ERC20 Transactions over Ethereum Blockchain: Network Analysis and Predictions

Shahar Somin[1,2], Goren Gordon[2,3], Alex Pentland[1],

Erez Shmueli[2], and Yaniv Altshuler[1,3]

[1] MIT Media Lab, Cambridge, MA, USA
shaharso@media.mit.edu
[2] Industrial Engineering Department, Tel Aviv University, Israel
[3] Endor Ltd.

Abstract. Following the birth of Bitcoin and the introduction of the Ethereum ERC20 protocol a decade ago, recent years have witnessed a growing number of cryptographic tokens that are being introduced by researchers, private sector companies and NGOs. The ubiquitous of such Blockchain based cryptocurrencies give birth to a new kind of rising economy, which presents great difficulties to modeling its dynamics using conventional semantic properties. Our work presents the analysis of the dynamical properties of the ERC20 protocol compliant crypto-coins trading data using a network theory prism. We examine the dynamics of ERC20 based networks over time by analyzing a meta-parameter of the network — the power of its degree distribution. Our analysis demonstrates that this parameter can be modeled as an under-damped harmonic oscillator over time, enabling a year forward of network parameters predictions.

Keywords: Network Dynamics, Complex Systems, Network Analysis, Blockchain, Ethereum, Smart contracts, ERC20 tokens, cryptocurrency

1 Introduction

Blockchain technology, which has been known by mostly small technological circles up until recently, is bursting throughout the globe, with a potential economic and social impact that could fundamentally alter traditional financial and social structures. Launched in July 2015 [11], the Ethereum Blockchain is a public ledger that keeps publicly accessible records of all Ethereum related transactions. The ability of the Ethereum Blockchain to store not only ownership, similarly to the Bitcoin Blockchain, but also execution code, in the form of *"Smart Contracts"*, has recently led to the creation of an immense number of new types of "tokens", based on the Ethereum ERC20 protocol.

Apart from providing full data of prices, volumes and holders distribution, the ERC20 transactional data also presents the monetary activity of anonymous individuals, which is otherwise scarce and hard to obtain due to confidentiality and privacy control. Thereby, this ERC20 digital ecosystem intrinsically provides a rare opportunity to analyze and model financial behavior in an evolving market over a long period of time. Specifically, understanding the governing forces upon this emerging economy, and in turn being able to perform accurate predictions of the economy's state are fundamental, as this market is becoming increasingly relevant to the traditional financial world.

In this work we aim to broaden our comprehension of the dynamics this financial ecosystem undergoes, from a network theory perspective. Specifically, we first demonstrate how the dynamics of the degree distribution's power parameter can be modeled by an under-damped oscillator with zero-mean Gaussian noise. In turn, this analytical model enables us to predict network's γ dynamics, reliably predicting a whole year forward in time.

2 Background and Related Work

Blockchain's ability to process transactions in a trust-less environment, apart from trading its official cryptocurrency, the *Ether*, presents the most prominent framework for the execution of "*Smart Contracts*" [33]. Smart Contracts are computer programs, formalizing digital agreements, automatically enforced to execute any predefined conditions using the consensus mechanism of the Blockchain, without relying on a trusted authority. Moreover, Smart Contracts allow companies or entrepreneurs to create their own proprietary tokens on top of the Blockchain protocol [13]. The most widely used token standard is Ethereums *ERC20* (representing Ethereum Request for Comment), issued in 2015. The protocol defines technical specifications giving developers the ability to program how new tokens will function within the Ethereum ecosystem.

There has been a surge in recent years in the attempt to model social dynamics via statistical physics tools [12, **?**,**?**]. Econophysics have attempted to describe the dynamical nature of the economy with different, and increasingly sophisticated physical models. Frisch [15], who started this trend, has suggested to use a damped oscillator model to the economy post wars or disasters, with the assumption that there is an equilibrium state that has been perturbed. Since then, many new models have been suggested, ranging from quantum mechanical models [34, 16] to chaos theory [17, 25]. However, all of these models have attempted to describe the economy, represented by a singular *value*, e.g. stock market prices, whereas the underlying network of the economy has not been addressed. Network science, however, has exceedingly contributed to multiple and diverse scientific disciplines in the past two decades, by examining exactly diverse network related parameters. Applying network analysis and graph theory have assisted in revealing the structure and dynamics of complex systems by representing them as networks, including social networks [8, 21, 22], computer communication networks [24], biological systems [7], transportation [28, 4], IOT [1], emergency detection [3] and financial trading systems [2, 23, 27]. Most of the research conducted in the Blockchain world, was concentrated in Bitcoin, spreading from theoretical foundations [10], security and fraud [20, 29] to some comprehensive research in network analysis [26, 19, 18]. The world of Smart contracts has recently inspired research in aspects of design patterns, applications and security [9, 5, 14, 6], policy towards ICOs has also been studied [13]. Some preliminary results examining network theory's applicability to ERC20 tokens have been made in [31, 32], specifically by validating that this financial ecosystem, when considered as a network of interactions, adheres to key network theory principles, such as power-law degree distribution. In this paper we aim to examine how this prominent field can enhance the understanding of the underlying structure of the ERC20 tokens trading data, model it's stabilization process as a network over time and achieve predictive abilities.

