Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Application of Benford's Law on Cryptocurrencies

Jernej Vičič [1,2,*,†] and Aleksandar Tošić [1,3,†]

[1] Faculty of Mathematics Natural Sciences and Information Technologies, University of Primorska, 6000 Koper, Slovenia; aleksandar.tosic@upr.si
[2] Research Centre of the Slovenian Academy of Sciences and Arts, The Fran Ramovš Institute, 1000 Ljubljana, Slovenia
[3] InnoRenew CoE, 6310 Izola, Slovenia
* Correspondence: jernej.vicic@upr.si
† These authors contributed equally to this work.

Abstract: The manuscript presents a study of the possibility of use of Benford's law conformity test, a well proven tool in the accounting fraud discovery, on a new domain: the discovery of anomalies (possibly fraudulent behaviour) in the the cryptocurrency transactions. Blockchain-based currencies or cryptocurrencies have become a global phenomenon known to most people as a disruptive technology, and a new investment vehicle. However, due to their decentralized nature, regulating these markets has presented regulators with difficulties in finding a balance between nurturing innovation, and protecting consumers. The growing concerns about illicit activity have forced regulators to seek new ways of detecting, analyzing, and ultimately policing public blockchain transactions. Extensive research on machine learning, and transaction graph analysis algorithms has been done to track suspicious behaviour. However, having a macro view of a public ledger is equally important before pursuing a more fine-grained analysis. Benford's law, the law of first digit, has been extensively used as a tool to discover accountant frauds (many other use cases exist). The basic motivation that drove our research presented in this paper was to test the applicability of the well established method to a new domain, in this case the identification of anomalous behavior using Benford's law conformity test to the cryptocurrency domain. The research focused on transaction values in all major cryptocurrencies. A suitable time-period was identified that was long enough to present sufficiently large number of observations for Benford's law conformity tests and was also situated long enough in the past so that the anomalies were identified and well documented. The results show that most of the cryptocurrencies that did not conform to Benford's law had well documented anomalous incidents, the first digits of aggregated transaction values of all well known cryptocurrency projects were conforming to Benford's law. Thus the proposed method is applicable to the new domain.

Keywords: cryptocurrency; Benford's law; anomaly detection; method application

Citation: Vičič, J.; Tošić, A. Application of Benford's Law on Cryptocurrencies. *J. Theor. Appl. Electron. Commer. Res.* **2022**, *17*, 313–326. https://doi.org/10.3390/jtaer17010016

Academic Editor: Jani Merikivi

Received: 7 November 2021
Accepted: 8 February 2022
Published: 25 February 2022

1. Introduction

Benford's law [1], also known as the first-digit law, has been widely used as a tool to discover anomalies in various data ranging from accounting fraud detection, stock prices, house prices to electricity bills, population numbers, natural phenomena, death rates and recently so popular COVID-19 cases reports. Cryptocurrencies, also referred to as Blockchain-based currencies or crypto coins, have become a global phenomenon known to most people.Throughout the paper we will rely on the definition presented by [2] (cryptocurrency). A cryptocurrency is in fact quite a narrow, albeit recognizable, description of a subset of an umbrella class of cryptoassets. While still somehow geeky and not understood by most people, banks, governments and many companies are aware of its importance.

Since the inception of Bitcoin, many alternative systems have been developed. Some remain blockchain-based, where transactions are stored and consequently timestamped

in blocks to create a canonical chain through consensus. Others employ a directed acyclic graph based data structures, where there is no single canonical chain. Instead, transactions reference and confirm previous transactions in order to increase the system's throughput by sacrificing some security features. Moreover, transaction structure can be changed to achieve privacy, i.e., using ring signatures in Monero [3]. Regardless of the underlying data structure, consensus mechanism, or network protocol, cryptocurrencies are decentralized and permissionless computer networks that maintain a transparent ledger of transactions. Unlike cryptocurrencies, where a user can have an arbitrary number of wallets (identities), centralized and permissioned systems are easier to monitor, detecting suspicious behaviour or anomalies where approaches are analogue to traditional banking systems, as users are assumed to have a verifiable identity.

