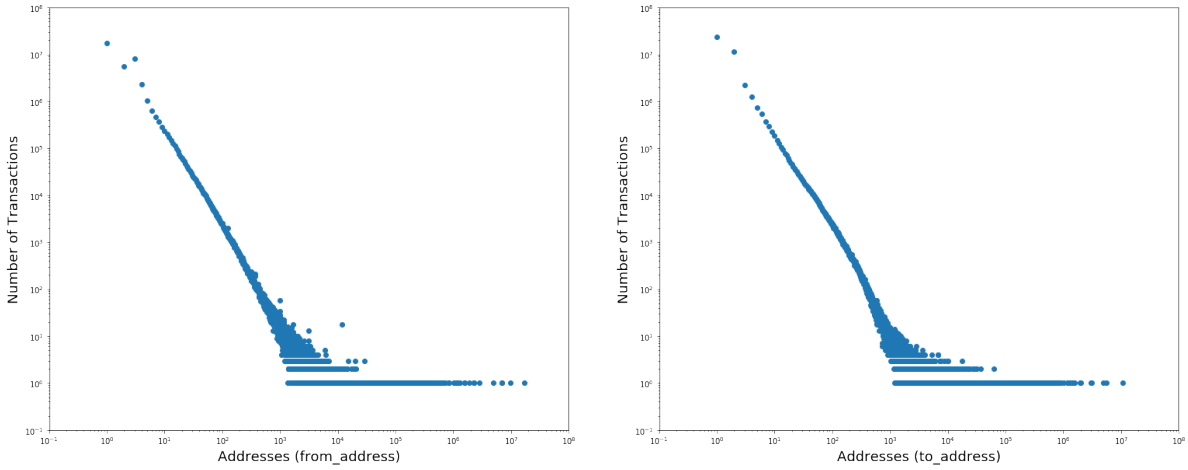

Figure 2: Log-log plots of the distributions of transactions against `from_addresses` (at left) and `to_addresses` (at right)

Parameter	`from_addresses`	`to_addresses`
Maximum number of transactions	17,509,218	23,404,261
Average number of transactions	5,640.76	5,913.37
α (power law parameter)	1.477	1.565
δ (power law parameter)	0.013	0.074

Table 2: Values of the parameters of transaction distributions against addresses

From the analysis of both Figure 2 and Table 2 we can observe that the two power law distributions are similar.

The second analysis that we take into consideration concerns the variation of the number of transactions over time. The purpose of this analysis is the identification of the pre-bubble, bubble and post-bubble periods. This trend is shown in Figure 3. From the analysis of this figure we can see that from January 2017 to October 2017 there is a substantially linear growth of the number of transactions. From November 2017 to March 2018 there is first an impressive increase and then an impressive decrease of the same variable. Finally, from April 2018 to December 2018 the number of transactions has an irregular trend, but on average its values are lightly higher than the ones observed before November 2017. Based on these observations, in the following, we assume as pre-bubble period the time interval January - October 2017, as bubble period the time interval November 2017 - March

2018, and as post-bubble period the time interval April - December 2018. This analysis performs exactly what is expected by the function `Determine_Intervals()` of our algorithm.

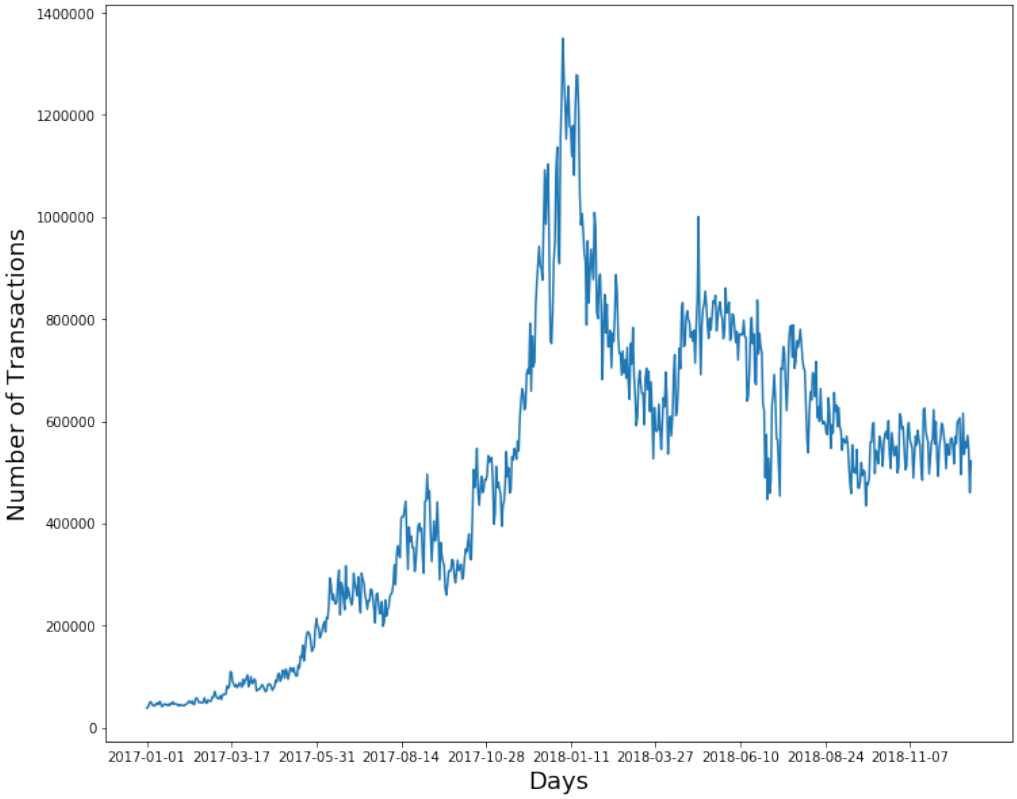

Figure 3: Number of transactions over time

The next analysis focuses on power addresses, i.e., those addresses that have made the most transactions. The analysis of these addresses is extremely relevant for two reasons. First, since the distributions of transactions against addresses follow a power law, the analysis of power addresses covers most of the phenomenon we want to examine. Second, since the number of power addresses is very small, compared to the total number of addresses, it is possible to make very precise and detailed analyses on them, which would be impossible to conduct on all addresses or on a very high fraction of them.

In particular, for each period (pre-bubble, bubble and post-bubble) and for each type of addresses (`from` and `to`), we decided to take the top 1000 addresses as the power ones. This activity corresponds to what is performed by the functions `Detect_From_Power_Addresses()` and `Detect_To_Power_Addresses()` of our algorithm. For each set thus selected, Table 3 shows: *(i)* what percentage of the total number of addresses operating in the reference period the top 1000 addresses correspond to; *(ii)* what percentage of the total number of transactions performed in the reference period the transactions carried out by the top 1000 addresses correspond to. From the analysis of this table, we can deduce that the previous conjectures on the opportunity to carry out the power address analyses were correct.

A first analysis of power addresses concerned the possible overlap between `from_addresses` and

Set	*Percentage of addresses*	*Percentage of transactions*
Pre-bubble, top 1000 `from_addresses`	0.01549%	89.81%
Bubble, top 1000 `from_addresses`	0.00599%	78.48%
Post-bubble, top 1000 `from_addresses`	0.00534%	77.87%
Pre-bubble, top 1000 `to_addresses`	0.01325%	86.02%
Bubble, top 1000 `to_addresses`	0.00495%	82.29%
Post-bubble, top 1000 `to_addresses`	0.00548%	86.34%

Table 3: Percentage of the addresses and transactions covered by each set of power addresses

`to_addresses`. For this purpose, for each period, we computed the intersection between the top 1000 `from_addresses` and the top 1000 `to_addresses`. This activity performs exactly what is expected by the function `Detect_Super_Power_Addresses()` of our algorithm. The cardinality of this intersection is equal to 173 (resp., 115, 81) for pre-bubble (resp., bubble, post-bubble) period. This says that only a small fraction of power addresses is simultaneously present in the top 1000 `from_addresses` and in the top 1000 `to_addresses` and this fraction significantly decreases in the transition from pre-bubble to bubble and from bubble to post-bubble periods.

A further analysis on power addresses led us to compute the possible intersections of the top 1000 addresses during the pre-bubble, bubble and post-bubble periods. This activity corresponds to what is performed by the functions `Detect_Multi_Interval_From_Power_Addresses()` and `Detect_Multi_Interval_To_Power_Addresses()` of our algorithm. The results obtained are reported in Table 4. Here, T^F_{Pre} (resp., T^F_B, T^F_{Post}) is the set of the top 1000 `from_addresses` during the pre-bubble (resp., bubble, post-bubble) period. Analogously, T^T_{Pre}, T^T_B and T^T_{Post} are the corresponding sets for `to_addresses`. From the analysis of this table we can see that:

Set	*Cardinality*
$\lvert T^F_{Pre} \cap T^F_B \rvert$	267
$\lvert T^F_B \cap T^F_{Post} \rvert$	268
$\lvert T^F_{Pre} \cap T^F_{Post} \rvert$	107
$\lvert T^T Pre \cap T^T_B \rvert$	288
$\lvert T^T_B \cap T^T_{Post} \rvert$	309
$\lvert T^T_{Pre} \cap T^T_{Post} \rvert$	114
$\lvert T^F_{Pre} \cap T^F_B \cap T^F_{Post} \rvert$	102
$\lvert T^T_{Pre} \cap T^T_B \cap T^T_{Post} \rvert$	112

Table 4: Cardinalities of the possible intersections of the top 1000 addresses during the pre-bubble, bubble and post-bubble periods

- The trends of `from_addresses` and `to_addresses` are very similar.
- The bubble has changed the power address scenario considerably. In fact, while the cardinality of the sets $\lvert T^F_{Pre} \cap T^F_B \rvert$, $\lvert T^F_B \cap T^F_{Post} \rvert$, $\lvert T^T Pre \cap T^T_B \rvert$ and $\lvert T^T_B \cap T^T_{Post} \rvert$ is quite large, the one of the sets $\lvert T^F_{Pre} \cap T^F_{Post} \rvert$ and $\lvert T^T_{Pre} \cap T^T_{Post} \rvert$ is much smaller. This tells us that, during the bubble period, most of the power addresses present in the pre-bubble period disappeared and new power addresses appeared; these last continued to exist during the post-bubble period. Finally, we observe that there are some power addresses, which we call "Survivors", that are present in the pre-bubble, bubble and post-bubble periods.

Based on the intersections introduced in Table 4, we can define three categories of addresses whose analysis appears extremely interesting for the extraction of knowledge on the bubble of Ethereum (and, presumably, of other cryptocurrencies). These categories are:

- *the Survivors*, which are the power addresses present in the pre-bubble, bubble and post-bubble periods;
- *the Missings*, which are the power addresses present in the pre-bubble period, but absent in the bubble and post-bubble ones;
- *the Entrants*, which are the power addresses absent in the pre-bubble period, but present in the bubble and post-bubble ones.

This activity performs exactly what is expected by the functions `Detect_Survivors()`, `Detect_Missings()` and `Detect_Entrants()` of our algorithm.

In the following of this paper, we aim at extracting knowledge patterns about these categories of addresses (and, ultimately, of users).

The next analysis aims at identifying how many power addresses are present in each category. We conducted this analysis for `from_addresses`, `to_addresses` and the intersection of these two sets. The results obtained are shown in Table 5.