3 Results

In this work we analyze the dynamics of ERC20 tokens' trading over the Ethereum Blockchain. We obtain the ERC20 transactions using the methodology thoroughly explained in [32]. We have retrieved all ERC20 tokens transactions spreading between February 2016 and June 2018, resulting in $88,985,493$ token trades, performed by $17,611,649$ unique wallets, trading $51,281$ token addresses.

During the examined timespan of 2.5 years of ERC20 transactions, the economy keeps evolving and changing its dynamics. Not only does the rising public interest in Blockchain and tokens induce an exponential growth in transactions' volume, but the traded tokens in this economy change as well, as new tokens are established and others lose their impact and decay. A thorough discussion of the dynamics of the economical properties of the ERC20 economy was conducted in [30].

3.1 Temporal Dynamics: The Oscillating Network Model

We apply temporal graph analysis to a sliding window of weekly graphs. Namely, we define a weekly transactions graph G_t based on ERC20 trading activity during a week $[t-7,\, t)$ as follows:

Definition 1. *The weekly transactions graph for a given day t, $G_t(V_t, E_t)$ is the directed graph constructed from all trading transactions over any ERC20 token, made during the time period $[t-7, t)$. The set of vertices V_t consists of all wallets trading during that period:*

$$V_t := \{v \| \text{ wallet } v \text{ bought or sold any token during } [t-7, t)\} \tag{1}$$

and the set of edges $E_t \subseteq V_t \times V_t$ is defined as:

$$E_t := \{(u, v) \| \text{ wallet } u \text{ sold to wallet } v \text{ any token during } [t-7, t)\} \tag{2}$$

Over the examined period of 2.5 years, we construct 1000 such weekly transactions graphs, using daily rolling windows, each containing one week of transactional data. As seen in [31] the incoming and outgoing degree distributions of G_t clearly adhere to a power-law model.

Next, we turn to model the degree distribution over time, as captured by its associated γ values. We postulate that any network of human related transactions, has a characteristic *stable state*, in the form of γ_∞^{in} and γ_∞^{out}, to which the network strives to converge:

$$\gamma_t^{in} \underset{t\to\infty}{\longrightarrow} \gamma_\infty^{in},\ \ \gamma_t^{out} \underset{t\to\infty}{\longrightarrow} \gamma_\infty^{out}$$

Empirical observations of both γ_t^{in} and γ_t^{out} coincide with this hypothesis, as can be seen in Fig. 1, and can be efficiently modeled as an Harmonic Under-Damped Oscillator, formally $\forall t$:

$$osc(t) = A \cdot e^{-\omega_0 \zeta t} \cdot \sin(\omega_0 \sqrt{1-\zeta^2} t + \varphi) + \gamma_\infty$$

Where A represents the maximal amplitude of the oscillation, ω_0 stands for the resonant frequency of the system, ζ for the damping ratio, φ is the phase of the oscillation and γ_∞ is the equilibrium state. The parameters of fitted oscillators to γ_t^{in} and γ_t^{out}, $osc^{in}(t)$ and $osc^{out}(t)$ correspondingly, are presented in Table 1