A report from The World Economics Forum [4] predicts 10% of the global domestic product to be stored on blockchain based public ledgers. The growing interest has made many developers, research, and innovators dedicate their time in an effort to improve on the existing systems. The effects can be observed through the thousands of cryptocurrencies and networks that exist presently. The growing velocity of these networks further increases the risk for the regulator to protect the consumer and the stability of the financial system. The United Nations Office on Drugs and Crime estimated up to 5% of the global GDP of laundered money [5]. Assuming frauds grow in parallel with the velocity and total value locked in the underlying network, a method for fast and efficient anomaly detection is paramount. However, with the growth of innovation in this space, the techniques employed must search for a generic solution that makes little or no assumptions about the underlying network.

Our approach attempts to provide a technology agnostic tool to analyze open ledgers to alert of potential suspicious behaviour which requires further, more fine-grained analysis. Although more than 12 years have passed since the first transaction of the first cryptocurrency—Bitcoin (BTC) [6]—only the last few years have seen a big enough number of transactions and a large enough time frame for some statistical analysis to be carried out. Our research focused on empirical proof whether Benford law [1], a law of anomalous numbers, could be used in a non-altered form for discovering fraudulent or at least suspicious activity on cryptocurrencies in the same way it is used in standard financial forensics.

Although we could observe the cryptocurrency transactions as just another financial tool that should comply to all the used mechanisms (among them also the Benford law conformity for identifying frauds and other anomalous behavior), there are some properties that must be addressed or at least be observed:

- Mining transactions (mostly with mining pools) for all cryptocurrency assets that are based on the Proof of Work (PoW) [7] consensus mechanism, by which the cryptocurrency blockchain network achieves distributed consensus. Mining pools, where most of the miners are concentrated, pay out rewards to miners based on the computing power contributed. The payouts are mostly scheduled to occur once the miner is owed more than the threshold to save up on transaction fees. As many miners keep the default threshold, many transactions are possibly of the same value;
- Default transaction fees (GAS) are the same. GAS refers to the pricing value required to successfully conduct a transaction or execute a contract on the Ethereum blockchain platform.

The basic idea of the research was to test if Benford's law conformity can be used as a tool to detect anomalies in cryptocurrencies. The paper is structured as follows: Section 2 presents the basic properties of Benford's law and its usages, Section 3 presents the state of the art, followed by Methods and Materials in Section 4. The results are presented in Section 5 and are discussed in Section 6.

2. Benford's Law

Benford's law, also called the Newcomb–Benford law or the first-digit law, is an observation about the frequency distribution of leading digits. The observation was first

discovered by [8] and later rediscovered by [1]. Benford's law defines a fixed probability distribution for leading digits of any kind of numeric data with the following properties [9]:

- Data with values from several distributions;
- Data that has a wide variety in the number of digits (e.g., data with plenty of values in the hundreds, thousands, tens of thousands, etc.);
- A data set that is fairly large, as a rule of a thumb consisting of at least 50–100 observations [10], although usually thousands of observations;
- Data is right-skewed (i.e., the mean is greater than the median), and the distribution has a long right-tail rather than being symmetric;
- Data has no predefined maximum or minimum value (with the exception of a zero minimum).

The distribution of digits is presented in Figure 1; the digit 1 occurs in roughly 30% of the cases, and the other digits follow in a logarithmic curve. It has been shown that this result applies to a wide variety of data sets [9]. Some examples are presented in Section 3. The equation for the distribution of the first digits of observed data is presented in Equation (1).

$$P(d) = log_{10}(d+1) - log_{10}(d) = log_{10}(1+\frac{1}{d}) \quad (1)$$

The quantity $P(d)$ is proportional to the space between d and $d+1$ on a logarithmic scale. Therefore, this is the distribution expected if the logarithms of the numbers (but not the numbers themselves) are uniformly and randomly distributed.