Addresses	*Survivors*	*Entrants*	*Missings*
`from_addresses`	102	166	728
`to_addresses`	112	197	710
Intersection of `from_addresses` and `to_addresses`	21	17	114

Table 5: Number of power addresses belonging to the Survivors, Entrants and Missings categories

To fully understand the knowledge that can be extracted from this table, we must recall that: *(i)* the maximum number of power addresses for each category is equal to 1000; *(ii)* the Survivors, the Entrants and the Missings are obtained carrying out intersection operations. According to this reasoning, we can observe that the Survivors are very few; this result was expected because this category of addresses is obtained performing the intersection of three sets. The Entrants are also few while the Missings are many. This confirms that the bubble completely revolutionized the power address scenario in Ethereum, making the previous "main actors" (i.e., power addresses) disappear while introducing new ones.

Observe that, for all categories, the intersections between `from_addresses` and `to_addresses` are very small. This is totally in line with the result found previously saying that only a few addresses are `from_addresses` and `to_addresses` simultaneously.

4.3 Detecting the main features of the user categories of interest

Given a period (pre-bubble, bubble and post-bubble) and the set of the corresponding power addresses, we build a support social network. More specifically, let

$$\mathcal{N}_{Pre} = \langle NS_{Pre}, AS_{Pre} \rangle \qquad \mathcal{N}_{B} = \langle NS_{B}, AS_{B} \rangle \qquad \mathcal{N}_{Post} = \langle NS_{Post}, AS_{Post} \rangle$$

be the social networks associated with the pre-bubble, bubble and post-bubble periods.

NS_{Pre} (resp., NS_B, NS_{Post}) represents the set of the nodes of $\mathcal{N}_{Pre}$ (resp., $\mathcal{N}_B$, $\mathcal{N}_{Post}$). In this set, there is a node n_i for each power address. A label is associated with n_i; it allows us to specify if the corresponding address belongs to one of the categories of interest (Survivors, Entrants, Missings) or to none of them. Since there is a biunivocal correspondence between power addresses and nodes, in the following we will use these two terms interchangeably.

AS_{Pre} (resp., AS_B, AS_{Post}) denotes the set of the arcs of $\mathcal{N}_{Pre}$ (resp., $\mathcal{N}_B$, $\mathcal{N}_{Post}$). There is an arc $(n_i, n_j, TS_{ij}) \in AS_{Pre}$ (resp., AS_B, AS_{Post}) if there was at least one transaction from n_i to n_j. TS_{ij} represents the set of transactions from n_i to n_j made during the pre-bubble (resp., bubble, post-bubble) period. It consists of a set of pairs (t_{ij_k}, τ_{ij_k}), where t_{ij_k} represents the k^{th} transaction and τ_{ij_k} indicates the corresponding timestamp. The construction of these social networks corresponds to what is performed by the function `Construct_Social_Networks()` of our algorithm.

Having defined the support social networks, we can start our analyses on the address categories of interest. Below, we use the following notations: *(i)* $\mathcal{S}^F$ (resp., $\mathcal{S}^T$), to indicate the Survivors `from_addresses` (resp., `to_addresses`); *(ii)* $\mathcal{E}^F$ (resp., $\mathcal{E}^T$), to denote the Entrants `from_addresses` (resp., `to_addresses`); *(iii)* $\mathcal{M}^F$ (resp., $\mathcal{M}^T$), to represent the Missings `from_addresses` (resp., `to_addresses`).

In order to conduct our analyses on the address categories, we have considered the adoption of ego networks extremely useful. We recall that the ego network of a node n_i (called, precisely, "ego") consists of n_i, the nodes (called "alters") to which n_i is directly connected, the arcs connecting the ego to the alters and the arcs connecting the alters to each other. An ego network provides a clear indication of the relationships the corresponding ego is involved in, the nodes it interacts with, and the relationships existing between these last ones. In our analysis, which aims at detecting the features of each address category, ego network can play an important role because, due to the principle of homophily characterizing social networks [30], the features of a node are strongly influenced by the nodes belonging to its neighborhood. This activity corresponds exactly to what is expected by the functions `Construct_Survivors_Ego_Network_Pre()`, `Construct_Survivors_Ego_Network_Bubble()`, `Construct_Survivors_Ego_Network_Post()`, `Construct_Missings_Ego_Network_Pre()`, `Construct_Missings_Ego_Network_Bubble()`, `Construct_Missings_Ego_Network_Post()`, `Construct_Entrants_Ego_Network_Pre()`, `Construct_Entrants_Ego_Network_Bubble()`, and `Construct_Entrants_Ego_Network_Post()` of our algorithm.

As a first task, we computed the average number of nodes, the average number of arcs and the average density of the ego networks of the nodes belonging to each address category of interest. First, we examined the pre-bubble period. The results obtained are reported in the first three rows of Table 6.

From the analysis of these rows we can see that the ego networks of the Survivors nodes have an average number of nodes and arcs significantly higher than the ego networks of the nodes belonging to the other two categories. If such a result was expected for the Entrants (because the corresponding nodes were not power addresses during the pre-bubble period), it is instead surprising for the Missings. In fact, the latter, like the Survivors, were power addresses during the pre-bubble period. This allows us to conclude that having a very large ego-network during the pre-bubble period increases the possibility of remaining power addresses during the bubble and post-bubble periods. As far as density

Parameter	$\mathcal{S}^F$	$\mathcal{S}^T$	$\mathcal{M}^F$	$\mathcal{M}^T$	$\mathcal{E}^F$	$\mathcal{E}^T$
Average number of nodes	36,177.84	27,335.21	1,710.52	2,864.44	537.69	886.02
Average number of arcs	115,290.27	68,051.82	4,561.86	7,342.89	795.53	1,718.39
Average density	0.1120	0.0639	0.3852	0.2423	0.2125	0.1568
Average number of nodes	82,832.51	59,339.83	366.58	798.29	17,180.69	18,945.69
Average number of arcs	325,179.44	172,713.37	587.84	2563.00	59,733.11	67,956.61
Average density	0.074	0.019	0.401	0.282	0.211	0.031
Average number of nodes	47,237.20	46,661.02	162.10	572.93	19,686.75	22,373.64
Average number of arcs	174,537.78	148,359.25	425.70	1,360.52	93,099.84	70,518.77
Average density	0.1045	0.039	0.411	0.233	0.178	0.0157

Table 6: Average number of nodes, average number of arcs and average density of the ego networks of the nodes belonging to the address categories of interest

is concerned, there are no particular observations to make taking into account that the low density of Survivor's ego networks can be explained simply by the large number of nodes characterizing them.

After this, we examined the bubble period. The results obtained are reported in the next three rows of Table 6. From the analysis of these rows we can observe that both the Survivors and the Entrants have much larger ego networks than the Missings. Actually, this result was expected since, in the bubble period, the nodes belonging to the Survivors and the Entrants are power addresses. Instead, it is unexpected that the Survivors have much larger ego networks than the Entrants. In fact, the addresses of both categories are power addresses during the bubble period. However, it seems that the Survivors tend to include the strongest power addresses. Note also that the one of the Survivors' ego networks during the bubble period is about twice the size of the Survivors' ego networks during the pre-bubble period. Also, the Survivors' ego networks have by far the largest size during the bubble period. This allows us to conclude that it is exactly the activity of the Survivors that could have caused the bubble; this activity has led to the exit of the Missings from the power addresses and to the arrival of the Entrants among them. However, these last ones enter into the power addresses "on tiptoe"; in fact, they are not the ones who dictate the line and cause the bubble; this task is carried out by the Survivors.

Finally, we considered the post-bubble period. The results obtained are reported in the last three rows of Table 6. The analysis of this table confirms the trends we observed for the bubble period. This is not surprising because also during the post-bubble period both the Survivors and the Entrants are power addresses. Note that, during this period, the size of the Survivors' ego networks is much smaller than the one of the Survivors' ego networks during the bubble period, although it is slightly larger than the size of the Survivors' ego networks during the pre-bubble period. This trend perfectly reflects the one of the number of transactions reported in Figure 3. This is a further confirmation that the trend shown by Ethereum in the years 2017 and 2018, which led to a bubble, was mainly caused by the Survivors. We note that the size of the Entrants' ego network during the post-bubble period shows a slight growth compared to the bubble period. This is an indication that, during the post-bubble period, the Entrants consolidate their presence among the power addresses, even though they are not dictating the line yet: this is still a responsibility of the Survivors. All the activities described above correspond to what is performed by the functions `Analyze_Ego_Pre()`, `Analyze_Ego_Bubble()` and `Analyze_Ego_Post()` of our algorithm.

The analysis of Table 6, along with the previous reasoning, indicates that having very large ego

networks seems to be an intrinsic feature of the Survivors, regardless of the pre-bubble, bubble or post-bubble period.

4.4 Evaluating the existence of backbones linking users of a certain category

The ego networks introduced previously represent a considerable tool to also estimate the possible existence of backbones linking addresses of the same category. In fact, a way to do this consists in verifying, given an address category, the fraction of the corresponding ego networks having, among the alters, at least k addresses belonging to it. Clearly, the higher the value of k and the fraction of the ego networks satisfying this property, the stronger the hypothesis that a backbone exists among the addresses of the category into examination.

To better clarify this idea, let us consider Table 7 that refers to the Survivors' ego networks during the pre-bubble period. In the left part of this table, we examine the set $\mathcal{S}^F$ of the Survivors `from_addresses`. The fifth row of this table tells us that 19.6% of the ego networks of the nodes of $\mathcal{S}^F$ contains at least 5 nodes of $\mathcal{S}^F$ among the alters. This percentage decreases to 0.9% if we consider the presence of at least 5 nodes of $\mathcal{E}^F$ and increases to 33.3% if we take into account the presence of at least 5 nodes of $\mathcal{M}^F$.

	Ego networks of $\mathcal{S}^F$			*Ego networks of* $\mathcal{S}^T$		
	Nodes of $\mathcal{S}^F$	*Nodes of* $\mathcal{E}^F$	*Nodes of* $\mathcal{M}^F$	*Nodes of* $\mathcal{S}^T$	*Nodes of* $\mathcal{E}^T$	*Nodes of* $\mathcal{M}^T$
$k = 1$	0.755	0.088	0.676	0.580	0.223	0.696
$k = 2$	0.512	0.058	0.529	0.339	0.071	0.509
$k = 3$	0.392	0.049	0.402	0.169	0.0	0.348
$k = 4$	0.294	0.019	0.353	0.098	0.0	0.304
$k = 5$	0.196	0.009	0.333	0.080	0.0	0.277
$k = 6$	0.147	0.0	0.284	0.062	0.0	0.268
$k = 7$	0.118	0.0	0.265	0.053	0.0	0.241
$k = 8$	0.078	0.0	0.235	0.036	0.0	0.196
$k = 9$	0.078	0.0	0.216	0.027	0.0	0.196

Table 7: Analysis of the presence of backbones linking the Survivors during the pre-bubble period

Once we have clarified the kind of information we want to look for, let us consider Table 7, which concerns the Survivors' ego networks during the pre-bubble period. From the analysis of this table we can see that many of the ego-networks of $\mathcal{S}^F$ (resp., $\mathcal{S}^T$) have, among their alters, several nodes belonging to $\mathcal{S}^F$ (resp., $\mathcal{S}^T$), along with several nodes belonging to $\mathcal{M}^F$ (resp., $\mathcal{M}^T$). Instead, the number of ego networks of $\mathcal{S}^F$ (resp., $\mathcal{S}^T$) having one or more nodes of $\mathcal{E}^F$ (resp., $\mathcal{E}^T$) among the alters is very small. This allows us to assume that there is a backbone linking the nodes of $\mathcal{S}^F$ (resp., $\mathcal{S}^T$). The presence of many nodes of $\mathcal{M}^F$ (resp., $\mathcal{M}^T$) among the alters of the ego networks of $\mathcal{S}^F$ (resp., $\mathcal{S}^T$) is not surprising because, during the pre-bubble period, the nodes of $\mathcal{M}^F$ (resp., $\mathcal{M}^T$) were power addresses. Finally, we observe that the presence of Survivors and Missings nodes among the alters of the ego networks of Survivors nodes is more marked for `from_addresses` than for `to_addresses`, as we can see comparing the first three and the last three columns of Table 7. This activity corresponds to what is performed by the function `Detect_Backbones_Survivor_Pre()` of our algorithm.