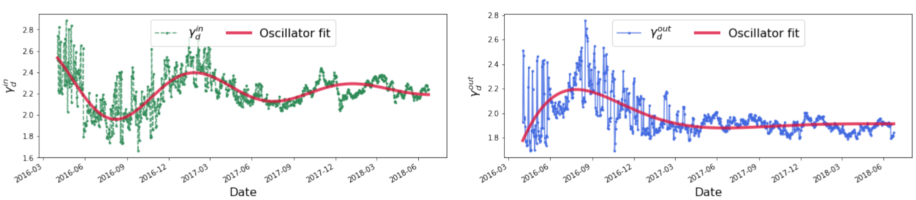

Fig. 1: ERC20 transactional network temporal development, in a network related prism, demonstrating the underlying consolidation process the network undergoes. Evolvement of incoming degree distribution gradient, γ_t^{in}, is depicted in the upper panel and out-degree distribution gradient γ_t^{out} is displayed in the lower panel. Both gradients converge to their *stable states* γ_∞^{in} and γ_∞^{out} correspondingly, following a Harmonic Under-Damped Oscillator model.

Table 1: Under-Damped Oscillator Models Parameters

Type	A	φ	γ_∞	$\frac{2\pi}{\omega_0}$ **(days)**	ζ
$osc_{FT}^{out}(t)$	-0.77	2.96	1.91	530.2	0.577
$osc_{FT}^{in}(t)$	0.39	2.23	2.23	341.1	0.152

3.2 The Oscillating Network Model: Predictive Ability

The under-damped oscillating model can be used for predictive purposes. We fit partial γ observations to an oscillator model, considering data restricted by date, in order to predict future γ dynamics, formally defined as :

Definition 2. *Let T_0 stand for the minimal date available in our dataset, April 1st, 2016. Given any time-stamp $T_i > T_0$, we define osc_{T_i} to be the partial oscillator model, representing the under-damped oscillator model fitted to γ values between T_0 and T_i. The parameters characterizing osc_{T_i} are denoted by $A(T_i)$, $\varphi(T_i)$, $\gamma_\infty(T_i)$, $\omega_0(T_i)$ and $\zeta(T_i)$.*

We select 5 different *inspection dates*, referred as e_d:

- September 28th, 2016
- December 27th, 2016
- March 27th, 2017
- June 25th, 2017

and analyze the predictive ability of osc_{T_i}' as to γ dynamics over the $[e_d, \text{June 2018})$ timespan.

In order to establish the confidence levels for each e_d associated prediction, we analyze the performance of 90 *partial oscillator models* for each e_d, forming a set, we'd refer to as O_{e_d}:

$$O_{e_d} = \bigcup_{T_i \in [e_d - 90, e_d)} osc_{T_i} \tag{3}$$

The mean and standard deviation of O_{e_d}'s prediction of γ for a given time $t \in FT$ are defined as:

$$\begin{aligned} \text{mean}(O_{e_d})(t) &\equiv \underset{T_i \in [e_d - 90, e_d)}{\text{mean}} (osc_{T_i}(t)) \\ \text{STD}(O_{e_d})(t) &\equiv \underset{T_i \in [e_d - 90, e_d)}{\text{STD}} (osc_{T_i}(t)) \end{aligned} \tag{4}$$

Fig. 2 depicts the predictions made for both γ^{in} and γ^{out}, presenting mean(O_{e_d}) and STD(O_{e_d}) along $[e_d, \text{June 2018})$, for each of the 5 *inspection dates*. The prediction analysis demonstrates that predictions for γ^{in} stabilize as e_d advances, until finally presenting high reliability, for predicting a whole year of data, starting from e_d = June 25, 2017. We further observe that confidence levels for the predictions increase with e_d for both γ^{in} and γ^{out}, manifested by the decreasing standard deviation as e_d progresses.

We note however, that the predictive ability for γ^{out} when estimated at e_d = June 25, 2017 isn't as strong, manifested by the evident over-estimation produced by mean(O_{e_d}) while predicting γ^{out} values during [June 2017, June 2018). This apparent bias in the prediction of γ^{out} values manifests the lack of oscillations in the actual observed γ^{out} values, as well as an overestimation of γ_∞ by mean(O_{e_d}). This may suggest that there are other *forces* influencing the dynamics of γ^{out}, rather than just *'spring and friction'*-like forces.

We further wish to analyze the *'goodness-of-fit'* of osc_{T_i} for any given T_i, for the entire $[T_i, \text{June 2018})$ timespan. We therefore calculate the *Root Mean Squared Error* of each osc_{T_i}, namely:

$$\text{RMSE}(osc_{T_i}) = \sqrt{\frac{1}{n_{T_i}} \sum_{t \in [T_i, \text{June 2018})} (\gamma(t) - osc_{T_i}(t))^2} \tag{5}$$

where n_{T_i} is the length of $[T_i$, June 2018) period, measured in days.