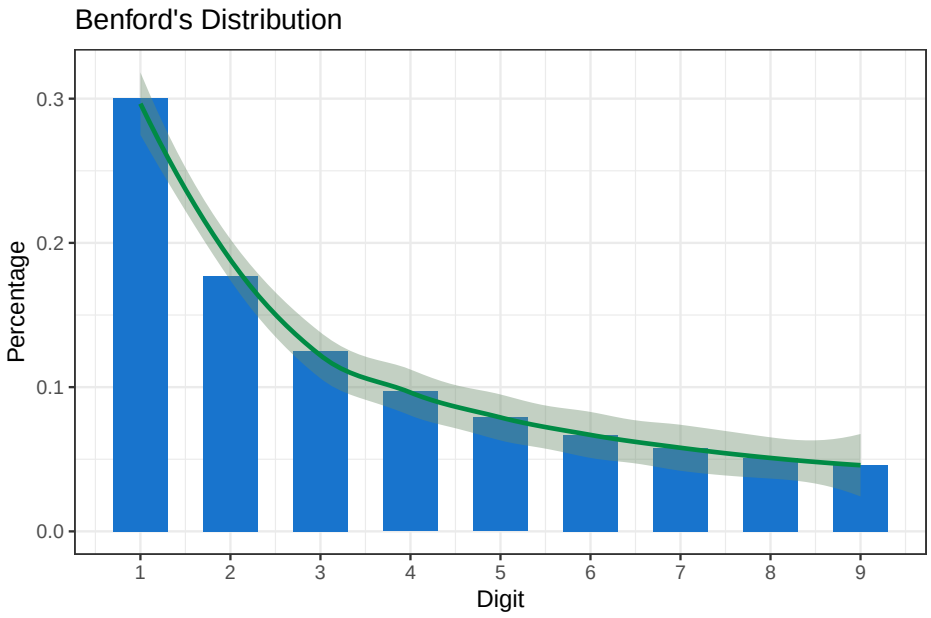

Figure 1. The distribution of digits in accordance to Benford's law [9]. Blue colored bars represent digits that conform to Benford's law.

3. State of the Art

Benford's law has been thoroughly researched and its theoretical grounds have been proved in many scientific papers. The methodology and basic mathematical grounds are discussed in greater detail by [11]. Many researchers have verified for themselves that the law is widely obeyed but have also noted that the popular explanations are not completely satisfying [12]. To the authors' knowledge, there has been no research in using Benford's law as a tool for the detection of anomalies in cryptocurrency transactions.

Benford's law has been extensively used in the accountant fraud detection and prevention, and there has been a lot of research in the area, such as [13,14], who present a literature overview of the area. Ref. [15] introduces Benford's Law and Digital Analysis (analysis of

digit and number patterns of a data set), which can be used as an analytical procedure and fraud detection tool. Ref. [16] presents Benford's law as a simple and effective tool for the detection of fraud. The purpose of the paper is to assist auditors in the most effective use of digital analysis based on Benford's law by identifying data sets, which can be expected to follow Benford's distribution, and presenting types of frauds that would be "detected/not detected" by such analysis. However, there are some research findings that point out some inherent problems that potentially arise in the use of Benford's law in the auditing process such as [17].

The simplicity of Benford's law as a tool allows for a broad range of uses. Ref. [18] examined crime statistics at the USA National, State, and local level in order to test the conformity to Benford's law distribution. Ref. [19] observed the distribution of initial digits of physical constants; however, their results were inconclusive.

One of the more recent researches involving Benford's law is [20]. The authors proposed a test of the reported number of cases of coronavirus disease in 2019 in China with Benford's law and report that the reported numbers of affected people abide to Benford's law.

Ref. [21] presented an overview of identified frauds that can be committed in the cryptocurency paradigm. Identified frauds include Ponzi schemes [22], fake initial coin offering schemes, pump and dump schemes, as well as cryptocurrency theft. Ref. [23] identified the main reasons for frauds and manipulation in cryptocurrencies: lack of consistent regulation, relative anonymity, low barriers of entry, exchange standards, and sophistication. Ref. [24] performed an end-to-end characterization of the counterfeit token in the Ethereum network, targeting Erc20 coins. Ref. [25] aimed to demonstrate that Bitcoin, the most known cryptocurrency, constitutes a substantial danger in terms of criminal enterprise. Ref. [26] presented an economic analysis of money laundering schemes utilizing cryptocurrencies, which aims at providing an answer to the open question of whether cryptocurrencies constitute a driver for money laundering. Ref. [27] proposed an approach to detect illicit accounts on the Ethereum blockchain using well proven machine learning techniques. Recent anomaly detection makes use of machine learning approaches. Support Vector Machines (SVM) were used to detect anomalies in the Bitcoin network [28]. However, the analysis is on the network level, and not on individual transactions. A clustering approach with Random Forest (RF) was used to detect wallets with anomalous behaviour [29]. However, the approach makes assumptions on the underlying structure of transactions to extract the features needed, and thereby lacks generality. A recent study showed that neural networks can be used to detect abnormalities with good stability and effectiveness, but the technique is limited to smart contract platforms, and not general transaction networks. Kamišalić et al. [30] presented a detailed overview of various techniques used for anomaly detection. This highlights the need for a simpler implementation agnostic technique for preliminary screening of public ledgers.