Consider, now, Table 8 that refers to the Missings' ego networks during the pre-bubble period. The structure and the semantics of this table are analogous to the ones of Table 7. From the analysis of this table, we can observe that many ego networks of $\mathcal{M}^F$ (resp., $\mathcal{M}^T$) have one or two nodes of $\mathcal{M}^F$ (resp., $\mathcal{M}^T$) or of $\mathcal{S}^F$ (resp., $\mathcal{S}^T$) among their alters. However, compared to the case of the Survivors,

reported in Table 7, this phenomenon is much smaller both as fraction of ego-networks and as value of k. Therefore, we can conclude that there is a backbone also among the nodes of $\mathcal{M}^F$ (resp., $\mathcal{M}^T$), although this is less strong than the one observed for the nodes of $\mathcal{S}^F$ (resp., $\mathcal{S}^T$). The presence of many nodes of $\mathcal{S}^F$ (resp., $\mathcal{S}^T$) among the alters of the ego networks of $\mathcal{M}^F$ (resp., $\mathcal{M}^T$) is justified by the fact that both these categories of nodes were power addresses during the pre-bubble period. The difference between `from_addresses` and `to_addresses` in the Missings' ego networks is much smaller than the one observed in the Survivors' ego networks. This task performs exactly what is expected by the function `Detect_Backbones_Missings_Pre()` of our algorithm.

	Ego networks of $\mathcal{M}^F$			*Ego networks of* $\mathcal{M}^T$		
	Nodes of $\mathcal{S}^F$	*Nodes of* $\mathcal{E}^F$	*Nodes of* $\mathcal{M}^F$	*Nodes of* $\mathcal{S}^T$	*Nodes of* $\mathcal{E}^T$	*Nodes of* $\mathcal{M}^T$
$k=1$	0.466	0.010	0.497	0.390	0.024	0.406
$k=2$	0.277	0.0	0.214	0.162	0.0	0.225
$k=3$	0.165	0.0	0.115	0.093	0.0	0.138
$k=4$	0.098	0.0	0.070	0.056	0.0	0.089
$k=5$	0.059	0.0	0.049	0.039	0.0	0.068
$k=6$	0.040	0.0	0.033	0.031	0.0	0.052
$k=7$	0.018	0.0	0.029	0.025	0.0	0.037
$k=8$	0.004	0.0	0.027	0.021	0.0	0.032
$k=9$	0.004	0.0	0.027	0.018	0.0	0.028

Table 8: Analysis of the presence of backbones linking the Missings during the pre-bubble period

Now, we conduct the same analysis for the Entrants' ego networks. The results obtained are shown in Table 9. The structure and the semantics of this table are similar to the ones of Tables 7 and 8. From the analysis of Table 9 we can conclude that there is no backbone linking the Entrants during the pre-bubble period. This result is justified considering that, during this period, the Entrants were not power addresses. The presence of some nodes of the Survivors or of the Missings in the alters of the Entrants is simply due to the fact that the Survivors and the Missings were power addresses during the pre-bubble period. This activity corresponds to what is performed by the function `Detect_Backbones_Entrants_Pre()` of our algorithm.

	Ego networks of $\mathcal{E}^F$			*Ego networks of* $\mathcal{E}^T$		
	Nodes of $\mathcal{S}^F$	*Nodes of* $\mathcal{E}^F$	*Nodes of* $\mathcal{M}^F$	*Nodes of* $\mathcal{S}^T$	*Nodes of* $\mathcal{E}^T$	*Nodes of* $\mathcal{M}^T$
$k=1$	0.326	0.140	0.163	0.194	0.0	0.222
$k=2$	0.140	0.0	0.023	0.083	0.0	0.056
$k=3$	0.070	0.0	0.0	0.056	0.0	0.056
$k=4$	0.0	0.0	0.0	0.056	0.0	0.056
$k=5$	0.0	0.0	0.0	0.056	0.0	0.028
$k=6$	0.0	0.0	0.0	0.056	0.0	0.028
$k=7$	0.0	0.0	0.0	0.056	0.0	0.028
$k=8$	0.0	0.0	0.0	0.056	0.0	0.028
$k=9$	0.0	0.0	0.0	0.056	0.0	0.028

Table 9: Analysis of the presence of backbones linking the Entrants during the pre-bubble period

To also give a graphical idea of the results on the presence of backbones obtained above, we consider a social network $\mathcal{N}^F_{Pre}$, obtained from $\mathcal{N}_{Pre}$ considering only the power `from_addresses`.

In order to extract a subnet of $\mathcal{N}^F_{Pre}$ containing nodes strongly connected to each other, we should consider the cliques of $\mathcal{N}^F_{Pre}$. However, since clique computation is an NP-hard problem, we decided to adopt a relaxation of the clique concept and focused on k-core. We recall that a k-core of a network $\mathcal{N}$ is a connected maximal induced subnetwork of $\mathcal{N}$ in which all nodes have degree at least k. A k-core

can be used as an indicator of the presence of backbones. In fact, if some nodes, say $n_1, n_2, \ldots, n_q$, belong to a k-core, then each of them will be connected to at least k of the other ones.

Consider the 5-core of $\mathcal{N}^F_{Pre}$ shown in Figure 4. In it, we indicate in yellow the Survivors nodes, in red the Missings nodes and in blue all the other ones. The 5-core consists of 175 nodes. As we can see from the figure, there is a strong backbone connecting 32 Survivors nodes and another weaker backbone connecting 13 Missings nodes. Consider, now, the 7-core of $\mathcal{N}^F_{Pre}$ shown in Figure 5. It contains even more strongly connected nodes than the 5-core. The total number of its nodes is 86. Again, there is a strong backbone connecting 19 Survivors nodes and a weaker backbone connecting 5 Missings nodes. Both these figures provide a graphical idea of the analytical results found previously.

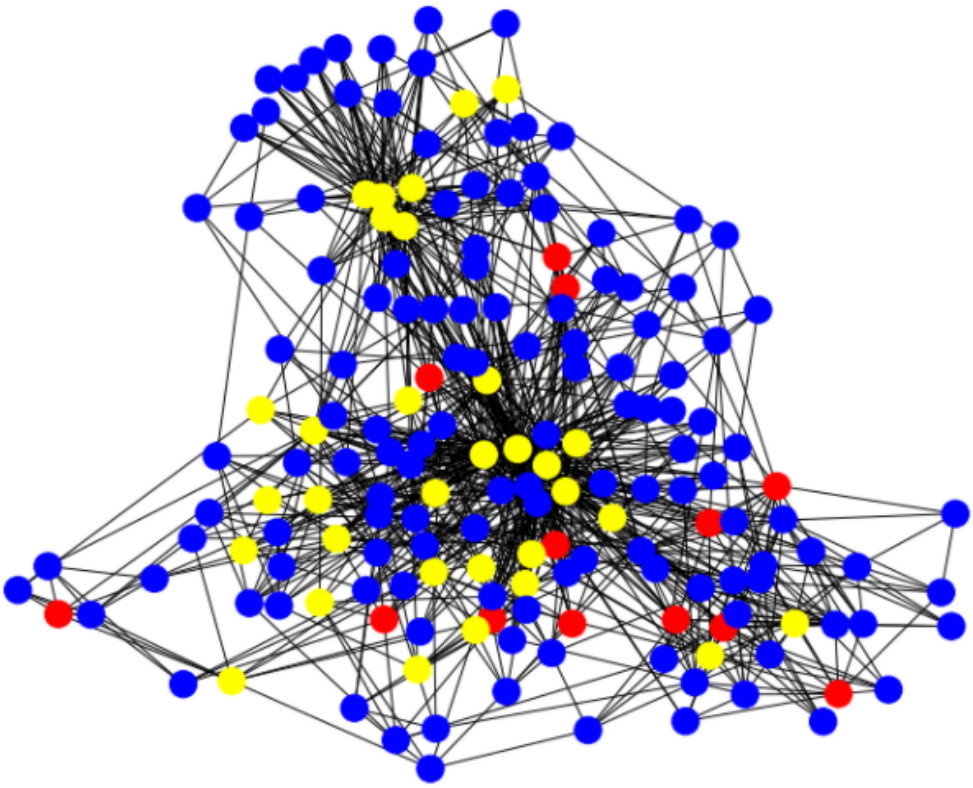

Figure 4: A 5-core of $\mathcal{N}^F_{Pre}$

The next analysis concerns the Survivors', the Missings' and the Entrants' ego networks during the bubble period. The results obtained by carrying out the same tasks seen for the pre-bubble period are reported in Tables 10, 11 and 12. These activities correspond to what is performed by the functions `Detect_Backbones_Survivors_Bubble()`, `Detect_Backbone_Missings_Bubble()` and `Detect_Backbones_Entrants_Bubbles()` of our algorithm.

From the analysis of these tables we can detect the following knowledge patterns: *(1)* There is a very strong backbone linking the Survivors, as can be seen by examining Table 10. *(2)* In the same table, we can observe that there are some Entrants and Missings nodes among the alters of the Survivors' ego networks. This can be explained taking into account that the Entrants are power