In order to smoothen the RMSE signal, we calculate a 90-days rolling mean over the RMSE of the *partial oscillator model*:

$$\begin{aligned} \text{mean}_{T_i}(\text{RMSE}(osc)) &\equiv \underset{t \in [T_i - 90, T_i)}{\text{mean}} (\text{RMSE}(osc_t)) \\ \text{STD}_{T_i}(\text{RMSE}(osc)) &\equiv \underset{t \in [T_i - 90, T_i)}{\text{STD}} (\text{RMSE}(osc_t)) \end{aligned} \tag{6}$$

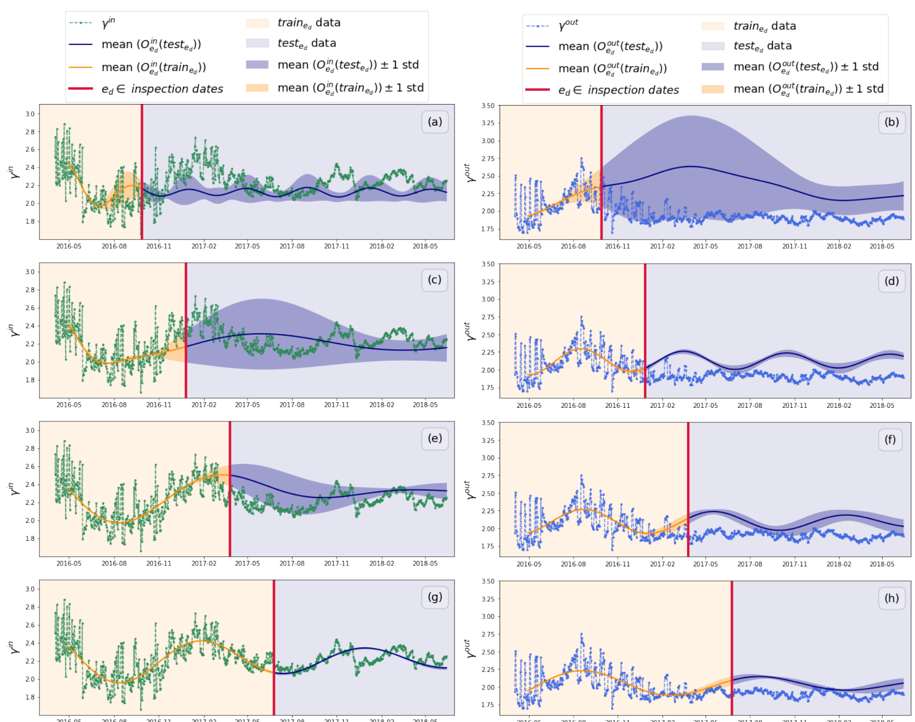

Fig. 2: Analyzing the abilities of the partial oscillator model to predict γ^{in} and γ^{out} dynamics over $[e_d$, June 2018), along 5 different inspection dates e_d. Presenting mean and standard deviation of O_{e_d}, the set of 90 consecutive partial oscillator models, fitted until e_d (see Eq. 4). Left panels depicted prediction of γ^{in} dynamics, and right panels present γ^{out} dynamics predictions by mean$(O^{in}_{e_d})$ and mean$(O^{out}_{e_d})$ respectively.

Fig. 3 depicts the Root Mean Squared Error of both $osc^{in}_{T_i}$ and $osc^{out}_{T_i}$, presenting both $\text{mean}_{T_i}(\text{RMSE}(osc))$ and $\text{STD}_{T_i}(\text{RMSE}(osc))$ for predicting γ^{in} and γ^{out} along time. This analysis assists in validating that in average, $osc^{in}_{T_i}$ has lower error values compared to $osc^{out}_{T_i}$ along most of the examined timespan and specifically over the predictions of the last year of data, starting from $e_d =$ June 25, 2017.

We note however that both $osc^{in}_{T_i}$ and $osc^{out}_{T_i}$ converge to similar RMSE values, starting from December 2017, yielding 7 months of similar and low error predictions for both γ^{in} and γ^{out} dynamics. The decreasing RMSE, and its standard deviation enhance even further the predictive abilities of the under-damped oscillator as a model of γ dynamics, concluding that osc^{in} is a better predictor compared to osc^{out} , as both its error rate and its confidence levels, present significant decrease earlier in time, enabling a reliable prediction of an entire year of data.