4. Methodology

As mentioned in Section 1, this paper proposes a methodology for identifying out-of-the-ordinary behavior and possibly detect frauds in blockchain-based currency. As such, the purpose is to present scientific grounds that allow feasibility and usefulness of the method as well as to propose a set of usage guidelines and a use case where our hypotheses were confirmed.

Our research experiment started with gathering all transactions on the Ethereum (ETH) network. Ethereum was chosen for these properties: It is one of the biggest cryptocurrencies by market capitalization and number of transactions processed; the network houses multiple cryptocurrencies (tokens) that could be compared directly (this part of the experiment is still open); and it is a well-documented and accessible blockchain. The first preliminary results revealed that transaction values (non-aggregated) of the whole Ethereum network do not conform to Benford's law [1] as is presented in Figure 2. Blue color depicts the leading digits that conform to Benford's law, red color depicts the non-conforming digits. The reasoning is further discussed in Section 4.1.

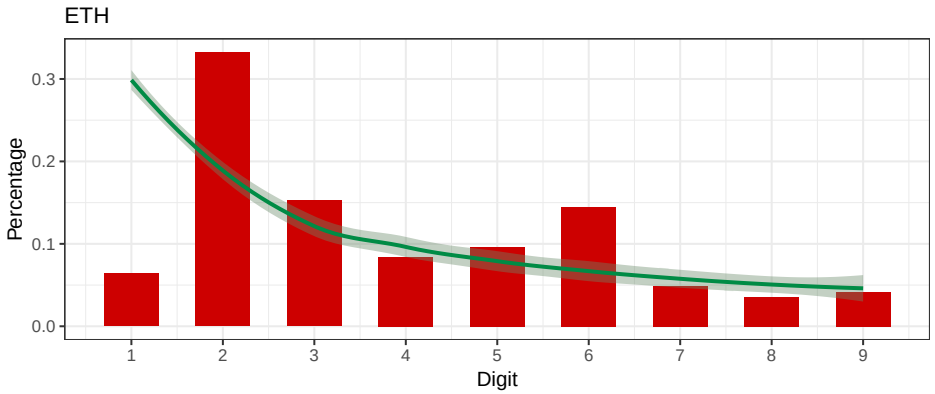

Figure 2. The leading digits of all ETH transaction values do not conform to Benford's law. The daily aggregated values conform to the same metric (see Figure 3), leading to a possible conclusion that there are too many automatic transactions in the network, but the aggregated values avoid this effect.

Although this does not mean that there was any artificial manipulation or any other kind of anomaly, we investigated further. According to [31] Benford's law metric can be used to achieve similar goals on aggregated data. We explored the same phenomenon on aggregated values (number of transactions in an observed period, aggregated transaction values, ...). Most of the aggregated values conform to Benford's law according to goodness of fit chi square ($\tilde{\chi}^2$) test [32], which in most literature, such as [16], is considered as a suitable tool to test Benford's law conformity. We extended our research to all major cryptocurrencies with enough transactions in the selected time-period.