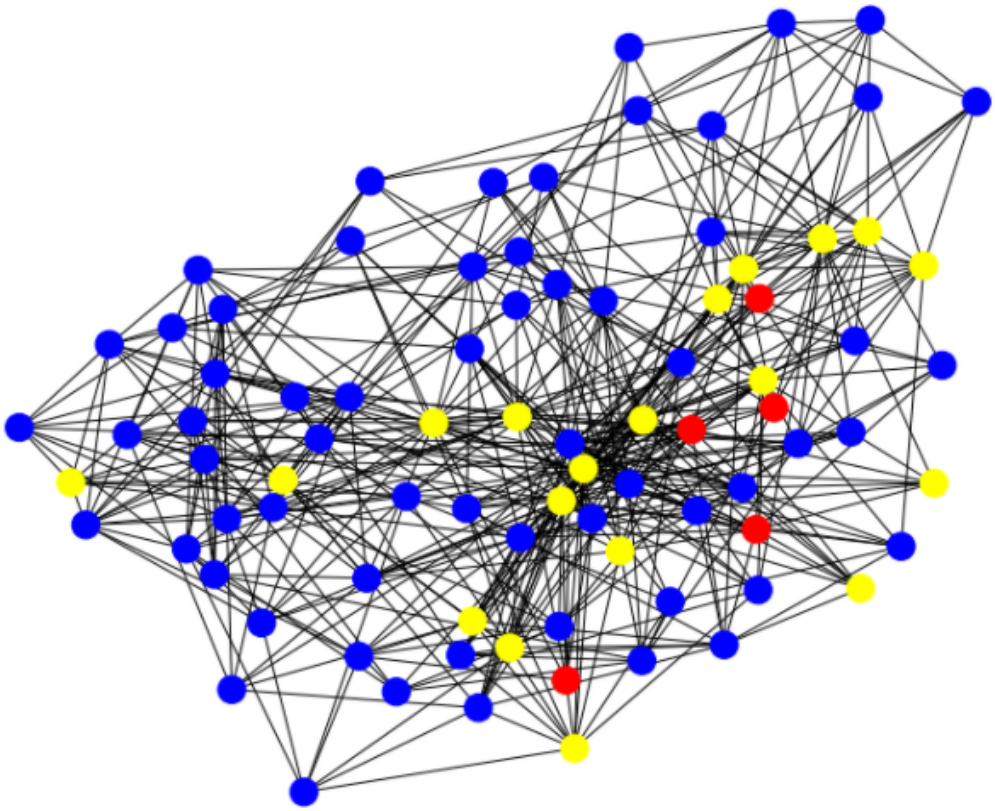

Figure 5: A 7-core of $\mathcal{N}^F_{Pre}$

	Ego networks of $\mathcal{S}^F$			*Ego networks of* $\mathcal{S}^T$		
	Nodes of $\mathcal{S}^F$	*Nodes of* $\mathcal{E}^F$	*Nodes of* $\mathcal{M}^F$	*Nodes of* $\mathcal{S}^T$	*Nodes of* $\mathcal{E}^T$	*Nodes of* $\mathcal{M}^T$
$k=1$	0.824	0.451	0.461	0.750	0.688	0.714
$k=2$	0.598	0.245	0.333	0.554	0.509	0.491
$k=3$	0.431	0.167	0.284	0.312	0.357	0.339
$k=4$	0.373	0.127	0.265	0.143	0.223	0.232
$k=5$	0.304	0.078	0.225	0.098	0.152	0.161
$k=6$	0.265	0.069	0.216	0.071	0.062	0.134
$k=7$	0.196	0.029	0.147	0.036	0.054	0.098
$k=8$	0.147	0.020	0.137	0.027	0.045	0.089
$k=9$	0.108	0.020	0.118	0.027	0.036	0.089

Table 10: Analysis of the presence of backbones linking the Survivors during the bubble period

	Ego networks of $\mathcal{M}^F$			*Ego networks of* $\mathcal{M}^T$		
	Nodes of $\mathcal{S}^F$	*Nodes of* $\mathcal{E}^F$	*Nodes of* $\mathcal{M}^F$	*Nodes of* $\mathcal{S}^T$	*Nodes of* $\mathcal{E}^T$	*Nodes of* $\mathcal{M}^T$
$k=1$	0.338	0.125	0.138	0.283	0.166	0.217
$k=2$	0.163	0.054	0.023	0.095	0.034	0.049
$k=3$	0.111	0.035	0.006	0.042	0.014	0.026
$k=4$	0.065	0.021	0.004	0.026	0.010	0.014
$k=5$	0.044	0.015	0.002	0.020	0.008	0.012
$k=6$	0.021	0.013	0.0	0.020	0.006	0.010
$k=7$	0.019	0.010	0.0	0.016	0.004	0.008
$k=8$	0.010	0.004	0.0	0.010	0.004	0.006
$k=9$	0.006	0.002	0.0	0.010	0.004	0.006

Table 11: Analysis of the presence of backbones linking the Missings during the bubble period

	Ego networks of $\mathcal{E}^F$			*Ego networks of* $\mathcal{E}^T$		
	Nodes of $\mathcal{S}^F$	*Nodes of* $\mathcal{E}^F$	*Nodes of* $\mathcal{M}^F$	*Nodes of* $\mathcal{S}^T$	*Nodes of* $\mathcal{E}^T$	*Nodes of* $\mathcal{M}^T$
$k = 1$	0.337	0.572	0.217	0.335	0.477	0.335
$k = 2$	0.175	0.295	0.127	0.152	0.284	0.152
$k = 3$	0.096	0.169	0.084	0.081	0.142	0.081
$k = 4$	0.066	0.096	0.054	0.061	0.076	0.051
$k = 5$	0.048	0.066	0.042	0.061	0.046	0.030
$k = 6$	0.036	0.030	0.036	0.056	0.030	0.025
$k = 7$	0.024	0.024	0.036	0.046	0.030	0.020
$k = 8$	0.024	0.0	0.036	0.041	0.025	0.015
$k = 9$	0.024	0.0	0.036	0.036	0.025	0.015

Table 12: Analysis of the presence of backbones linking the Entrants during the bubble period

addresses during the bubble period, while the Missings, although not anymore, were power addresses in the period immediately before. *(3)* Table 11 shows that there is no longer a backbone linking the Missings. *(4)* Table 12 reveals that a backbone linking the Entrants starts to exist, even if it is not very strong yet.

To also give a graphical idea of these results, we consider the $\mathcal{N}_B^F$ network. It is defined similarly to N_{Pre}^F, but taking the bubble period into account. We also consider the corresponding 5-core and 7-core, shown in Figures 6 and 7, respectively. In them, we represent the Survivors nodes in yellow, the Entrants nodes in green and all the other nodes in blue. The 5-core consists of 149 nodes. Here, there is a very strong backbone involving 47 Survivors nodes and a weaker one involving 17 Entrants nodes. The 7-core consists of 67 nodes. Also in this case there is a very strong backbone connecting 30 Survivors nodes and a weaker backbone connecting 13 Entrants nodes.

The last analysis concerns the Survivors', the Missings' and the Entrants' ego networks during the post-bubble period. The results obtained are reported in Tables 13, 14 and 15. These activities perform exactly what is expected by the functions `Detect_Backbones_Survivors_Post()`, `Detect_Backbones_Missings_Post()` and `Detect_Backbones_Entrants_Post()` of our algorithm.

	Ego networks of $\mathcal{S}^F$			*Ego networks of* $\mathcal{S}^T$		
	Nodes of $\mathcal{S}^F$	*Nodes of* $\mathcal{E}^F$	*Nodes of* $\mathcal{M}^F$	*Nodes of* $\mathcal{S}^T$	*Nodes of* $\mathcal{E}^T$	*Nodes of* $\mathcal{M}^T$
$k = 1$	0.716	0.490	0.353	0.741	0.768	0.518
$k = 2$	0.510	0.265	0.206	0.607	0.598	0.330
$k = 3$	0.363	0.167	0.167	0.384	0.446	0.188
$k = 4$	0.265	0.147	0.108	0.223	0.366	0.143
$k = 5$	0.216	0.137	0.088	0.116	0.268	0.089
$k = 6$	0.186	0.098	0.078	0.080	0.223	0.089
$k = 7$	0.108	0.069	0.059	0.062	0.134	0.080
$k = 8$	0.088	0.059	0.049	0.045	0.098	0.062
$k = 9$	0.059	0.039	0.039	0.045	0.062	0.045

Table 13: Analysis of the presence of backbones linking the Survivors during the post-bubble period

From the analysis of these tables we can deduce the following knowledge patterns: *(1)* There is a strong backbone linking the Survivors, as can be seen in Table 13. Comparing Tables 10 and 13 we can see that this backbone, while continuing to remain strong, un dergoes a weakening, compared to the pre-bubble period. This is physiological because, during the post-bubble period, the number of transactions made decreased considerably with respect to the ones of the bubble period. *(2)* We continue to observe the presence of some Entrants and Missings nodes among the alters of the Survivors' ego networks. The reasons for this fact are the same as those seen for the bubble period. *(3)* The backbone linking the Missings, which had already started to disappear during the bubble period,

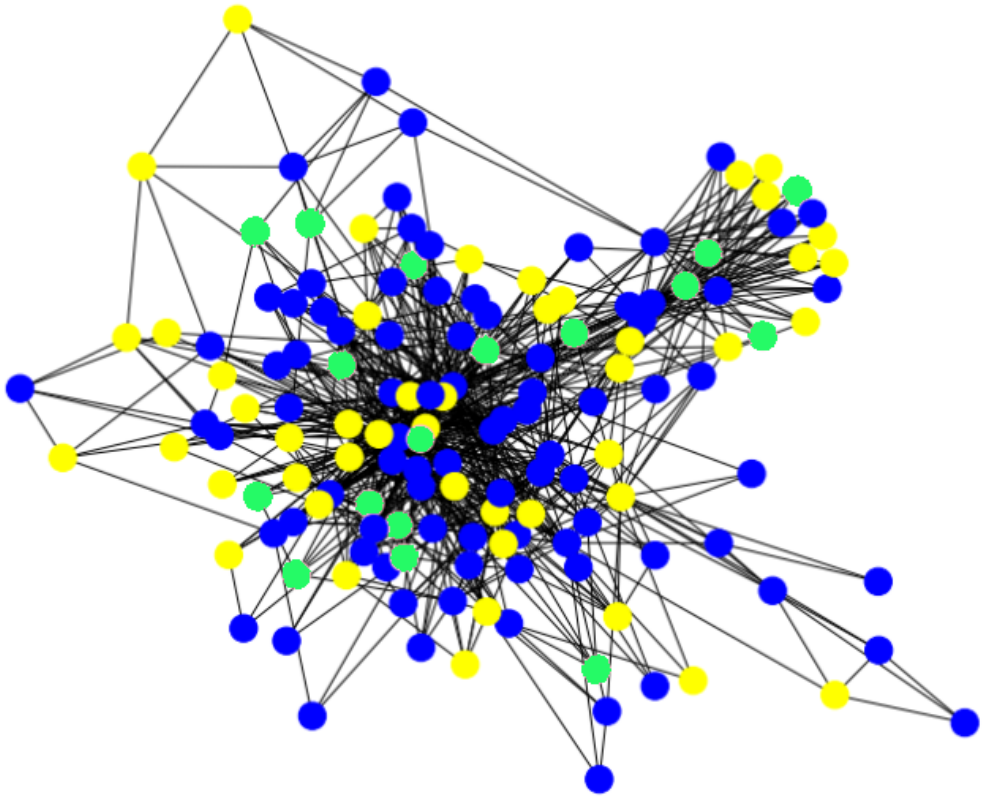

Figure 6: A 5-core of $\mathcal{N}_B^F$

	Ego networks of $\mathcal{M}^F$			*Ego networks of* $\mathcal{M}^T$		
	Nodes of $\mathcal{S}^F$	*Nodes of* $\mathcal{E}^F$	*Nodes of* $\mathcal{M}^F$	*Nodes of* $\mathcal{S}^T$	*Nodes of* $\mathcal{E}^T$	*Nodes of* $\mathcal{M}^T$
$k=1$	0.263	0.193	0.119	0.274	0.167	0.070
$k=2$	0.122	0.126	0.015	0.067	0.040	0.027
$k=3$	0.056	0.081	0.007	0.032	0.019	0.013
$k=4$	0.033	0.059	0.007	0.027	0.011	0.008
$k=5$	0.026	0.052	0.004	0.019	0.011	0.008
$k=6$	0.015	0.041	0.004	0.016	0.008	0.005
$k=7$	0.011	0.033	0.004	0.013	0.005	0.003
$k=8$	0.011	0.022	0.0	0.011	0.005	0.0
$k=9$	0.007	0.011	0.0	0.008	0.005	0.0

Table 14: Analysis of the presence of backbones linking the Missings during the post-bubble period

	Ego networks of $\mathcal{E}^F$			*Ego networks of* $\mathcal{E}^T$		
	Nodes of $\mathcal{S}^F$	*Nodes of* $\mathcal{E}^F$	*Nodes of* $\mathcal{M}^F$	*Nodes of* $\mathcal{S}^T$	*Nodes of* $\mathcal{E}^T$	*Nodes of* $\mathcal{M}^T$
$k=1$	0.331	0.651	0.211	0.431	0.675	0.376
$k=2$	0.187	0.380	0.133	0.223	0.457	0.096
$k=3$	0.133	0.193	0.084	0.091	0.310	0.036
$k=4$	0.090	0.108	0.048	0.076	0.198	0.020
$k=5$	0.054	0.078	0.048	0.071	0.122	0.015
$k=6$	0.036	0.066	0.048	0.061	0.086	0.015
$k=7$	0.036	0.042	0.048	0.061	0.056	0.015
$k=8$	0.030	0.018	0.048	0.056	0.051	0.015
$k=9$	0.024	0.018	0.042	0.056	0.046	0.010