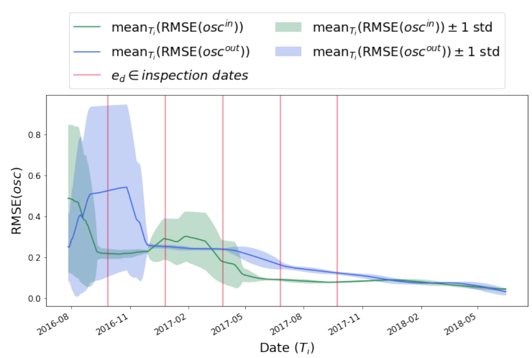

Fig. 3: Average and standard deviation of Root Mean Squared Error of both $osc^{out}_{T_i}$ and $osc^{in}_{T_i}$, over 90 consecutive partial oscillator models, osc_{T_i}, manifesting the ability of the under-damped oscillator to predict γ dynamics. This analysis presents both the improving accuracy of models as time advances, and their growing stability, manifested by a decreasing standard deviation.

As part of this novel approach to modeling the network's consolidation process, one should further note that the amplitude of the under-damped oscillator is governed by:

$$A \cdot e^{-\omega_0 \zeta t} \tag{7}$$

The latter enables establishing the time t_1 at which the network has reached a *stabilization of* $x\%$, formally:

Definition 3. *Let $G_t(V_t, E_t)$ be the directed graph based on all transactions made during $[t-7, t)$, trading any of the ERC20 tokens, for a given $t \in FT$. Let γ_t denote the power of the associated degree distribution of G_t, whose dynamics* *modeled by an oscillator γ_{fit}. We define the 'x% stabilization time of the network' w.r.t γ to be the time t_1 when the amplitude of γ_{fit} reaches at most x% of the initial amplitude, observed as time t_0:*

$$x = e^{\omega_0 \zeta (t_0 - t_1)} \implies t_1 = t_0 - \frac{\ln(x)}{\omega_0 \zeta} \tag{8}$$

This, in turn, enables us to establish the time required for the network to reach stabilization, in both aspects of γ_t^{in} and γ_t^{out}. For instance, using the fitted parameters of the under-damped oscillator depicted in Table 1, one can verify that a 70% stabilization occurs after 430 days for γ_t^{in}:

$$t_1^{\gamma^{in}} = -\frac{\ln(0.3)}{0.018 \cdot 0.152} = 429.3$$

Using $x = 0.3$, $\omega_0 = 0.018$, $\zeta = 0.152$, $t_0 = 0$. γ_d^{out} presents the same stabilization after merely 177 days:

$$t_1^{\gamma^{out}} = -\frac{\ln(0.3)}{0.011 \cdot 0.577} = 176.1$$

where $x = 0.3$, $\omega_0 = 0.011$, $\zeta = 0.577$, $t_0 = 0$.

References

1. Altshuler, Y., Fire, M., Aharony, N., Elovici, Y., Pentland, A.: How many makes a crowd? on the correlation between groups' size and the accuracy of modeling. In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction. pp. 43–52. Springer (2012)
2. Altshuler, Y., Pan, W., Pentland, A.: Trends prediction using social diffusion models. In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction. pp. 97–104. Springer (2012)
3. Altshuler, Y., Fire, M., Shmueli, E., Elovici, Y., Bruckstein, A., Pentland, A.S., Lazer, D.: The social amplifier—reaction of human communities to emergencies. Journal of Statistical Physics **152**(3), 399–418 (2013)
4. Altshuler, Y., Puzis, R., Elovici, Y., Bekhor, S., Pentland, A.S.: On the rationality and optimality of transportation networks defense: a network centrality approach. Securing Transportation Systems pp. 35–63 (2015)
5. Anderson, L., Holz, R., Ponomarev, A., Rimba, P., Weber, I.: New kids on the block: an analysis of modern blockchains. arXiv preprint arXiv:1606.06530 (2016)
6. Atzei, N., Bartoletti, M., Cimoli, T.: A survey of attacks on ethereum smart contracts (sok). In: International Conference on Principles of Security and Trust. pp. 164–186. Springer (2017)
7. Barabasi, A.L., Oltvai, Z.N.: Network biology: understanding the cell's functional organization. Nature reviews genetics **5**(2), 101 (2004)
8. Barrat, A., Barthelemy, M., Vespignani, A.: Dynamical processes on complex networks. Cambridge university press (2008)