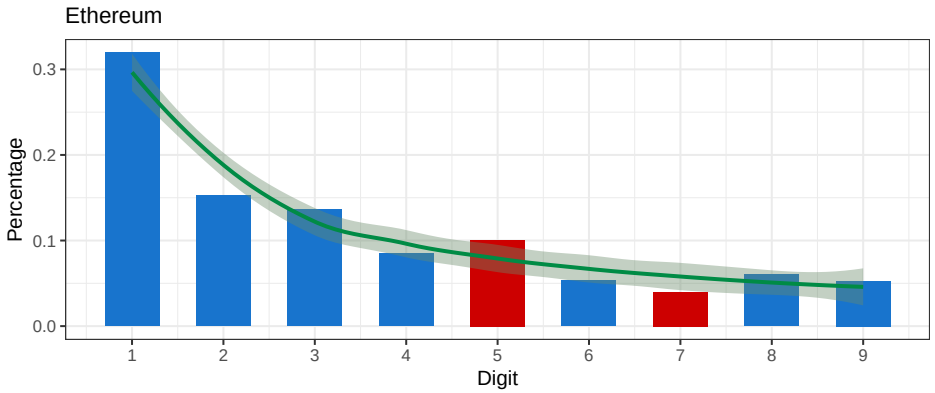

Figure 3. The leading digits of daily aggregated ETH transaction values in USD conform to Benford's law. Blue colored bars represent digits that conform and red colored bars represent digits that do not conform to Benford's law.

4.1. Methods

The observation sets need to conform to all the basic prerequisites for Benford's law as described in Section 2. This is the agenda for the executed research:

- Take all major cryptocurrencies into consideration;
- Express all aggregated daily transactions in one currency—we selected USD ($) as the most used fiat currency in comparisons;
- Select a viable observation period:
  - Starting date for each currency was the date of the first successful transaction;
  - Ending date for the observation period was set long enough into the past so that the frauds or abnormal behavior were well documented (in the forms of law-suits, scandals, vanished cryptocurrencies, well-documented special properties of specific currencies). We selected the year as the end of 2018, almost three years in the past;

- A long enough observation period that makes Benford's law conformity observation feasible (as presented in Section 2). In the body of surveyed literature, the sample size varies from 200 [33] to a few hundred thousand. We opted for doubling the minimum sample size—selecting all cryptocurrencies with 400 or more transaction days;
- Perform the MAD test [34] and classify all the cryptocurrencies according to [35] and visually observe all conformity graphs;
- Perform a literature review for all the currencies that do not conform to Benford's law and establish if there are any abnormalities documented for the selected time frame.

Testing conformity to Benford's law distribution has been done with many goodness of fit tests ranging from Pearson's Chi squared [36], Kolmogorov-Smirnov D statistics [37], Freedman's modification of Watson U^2 statistics [38], euclidean distance d statistics, and many others. However, no real data will ever follow the exact distribution; hence, most analysis supplements statistical testing with graphical representations that help in pointing out suspicious patterns in the data for further investigation. Additionally, different tests have different reactions on sample sizes. The Chi square test suffers from an excess power problem in that when the number of observations becomes large (above 5000 records estimated by [35]) it becomes more sensitive to insignificant spikes, leading to the conclusion that the data does not conform. Ref. [39] suggested that some statistical tests can render misleading results when applied to large number of observations. On the other hand, ref. [40] conclude that the Mean Absolute Deviation MAD test [34] is reliable with as low as 200 observations (as additional safety measure, we opted doubling that value to 400 in our experiment). Ref. [41] proposed the Mantissa Arc test, which is a very interesting geometrical test. Unfortunately, it tolerates little deviation from Benford's distributions.

Ref. [35] concluded that the best test is Mean Absolute Deviation (MAD), and a lot of the state-of-the art literature agrees with this proposal. Ref. [35] also presents a list of thresholds to classify the observed conformity:

- Conformity (0.000);
- Acceptable conformity (0.006);
- Marginally acceptable conformity (0.012);
- Nonconformity (0.015 and above).

The adapted MAD is used to measure the average deviation between the heights of the bars and the Benford line. The higher the MAD, the lower the conformity. We opted to perform conformity tests using all three of the aforementioned tests as our sample sizes are well within the acceptable ranges. All presented statistical tests are also supplemented with graphical representations; the results are presented in Section 5.

The Criteria That the Objects under Scrutiny Must Meet

Select a big enough set of aggregated data that conforms to Benford's law prerequisites described in Section 2. Observing only ledgers, the prerequisites that must be met are:

- The ledger must have support querying for transactions that contain the sending address, receiving address, amount, and timestamp;
- The assets being transferred must be denominated in any universally comparable form (any fiat currency (i.e., US Dollars) meets this criterion) at the time of transfer.

Count leading digits and perform Mean Absolute Deviation (MAD) conformity [35] on the gathered data. Plot simple bar charts with the numbers for each leading digit and visually and manually observe the distribution. If the data does not conform to Benford's law, investigate further.