Table 15: Analysis of the presence of backbones linking the Entrants during the post-bubble period

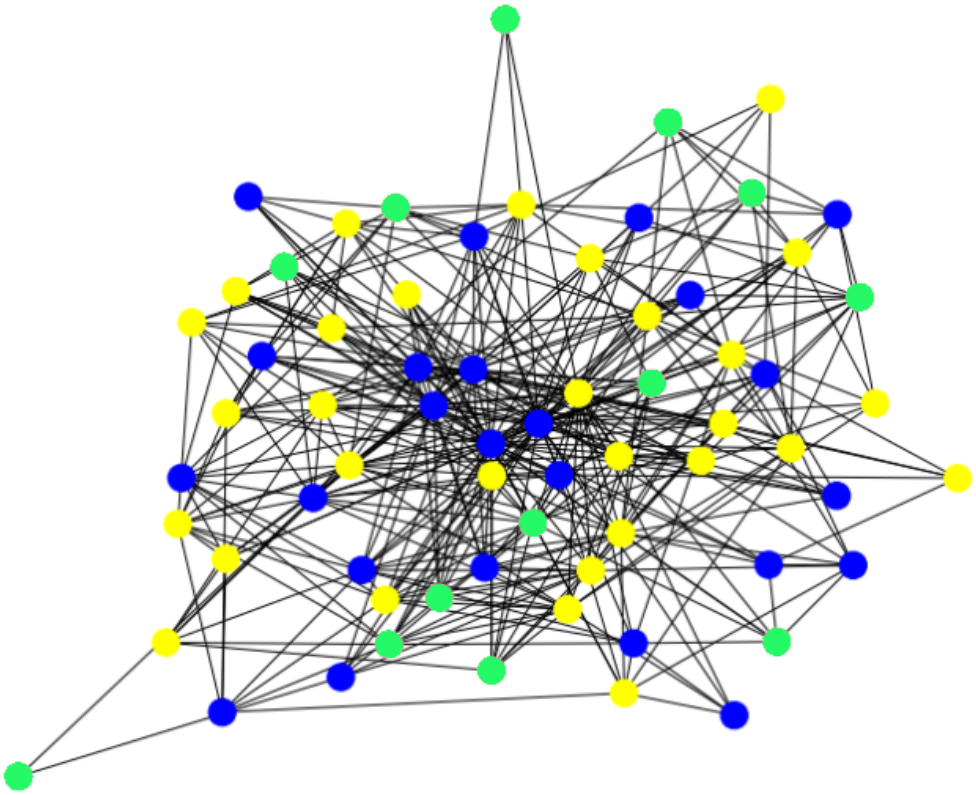

Figure 7: A 7-core of $\mathcal{N}_B^F$

has completely dissolved, as evidenced by the further decrease of the values in the fourth and seventh columns of Table 14, compared to the corresponding ones of Table 11. *(4)* The backbone linking the Entrants, which was already visible during the bubble period, is further consolidated during the post-bubble period, as can be seen by examining Table 15.

Also in this case we can use k-cores to give a graphical idea of the results obtained. For this purpose, we consider the network $\mathcal{N}_{Post}^F$, obtained similarly to $\mathcal{N}_{Pre}^F$ and $\mathcal{N}_B^F$. We also consider the corresponding 5-core and 7-core. Due to space limitations, we do not show them. In any case, we found that the 5-core consists of 202 nodes. Here, there is a strong backbone linking 42 Survivors and another one linking 31 Entrants. Note that, compared to the bubble period, the backbone linking the Entrants has strengthened. A similar reasoning also applies to the 7-core. It consists of 111 nodes. In it, we can observe a strong backbone linking 24 Survivors and another one linking 16 Entrants. Also this last backbone appears strengthened compared to the corresponding one relative to the 7-core during the bubble period, shown in Figure 7. All these graphical results are totally in line with the analytical ones relative to the post-bubble period presented above.

4.4.1 Graphical backbone evaluations through k-trusses

Till now, we have said that, in order to verify the possible existence of backbones among the Survivors, the Missings or the Entrants, the concept of clique could be used. We have also said that the computation of cliques was a NP-hard problem and, for this reason, we chose to replace cliques with k-cores. In fact, the k-core concept is a relaxation of the clique one and, unlike cliques, the computation of k-cores can be done in polynomial time. However, it is worth checking that the results obtained with k-core are not unduly influenced by the properties of this structure. One way to carry out this verification is to repeat the experiments performed with k-cores using another data structure that can be considered a relaxation of the clique concept and can be computed in polynomial time. To this end, we focused on the concept of k-truss [10]. A k-truss is a non-trivial, one component subgraph such that each edge is reinforced by at least $k-2$ pairs of edges making a triangle with it. Observe that each clique of order k is contained in a k-truss, whereas a k-truss does not necessarily contain a clique of order k. Furthermore, each k-truss is a subgraph of a (k-1)-core. All these properties support the idea that a k-truss is a concept that lies between the clique concept, which is too restrictive, and the k-core one, which is too lax.

After choosing the k-truss as the structure to use, similarly to what we did for k-cores, we computed the 5-truss of $\mathcal{N}^F_{Pre}$ and we saw that: *(i)* it consists of 152 nodes; *(ii)* there is a strong backbone connecting 27 Survivors; *(iii)* there is a weaker backbone connecting 7 Missings. Then, we computed the 7-truss of $\mathcal{N}^F_{Pre}$ and we obtained that: *(i)* it consists of 74 nodes; *(ii)* there is a strong backbone connecting 16 Survivors; *(iii)* there is no significant backbone among the Missings.

Proceeding with our investigation, we computed the 5-truss of $\mathcal{N}^F_B$; analyzing it, we obtained that: *(i)* it consists of 134 nodes; *(ii)* there is a very strong backbone involving 41 Survivors; *(iii)* there is an additional backbone involving 15 Entrants. The analysis of the 7-truss of $\mathcal{N}^F_B$ allows us to say that: *(i)* it consists of 61 nodes; *(ii)* there is a very strong backbone involving 26 Survivors; *(iii)* there is a weaker backbone involving 10 Entrants.

Our analysis on k-trusses ends with the computation of the 5-truss and 7-truss of $\mathcal{N}^F_{Post}$. Regarding the former, we obtained that: *(i)* it consists of 194 nodes; *(ii)* there is a strong backbone connecting 36 Survivors; *(iii)* there is an additional backbone connecting 26 Entrants. Regarding the latter, we saw that: *(i)* it consists of 96 nodes; *(ii)* there is a strong backbone connecting 22 Survivors; *(iii)* there is an additional backbone connecting 12 Entrants.

Comparing the results obtained through the k-truss analysis with those regarding the k-core one shown above, we can observe that they are similar. In fact, the k-truss analysis confirms everything we found through the k-core one. The only exception regards the fact that the k-core analysis detected a backbone (albeit a very weak one) between the Missings in the 7-core associated with $\mathcal{N}^F_{Pre}$. Such a backbone was not detected in the corresponding 7-truss analysis. However, this minimal difference can be explained considering that: *(i)* the detected backbone of the 7-core analysis is anyway very weak; *(ii)* the concept of k-truss is more "severe" than the one of k-core.

At the end of this analysis, we can conclude that the strong similarity of the results obtained using k-cores and k-trusses allows us to say that the information found is intrinsic in the data and is not unduly generated by the properties of k-cores.

4.5 Predicting the characteristics of the main future actors

All the previous analyses are mainly descriptive and diagnostic. In this section, instead, we want to go one step further proposing a predictive analysis with the aim of understanding, during a period (specifically, pre-bubble, bubble), what are the features of the addresses that will probably play a leading role during the next period (specifically, bubble, post-bubble). The importance of this analysis (in itself already evident) is reinforced by the results obtained in the previous section, telling us that these main actors are often connected by backbones. Consequently, identifying (and possibly acting on) some of them gives the possibility to identify (and act on) most of the others connected through the backbones.

In Table 16, we show the number of transactions, the number of contacts and the average value of transactions for the following addresses: *(1)* T^F_{Pre}: the power `from_addresses` in the pre-bubble period. *(2)* $\mathcal{S}^F$: the Survivors `from_addresses`. By definition, each element of $\mathcal{S}^F$ must also be an element of T^F_{Pre} and an element of T^F_B, i.e., the power `from_addresses` in the bubble period. *(3)* $\mathcal{M}^F$: the Missings `from_addresses`. By definition, each element of $\mathcal{M}^F$ must also be an element of T^F_{Pre}, while it cannot belong to T^F_B. *(4)* $\mathcal{E}^F_{Pre}$: the `from_addresses` that appeared in the bubble period but were already present (albeit not as power addresses) in the pre-bubble period. By definition, each element of $\mathcal{E}^F_{Pre}$ must also be an element of T^F_B, while it cannot belong to T^F_{Pre}.

	T^F_{Pre}	$\mathcal{S}^F$	$\mathcal{M}^F$	$\mathcal{E}^F_{Pre}$
Average Number of Transactions	30,346.55	175,729.30	11,064.18	473.83
Average Number of Contacts	4,817.39	27,088.52	1,259.26	242.98
Average Value of Transactions (Eth)	8.65	8.18	7.32	106.53

Table 16: Average number of transactions, average number of contacts and average values of transactions for T^F_{Pre}, $\mathcal{S}^F$, $\mathcal{M}^F$ and $\mathcal{E}^F_{Pre}$

From the analysis of this table we can see that the addresses of $\mathcal{S}^F$ have a significantly higher number of transactions and contacts than the corresponding ones not only of $\mathcal{M}^F$ and $\mathcal{E}^F_{Pre}$ but also of T^F_{Pre}. Instead, the average value of transactions is smaller for $\mathcal{S}^F$, $\mathcal{M}^F$ and T^F_{Pre} than for $\mathcal{E}^F_{Pre}$.

This result is even more evident considering Figure 8 (resp., 9). Here, we show the distribution of the addresses of $\mathcal{S}^F$ and $\mathcal{M}^F$ against the number of transactions (resp., contacts) of T^F_{Pre}. The abscissae axis is divided into deciles. In the figure, we indicate the decile with the highest values with D_{10} and the one with the lowest value with D_1. Figure 8 shows that most of the addresses of $\mathcal{S}^F$ belong to the highest deciles of T^F_{Pre}. This does not happen for the addresses of $\mathcal{M}^F$ that show a rather uniform distribution among the deciles of T^F_{Pre}, except for the lowest decile where they are almost absent. Figure 9 shows a similar trend except for the lowest decile, which comprises a lot of addresses for both $\mathcal{S}^F$ and $\mathcal{M}^F$. The activities described above correspond to what is performed by the function `Detect_Top_Power_Addresses()` of our algorithm.