9. Bartoletti, M., Pompianu, L.: An empirical analysis of smart contracts: platforms, applications, and design patterns. In: International Conference on Financial Cryptography and Data Security. pp. 494–509. Springer (2017)
10. Bonneau, J., Miller, A., Clark, J., Narayanan, A., Kroll, J.A., Felten, E.W.: Sok: Research perspectives and challenges for bitcoin and cryptocurrencies. In: Security and Privacy (SP), 2015 IEEE Symposium on. pp. 104–121. IEEE (2015)
11. Buterin, V., et al.: A next-generation smart contract and decentralized application platform. white paper (2014)
12. Castellano, C., Fortunato, S., Loreto, V.: Statistical physics of social dynamics. Reviews of modern physics **81**(2), 591 (2009)
13. Catalini, C., Gans, J.S.: Initial coin offerings and the value of crypto tokens. Tech. rep., National Bureau of Economic Research (2018)
14. Christidis, K., Devetsikiotis, M.: Blockchains and smart contracts for the internet of things. IEEE Access **4**, 2292–2303 (2016)
15. Frisch, R., et al.: Propagation problems and impulse problems in dynamic economics (1933)
16. Gonçalves, C.P.: Quantum financial economics—risk and returns. Journal of Systems Science and Complexity **26**(2), 187–200 (2013)
17. Goodwin, R.M.: The economy as a chaotic growth oscillator. In: The Dynamics of the Wealth of Nations, pp. 300–310. Springer (1993)
18. Lischke, M., Fabian, B.: Analyzing the bitcoin network: The first four years. Future Internet **8**(1) (2016)
19. Maesa, D.D.F., Marino, A., Ricci, L.: Uncovering the bitcoin blockchain: an analysis of the full users graph. In: Data Science and Advanced Analytics (DSAA), 2016 IEEE International Conference on. pp. 537–546. IEEE (2016)
20. Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G.M., Savage, S.: A fistful of bitcoins: characterizing payments among men with no names. In: Proceedings of the 2013 conference on Internet measurement conference. pp. 127–140. ACM (2013)
21. Newman, M.E.: The structure and function of complex networks. SIAM review **45**(2), 167–256 (2003)
22. Newman, M.E.: Power laws, pareto distributions and zipf's law. Contemporary physics **46**(5), 323–351 (2005)
23. Pan, W., Altshuler, Y., Pentland, A.: Decoding social influence and the wisdom of the crowd in financial trading network. In: Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom). pp. 203–209. IEEE (2012)
24. Pastor-Satorras, R., Vespignani, A.: Evolution and structure of the Internet: A statistical physics approach. Cambridge University Press (2007)
25. Puu, T.: Attractors, bifurcations, & chaos: Nonlinear phenomena in economics. Springer Science & Business Media (2013)
26. Ron, D., Shamir, A.: Quantitative analysis of the full bitcoin transaction graph. In: International Conference on Financial Cryptography and Data Security. pp. 6–24. Springer (2013)
27. Shmueli, E., Altshuler, Y., et al.: Temporal dynamics of scale-free networks. In: International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction. pp. 359–366. Springer (2014)
28. Shmueli, E., Mazeh, I., Radaelli, L., Pentland, A.S., Altshuler, Y.: Ride sharing: a network perspective. In: International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction. pp. 434–439. Springer (2015)

29. Shrobe, H., Shrier, D.L., Pentland, A.: New Solutions for Cybersecurity. MIT Press (2018)
30. Somin, S., Altshuler, Y., Gordon, G., Shmueli, E., et al.: network dynamics of a financial ecosystem. Scientific reports **10**(1), 1–10 (2020)
31. Somin, S., Gordon, G., Altshuler, Y.: Network analysis of erc20 tokens trading on ethereum blockchain. In: International Conference on Complex Systems. pp. 439–450. Springer (2018)
32. Somin, S., Gordon, G., Altshuler, Y.: Social signals in the ethereum trading network. arXiv preprint arXiv:1805.12097 (2018)
33. Wood, G.: Ethereum: A secure decentralised generalised transaction ledger. Ethereum Project Yellow Paper **151**, 1–32 (2014)
34. Ye, C., Huang, J.: Non-classical oscillator model for persistent fluctuations in stock markets. Physica A: Statistical Mechanics and its Applications **387**(5-6), 1255–1263 (2008)