Both Table 16 and Figures 8 and 9 give us the same important following indication: "The addresses that will survive a bubble are to be searched among the ones that, in the pre-bubble period, have carried out the highest numbers of transactions and have the highest numbers of contacts". This indication is very strong for the number of transactions while it is a bit weaker for the number of contacts. In fact, as for this last parameter, we can see that the lowest decile contains a certain number not only

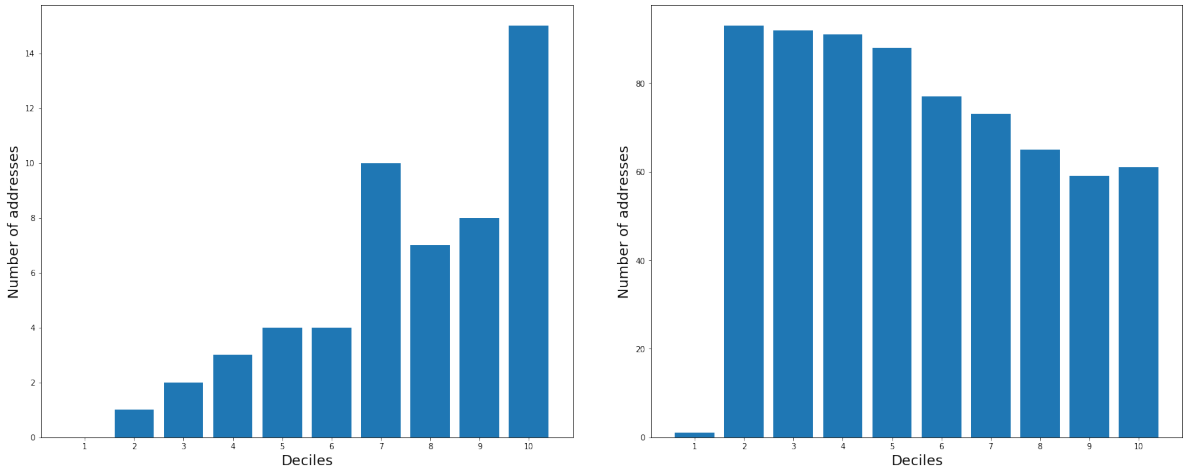

Figure 8: Distribution of the addresses of $\mathcal{S}^F$ (at left) and $\mathcal{M}^F$ (at right) against the number of transactions of T^F_{Pre}

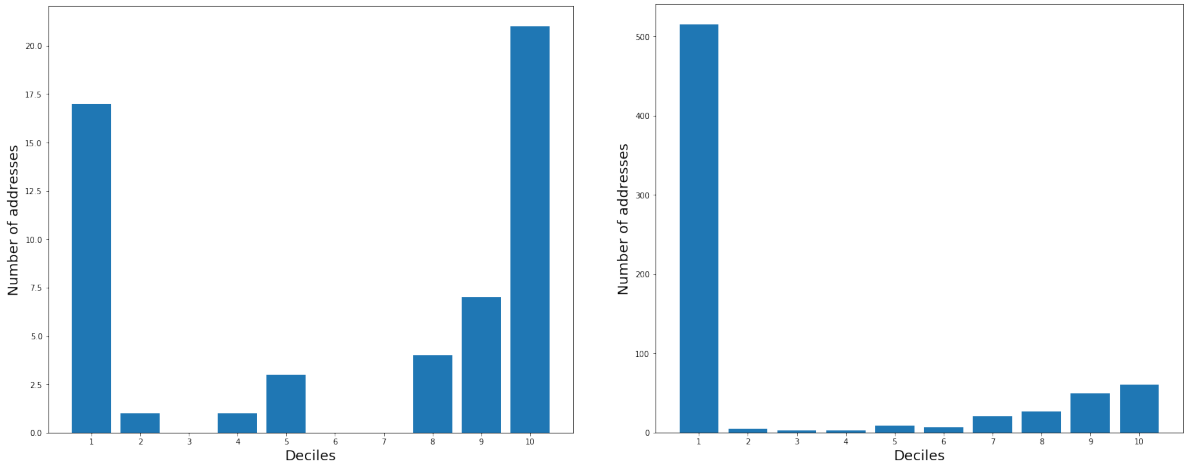

Figure 9: Distribution of the addresses of $\mathcal{S}^F$ (at left) and $\mathcal{M}^F$ (at right) against the number of contacts of T^F_{Pre}

of Missings nodes but also of Survivors ones. This indication represents the basis of the function `Predict_Bubble_Survivors()` of our algorithm.

Instead, Table 16 does not seem to give any indication on how searching, in the pre-bubble period, the future Entrants that will be among the main actors in the bubble and post-bubble periods.

All previous analyses performed for `from_addresses` in the pre-bubble period can be repeated for `to_addresses` in the same period. In Table 17, we report the average number of transactions, the average number of contacts and the average value of transactions for T^T_{Pre}, $\mathcal{S}^T$, $\mathcal{M}^T$ and $\mathcal{E}^T_{Pre}$ (the latter defined similarly to $\mathcal{E}^F_{Pre}$, but for `to_addresses` instead of `from_addresses`). Furthermore, in Figure 10, we show the distribution of the addresses of $\mathcal{S}^T$ and $\mathcal{M}^T$ against the number of transactions of T^T_{Pre}. Due to space limitations, we do not report the corresponding distributions against the number

of contacts of T^T_{Pre}. Both the table and the two distributions confirm, for `to_addresses`, the same results that we found previously for `from_addresses`.

	T^T_{Pre}	$\mathcal{S}^T$	$\mathcal{M}^T$	$\mathcal{E}^T_{Pre}$
Average Number of Transactions	28,035.76	138,663.66	10,121.69	599.78
Average Number of Contacts	5,329.76	23,007.33	2,165.56	294.28
Average Value of Transactions (Eth)	9.05	6.79	14.17	4.86

Table 17: Average number of transactions, average number of contacts and average value of transactions for T^T_{Pre}, $\mathcal{S}^T$, $\mathcal{M}^T$ and $\mathcal{E}^T_{Pre}$

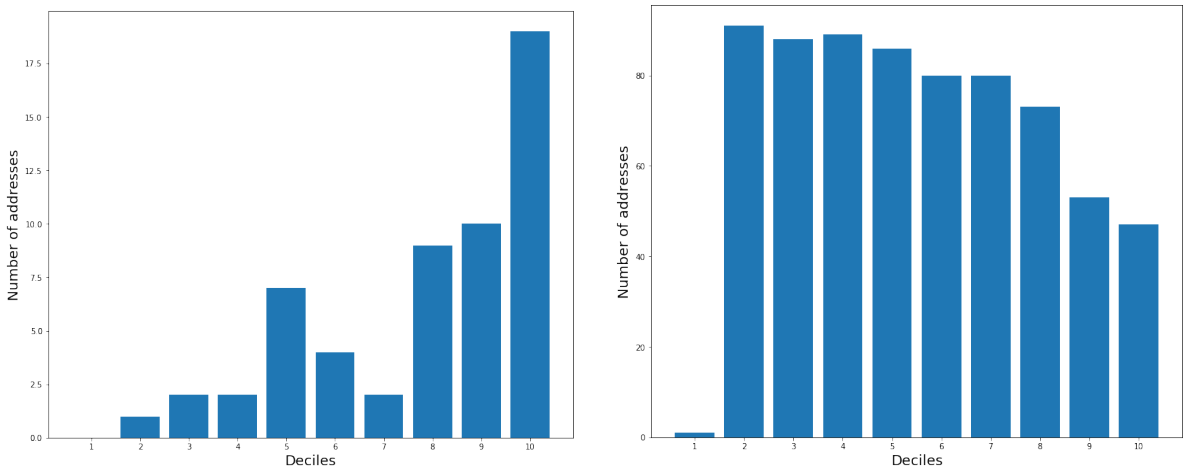

Figure 10: Distribution of the addresses of $\mathcal{S}^T$ (at left) and $\mathcal{M}^T$ (at right) against the number of transactions of T^T_{Pre}

So far we have examined pre-bubble data to identify some characteristics allowing us to predict who will be the main actors of the bubble period. Now, we want to do the same activity but examining bubble data to look for features allowing us to predict who will be the protagonists of the post-bubble period. In this analysis, we consider the following addresses: *(i)* T^F_B: the top 1000 `from_addresses` in the bubble period; *(ii)* $\mathcal{S}^F$: the Survivors `from_addresses`; *(iii)* $\mathcal{E}^F$: the Entrants `from_addresses`.

In Table 18, we show the average number of transactions, the average number of contacts and the average value of transactions for T^F_B, $\mathcal{S}^F$ and $\mathcal{E}^F$.

	T^F_B	$\mathcal{S}^F$	$\mathcal{E}^F$
Average Number of Transactions	45,418.29	266,183.77	46,010.31
Average Number of Contacts	10,100.95	55,029.89	12,851.75
Average Value of Transactions (Eth)	2.43	2.49	3.73

Table 18: Average number of transactions, average number of contacts and average value of transactions for T^F_B, $\mathcal{S}^F$ and $\mathcal{E}^F$

From the analysis of Table 18 we can see that, once again, it is easy to identify the Survivors of the post-bubble period. In fact, they generally have a significantly higher number of transactions and contacts than the other power `from_addresses`. Instead, the Entrants are not easily distinguishable,

because they have only slightly more transactions and contacts than the other power `from_addresses`. This represents a confirmation of what we had deduced from the analysis of Tables 10 - 15 and Figures 6 - 7, where we derived that the set of the Entrants is formed during the bubble period but it consolidates only during the post-bubble period.

This result is confirmed and substantially reinforced by Figure 11. In it, we can see that the Survivors are in the highest deciles, and this was expected considering the results of Table 18. However, a similar trend, although less marked, is also found for the Entrants. This represents a further important result because it allows us to define, at least partially, which nodes will be the Entrants in the post-bubble period. Similarly to what happened in the pre-bubble period, the distribution against the number of transactions is better than the one against the number of contacts in discriminating the Survivors and the Entrants against the other nodes during the post-bubble period. Indeed, in the case of the number of contacts, there is a certain number of addresses in the lowest decile, which, in fact, represents an outlier. Analogous conclusions could be drawn considering the distributions against the number of contacts of T_B^F. Due to space constraints, we do not report the corresponding graphs.

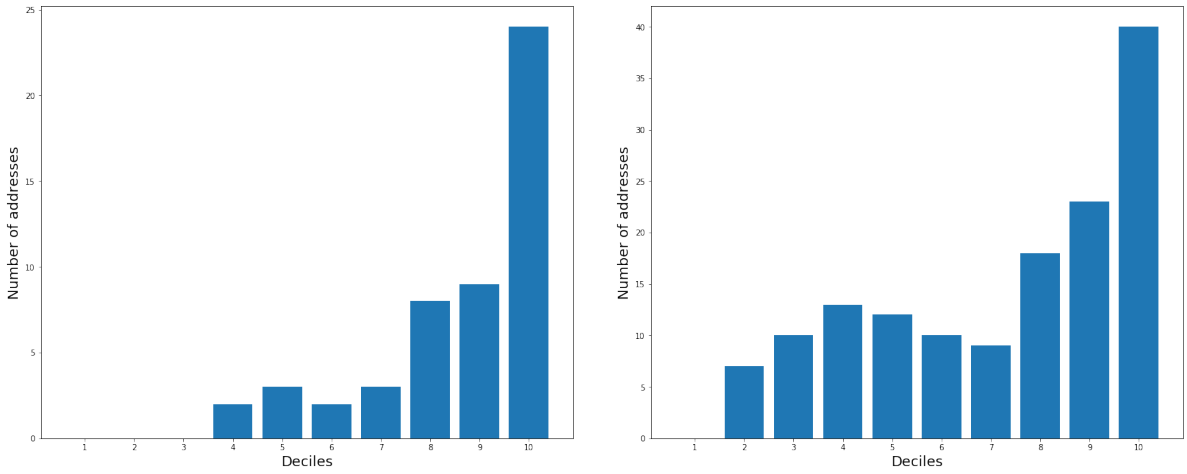

Figure 11: Distribution of the addresses of $\mathcal{S}^F$ (at left) and $\mathcal{E}^F$ (at right) against the number of transactions of T_B^F

Both Table 18 and Figure 11 give us the same important following indication: "The addresses that will survive a speculative bubble are to be searched among those that, in the bubble period, have carried out the highest numbers of transactions and have the highest numbers of contacts. If they also had this property in the pre-bubble period they belong to the Survivors, otherwise they belong to the Entrants.". This indication represents the basis of the functions `Predict_Post_Survivors()` and `Predict_Post_Entrants()` of our algorithm.

All previous analyses performed for `from_addresses` in the bubble period can be repeated for `to_addresses` in the same period. In Table 19, we report the average number of transactions, the average number of contacts and the average value of transactions for T_B^T, $\mathcal{S}^T$ and $\mathcal{E}^T$. This table confirms, for `to_addresses`, the same results we found previously for `from_addresses`.

	T_B^T	$\mathcal{S}^T$	$\mathcal{E}^T$
Average Number of Transactions	49,912.89	219,068.94	58,823.91
Average Number of Contacts	11,963.66	45,949.34	14,134.10
Average Value of Transactions (Eth)	1.90	1.98	1.71

Table 19: Average number of transactions, average number of contacts and average value of transactions for T_B^T, $\mathcal{S}^T$ and $\mathcal{E}^T$

4.6 What happened a few years later...

As a last analysis, we investigated how the power addresses of the post-bubble period behaved during the months following the time interval considered for our dataset, i.e., from January 2019 until today. For this purpose, we considered three subsets of the power addresses, i.e., the Survivors, the Entrants and the other nodes (hereafter, the Others), and we examined the date of the last transaction for them. The distribution of the Survivors (resp., the Entrants, the Others) against this date is shown in Figure 12 (resp., 13, 14) for `from_addresses`. Analogous distributions could be drawn for `to_address`. We do not report them due to space constraints. From the analysis of these figures we can observe that:

- As for `from_addresses`, we can see that most of the Survivors are still active. Many Entrants are also active but, unlike the Survivors, there is a fraction of them that ceased to operate in the second half of 2019. The date of the end of activity of the Others is, instead, more uniformly distributed. This is a further confirmation that the Survivors represent the vast part of the guiding users in Ethereum.
- As far as `to_addresses` are concerned, we can see that most of the Survivors and the Entrants are still active. The date of the end of activity of the Others is distributed in a more balanced way, even if there is a large amount of addresses still active also in this case. Therefore, as for `to_addresses`, we can deduce that the Survivors include most of the guiding users in Ethereum. However, differently from what happens for `from_addresses`, they have been flanked as leaders by the Entrants.

5 Applying our approach to hunt past, present and future speculators

In the Introduction, and, more generally, throughout this paper, we have seen that our approach aims at investigating user behavior during a cryptocurrency speculative bubble. In this section, we show how it is possible to know the identikit of speculators of past cryptocurrency bubbles by suitably integrating the knowledge patterns extracted by means of our approach (Subsection 5.1). Furthermore, we show how new speculative cryptocurrency bubbles are in the horizon (Subsection 5.2), and how our approach can be used to identify in advance who is maneuvering behind them to make big profits at the expense of all the other investors.

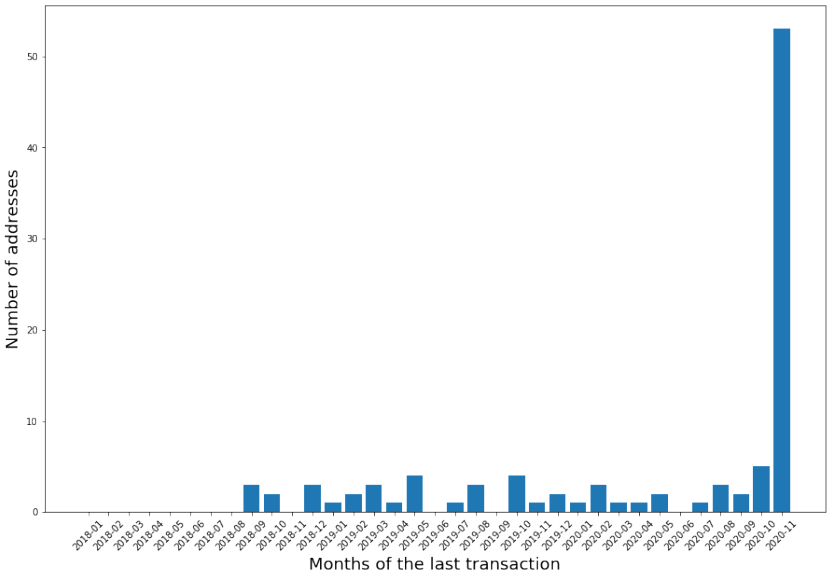

Figure 12: Distribution of the Survivors (`from_addresses`) against the date of the last transaction

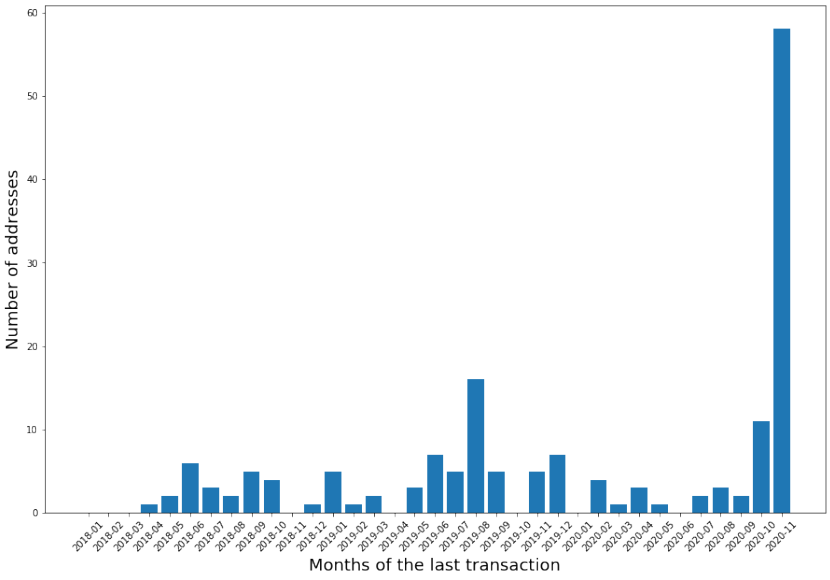

Figure 13: Distribution of the Entrants (`from_addresses`) against the date of the last transaction

5.1 Defining the identikit of cryptocurrency bubble speculators

In Section 4, we extracted some knowledge patterns involving various kinds of address present in a cryptocurrency blockchain by applying our approach to the Ethereum speculative bubble in the years 2017 and 2018. In this section, we want to verify whether the suitable integration and correlation of these patterns allow us to build an identikit of speculators.

Before proceeding in this task, we want to clarify that, although we will illustrate this demonstra-

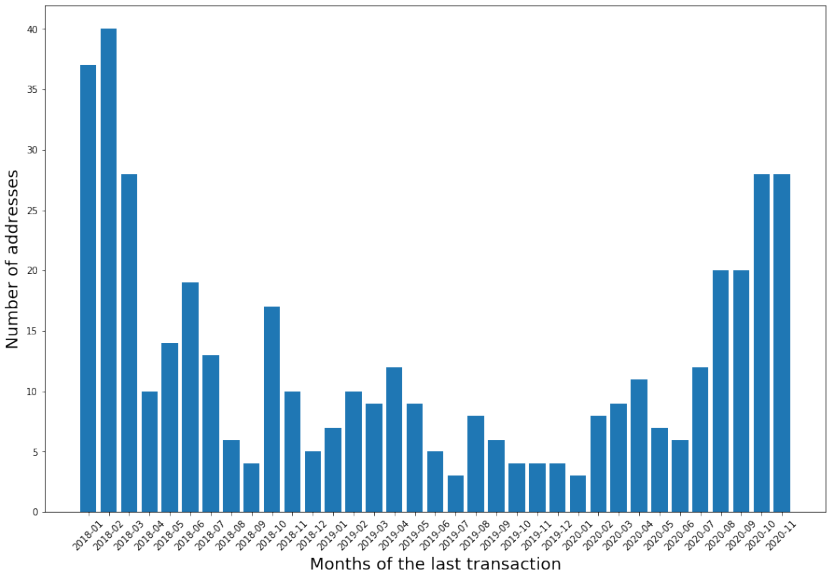

Figure 14: Distribution of the Others (`from_addresses`) against the date of the last transaction

tion on the knowledge patterns extracted from Ethereum, our discussion is general. In fact, it concerns a potentiality of our approach that is valid for any past, present and future speculative bubble.

We start with the information about the ego networks obtained in Section 4.3. It tells us that: *(i)* in the pre-bubble, bubble and post-bubble period, the Survivors have much larger ego networks than the other nodes; *(ii)* in the bubble period, the Survivors' ego networks are much larger than even the Entrants' ones; this difference fades in the post-bubble period. Recall that having a large ego network means having the possibility to influence a large number of nodes.

Now, we consider the information on backbones extracted in Section 4.4. It tells us that: *(i)* in the pre-bubble period, there is a strong backbone among the Survivors and a weaker backbone among the Missings; *(ii)* in the bubble period, there is a very strong backbone among the Survivors and a weaker backbone among the Entrants; this last is stronger than the corresponding one of the bubble period. Recall that the presence of a backbone among a set of nodes is an indicator that they tend to act in a coordinated way with each other.

We continue our investigation considering the characteristics of the future main actors, as extracted in Section 4.5. Here, we saw that the addresses that best survive a bubble must be sought among those that, in the pre-bubble and bubble periods, made the most transactions and had the most contacts. But, from what we saw in Section 4.3, the addresses with such characteristics are first of all the Survivors and then the Entrants.

Finally, the analysis of the nodes active in the period corresponding to the Ethereum bubble of the years 2017-2018 that are still active today, described in Section 4.5, also leads us to the same results, namely that most of the Survivors and a good portion of the Entrants present in the 2017-2018 Ethereum bubble are still active today.

All these considerations lead us to conclude that, indeed, in the Ethereum speculative bubble of the years 2017-2018, a group of speculators existed. Regarding the profile of the users belonging to

this group, we can conclude that most of them can be found among the Survivors and were already present in the pre-bubble period. They were flanked in the bubble period by a group of speculators that formed the Entrants. Initially, these were not the leaders of the phenomenon: at first, the leadership was of the Survivors alone. However, as time passed, the Entrants gradually consolidated and reached the leadership level that previously characterized only the Survivors.

5.2 Adoption of our approach in the next speculative bubble

The approach proposed in this paper is general and can be applied to any speculative bubble made on any cryptocurrency in the past, present and future. We believe that the ability to analyze past speculative bubbles and draw an identikit of the speculators is already a worthy property of our approach. But that property becomes even more interesting if, in addition to looking at the past, we look at the present and especially the future.

In fact, the cryptocurrency context is considered a highly speculative environment by many graduates of the Nobel Memorial Prize in Economic Sciences, central bankers and investors. Speculations on cryptocurrencies have also been observed recently. For example, on March 8^{th}, 2020 the price of Bitcoin was 8,901 USD. On March 12^{th}, 2020, it was 6,206 USD, with a decrease of about 30%. In October 2020 this price was already doubled again and was about 13,000 USD. On January 3^{rd}, 2021 the price of Bitcoin was 34,792 USD; the next day it decreased by 17%. On January 8^{th}, 2021 its value exceeded 40,000 USD and on February 16^{th}, 2021 it exceeded 50,000 USD. In March 2021 its value was 58,734 USD, while on May 8^{th}, 2021 it reached its highest value in history being 58,958 USD. On May 22^{th}, 2021 (which corresponds to the time of writing of this section) it had fallen again to 36,312 USD losing 38.41% of its value in 14 days.

Similar trends also characterize other cryptocurrencies. For example, the value of Ether was about 750 USD in December 2020, about 1,350 USD in January 2021, about 1,800 USD in March 2021 and about 2,700 USD in April 2021. On May 11^{th}, 2021 this value was equal to 4,179.81 USD, which represents the highest value reached by this cryptocurrency so far. On May 22^{th}, 2021 (which corresponds to the time of writing of this section) the value of the Ether was 2,300.01 USD with a collapse of 44.97% in 11 days.

The above examples highlight how prone the cryptocurrency world is to speculation. In addition, the trends of the last month lead analysts to believe that we are in the midst of a speculative bubble similar to the one of the years 2017-2018. If this is the case, the approach proposed in this paper would provide investors with many information about the behaviors of the various players operating in this market. It could even support analysts in understanding who are the current speculators behind these bubbles. Therefore, we believe that our approach has not only a value for the past but it provides useful predictive tools for the present and the future.

6 Conclusion

In this paper, we presented a Social Network Analysis based approach to investigate user behavior during a cryptocurrency speculative bubble in order to extract knowledge patterns about this phenomenon. The proposed approach is general and can be applied to any past, present and future

cryptocurrency speculative bubble. To verify its potential, we applied it to the study of the speculative bubble involving Ethereum in the years 2017 and 2018. In this way, we were able to extract several knowledge patterns about the behavior of important categories of users in the pre-bubble, bubble and post-bubble periods. Then, we showed how it is possible to define an identikit of the speculators who maneauvered behind the bubble by appropriately integrating these knowledge patterns. In the last part of the paper, we showed how this capability of supporting the hunting for speculators is not limited to the particular case of the 2017-2018 Ethereum bubble, but it is intrinsic of our approach and can cover past, present and future speculations involving cryptocurrencies.

The activities described in this paper are not to be considered as a point of arrival. Instead, they are a starting point for further researches in this field. For example, we might perform further studies on user behavior, taking into account labels identifying the type of addresses in a blockchain. Based on these labels, we would like to define a classification approach that first constructs a profile for all users of each label and, then, employs that profile to classify non-labeled users. In addition, we could think of upgrading from predictive to prescriptive analysis by defining the characteristics that a new user must take over time in a blockchain for quickly becoming one of the main actors in it. Last, but not the least, we could investigate the text data sent along with transactions. Indeed, it would be possible to analyze the shared contents through Natural Language Processing techniques in order to detect additional features allowing a more precise definition of the profiles of the main players in the blockchain.

References

[1] N. Antonakakis, I. Chatziantoniou, and D. Gabauer. Cryptocurrency market contagion: Market uncertainty, market complexity, and dynamic portfolios. *Journal of International Financial Markets, Institutions and Money*, 61:37–51, 2019. Elsevier.

[2] C. Baek and M. Elbeck. Bitcoins as an investment or speculative vehicle? A first look. *Applied Economics Letters*, 22(1):30–34, 2015. Taylor & Francis.

[3] B.M. Blau. Price dynamics and speculative trading in bitcoin. *Research in International Business and Finance*, 41:493–499, 2017. Elsevier.

[4] E. Bouri, C.K.M. Lau, B. Lucey, and D. Roubaud. Trading volume and the predictability of return and volatility in the cryptocurrency market. *Finance Research Letters*, 29:340–346, 2019. Elsevier.

[5] P. Chaim and M.P. Laurini. Is Bitcoin a bubble? *Physica A: Statistical Mechanics and its Applications*, 517:222–232, 2019. Elsevier.

[6] W. Chan and A. Olmsted. Ethereum transaction graph analysis. In *Proc. of the International Conference for Internet Technology and Secured Transactions (ICITST'17)*, pages 498–500, Cambridge, MA, USA, 2017. IEEE.

[7] E.T. Cheah and J. Fry. Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics Letters*, 130:32–36, 2015. Elsevier.

[8] C.Y.H. Chen and C.M. Hafner. Sentiment-induced bubbles in the cryptocurrency market. *Journal of Risk and Financial Management*, 12(2):53, 2019. Multidisciplinary Digital Publishing Institute.

[9] A. Cheung, E. Roca, and J. Su. Crypto-currency bubbles: an application of the Phillips–Shi–Yu (2013) methodology on Mt. Gox bitcoin prices. *Applied Economics*, 47(23):2348–2358, 2015. Taylor & Francis.

[10] J. Cohen. Trusses: Cohesive subgraphs for social network analysis. *National security agency technical report*, 16(3.1), 2008.

[11] S. Corbet, B. Lucey, A. Urquhart, and L. Yarovaya. Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, 62:182–199, 2019. Elsevier.

[12] S. Corbet, B. Lucey, and L. Yarovaya. Datestamping the Bitcoin and Ethereum bubbles. *Finance Research Letters*, 26:81–88, 2018. Elsevier.

[13] R. DeJordy and D. Halgin. Introduction to ego network analysis. *Boston MA: Boston College and the Winston Center for Leadership & Ethics*, 2008.

[14] S. Dorogovtsev, A. Goltsev, and J. Mendes. K-core organization of complex networks. *Physical Review Letters*, 96(4):040601, 2006. APS.

[15] A. ElBahrawy, L. Alessandretti, A. Kandler, R. Pastor-Satorras, and A. Baronchelli. Evolutionary dynamics of the cryptocurrency market. *Royal Society Open Science*, 4(11):170623, 2017. The Royal Society Publishing.

[16] Y. Eom. Premium and speculative trading in bitcoin. *Finance Research Letters*, page 101505, 2020. Elsevier.

[17] A. Esfandyari, M. Zignani, S. Gaito, and G.P. Rossi. User identification across online social networks in practice: Pitfalls and solutions. *Journal of Information Science*, 44(3):377–391, 2018. SAGE Publications.

[18] J. Fry and E.T. Cheah. Negative bubbles and shocks in cryptocurrency markets. *International Review of Financial Analysis*, 47:343–352, 2016. Elsevier.

[19] J.C. Gerlach, G. Demos, and D. Sornette. Dissection of Bitcoin's multiscale bubble history from January 2012 to February 2018. *Royal Society Open Science*, 6(7):180643, 2019. The Royal Society.

[20] M. Hosseini-Pozveh, K. Zamanifar, and A.R. Naghsh-Nilchi. A community-based approach to identify the most influential nodes in social networks. *Journal of Information Science*, 43(2):204–220, 2017. SAGE Publications.

[21] M. Ianni, E. Masciari, G.M. Mazzeo, M. Mezzanzanica, and C. Zaniolo. Fast and effective big data exploration by clustering. *Future Generation Computer Systems*, 102:84–94, 2020. Elsevier.

[22] L. Jiang and X. Zhang. BCOSN: A blockchain-based decentralized online social network. *IEEE Transactions on Computational Social Systems*, 6(6):1454–1466, 2019. IEEE.

[23] H. Kalodner, M. Möser, K. Lee, S. Goldfeder, M. Plattner, A. Chator, and A. Narayanan. Blocksci: Design and applications of a blockchain analysis platform. In *Proc. of the International Security Symposium (USENIX'20)*, pages 2721–2738, 2020. USENIX Association.

[24] Y. Ko, D. Chae, and S. Kim. Influence maximisation in social networks: A target-oriented estimation. *Journal of Information Science*, 44(5):671–682, 2018. SAGE Publications.

[25] X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen. A survey on the security of blockchain systems. *Future Generation Computer Systems*, 107:841–853, 2020. Elsevier.

[26] M. Lischke and B. Fabian. Analyzing the bitcoin network: The first four years. *Future Internet*, 8(1):7, 2016. Multidisciplinary Digital Publishing Institute.

[27] R. Liu, S. Wan, Z. Zhang, and X. Zhao. Is the introduction of futures responsible for the crash of Bitcoin? *Finance Research Letters*, 34:101259, 2020. Elsevier.

[28] D.D.F. Maesa, A. Marino, and L. Ricci. Uncovering the bitcoin blockchain: an analysis of the full users graph. In *Proc. of the International Conference on Data Science and Advanced Analytics (DSAA'16)*, pages 537–546, Montreal, Quebec, Canada, 2016. IEEE.

[29] E. Masciari. SMART: stream monitoring enterprise activities by RFID tags. *Information Sciences*, 195:25–44, 2012. Elsevier.

[30] M. McPherson, L. Smith-Lovin, and J.M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001. JSTOR.

[31] R.C. Phillips and D. Gorse. Predicting cryptocurrency price bubbles using social media data and epidemic modelling. In *Proc. of the International Symposium Series on Computational Intelligence (SSCI'17)*, pages 1–7, Honolulu, HI, USA, 2017. IEEE.

[32] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *The Cryptography Mailing List*, 2008.

[33] A. Sheikhahmadi and M.A. Nematbakhsh. Identification of multi-spreader users in social networks for viral marketing. *Journal of Information Science*, 43(3):412–423, 2017. SAGE Publications Sage UK: London, England.

[34] V.N. Soloviev and A. Belinskiy. Complex systems theory and crashes of cryptocurrency market. In *Proc. of the International Conference on Information and Communication Technologies in Education, Research, and Industrial Applications (ICTERI'18)*, pages 276–297, Kyiv, Ukraine, 2018. Springer.

[35] S. Somin, G. Gordon, and Y. Altshuler. Network analysis of ERC20 tokens trading on ethereum blockchain. In *Proc. of the International Conference on Complex Systems (ICCS'18)*, pages 439–450, Cambridge, MA, USA, 2018. Springer.

[36] H. Sun, N. Ruan, and H. Liu. Ethereum Analysis via Node Clustering. In *Proc. of the International Conference on Network and System Security (NSS'19)*, pages 114–129, Sapporo, Japan, 2019. Springer.

[37] M. Thelwall. Can social news websites pay for content and curation? The SteemIt cryptocurrency model. *Journal of Information Science*, 44(6):736–751, 2018. SAGE Publications.

[38] P. Wang and Y. Wen. Speculative bubbles and financial crises. *American Economic Journal: Macroeconomics*, 4(3):184–221, 2012.

[39] O.S. Yaya, A.E. Ogbonna, and O.E. Olubusoye. How persistent and dynamic inter-dependent are pricing of Bitcoin to other cryptocurrencies before and after 2017/18 crash? *Physica A: Statistical Mechanics and its Applications*, 531:121732, 2019. Elsevier.

[40] O.S. Yaya, E.A. Ogbonna, and R. Mudida. Market Efficiency and Volatility Persistence of Cryptocurrency during Pre-and Post-Crash Periods of Bitcoin: Evidence based on Fractional Integration. *International Journal of Finance and Economics*, 2020. John Wiley & Sons.

[41] J. Yli-Huumo, D. Ko, S. Choi, S. Park, and K. Smolander. Where is current research on blockchain technology? A systematic review. *PloS one*, 11(10):e0163477, 2016. PloS ONE.

[42] Z. Zheng, S. Xie, H.N. Dai, X. Chen, and H. Wang. Blockchain challenges and opportunities: A survey. *International Journal of Web and Grid Services*, 14(4):352–375, 2018. Inderscience.