4.2. Materials

DataHub cryptocurrency datasets (DataHub cryptocurrency datasets: https://datahub.io/cryptocurrency accessed on 1 March 2021) hosts daily aggregated data about all transac-

tions on all crypto coin networks from the first mined block on the Bitcoin network till the end of 2018. As such, it presents the perfect data source for our research. The problem that arises is how to get more recent data. The problem is further discussed in Section 6.

The data that support the findings of this study are openly available on Zenodo (Zenodo: https://zenodo.org/record/4682976 accessed on 1 January 2022, doi:10.5281/zenodo.4682976).

5. Results

This section presents the results of the experiment following the methodology from Section 4. All the figures in this section have the same format: a graph showing the distribution of leading digits. Red colored bars represent suspect values, which skew the distribution the most. Suspects are classified where the mean absolute deviation is above the threshold of 4. The threshold can be adjusted to increase the sensitivity. Suspects are useful as a starting point for further investigation in the case of nonconformity.

The time interval selected was between 2009 and 2018. Most of the cryptocurrencies were in an early development phase without a use-case or product, and consequently the amount of transactions recorded was negligible. Table 1 presents all cryptocurrencies that conformed to the prerequisites presented in Sections 2 and 4. The most discriminating factor in this phase was the minimum number of observations, which was set to 400 days (roughly double the minimal number of observations for Benford's law to be meaningful). This property eliminated all currencies that were started later than the last quarter of 2017. Each cryptocurrency is presented by its name and the ticker, number of observations (equal to the number of days), starting and ending date of the observation period and all the values from Benford's law conformance test. The currencies were grouped into four groups according to [35] and were also sorted according to this grouping from best to worst conformance.

All non-conformant cryptocurrencies were thoroughly observed and a list of publicly announced anomalies and even frauds was compiled for each of these cryptocurrencies. The two best performing and two cryptocurrencies with the biggest market cap were also observed in details. The results are presented in the remainder of the section. All the other cryptocurrencies can be further analyzed using the available accompanying data (Zenodo: https://zenodo.org/record/4682976 accessed on 1 January 2022, doi:10.5281/zenodo.4682976) in the raw aggregated data form, a list of Benford's law conformity values and charts.

Two "best conforming" cryptocurrencies, Ethereum classic (ETC) and Vertcoin (VTC), both still respectable projects, were classified as "Close conformity". The two biggest blockchain platforms regarding market capitalization, Bitcoin (BTC) and Ethereum (ETH), were classified as "Acceptable conformity" and "Marginally acceptable conformity", respectively. Figure 4 shows Benford's law conformance chart for further visual examination for all four cryptocurrencies.

Six of the currencies from Table 1 were classified as "non-conformant" to Benford's law: EOS (EOS), TENX token (TENX), Veritaseum (VERI), Basic Atention Token (BAT), PIVX (PIVX), and Dogecoin (DOGE). Each of the cryptocurrencies from this list will be presented and discussed.

Table 1. Conformity tests for all major cryptocurrencies in the observed time-period with more than 400 days of transactions on the blockchain. The records are sorted according to MAD Conformity column, from close conforming to nonconforming.

Currency	Obs.	Pearson's Chi-Squared Test		Mantissa Arc Test		MAD	MAD Conformity	Distortion Factor	Start Date	End Date
		X-Squared	*p*-Value	L2	*p*-Value					
Ethereum Classic (ETC)	750	1.766027	0.9873638	0.0000861	0.9374726	0.00351481	Close	−0.1409321	2015-07-30	2018-08-12
Vertcoin (VTC)	1666	7.30948	0.5036398	0.000165673	0.7588044	0.005795195	Close	−1.621333	2014-01-10	2018-08-12
Metal (MTL)	400	5.15115	0.7413057	0.001522525	0.543889	0.01089584	Acceptable	−0.1257945	2017-06-29	2018-08-12
Status (SNT)	411	7.692396	0.4640798	0.001050824	0.6492816	0.01005221	Acceptable	−1.560401	2017-06-19	2018-08-12
Aragon (ANT)	452	5.696092	0.6812311	0.005388913	0.08752867	0.01078389	Acceptable	3.448495	2017-05-15	2018-08-12
Waves (WAVES)	603	5.14964	0.7414692	0.003501951	0.1210349	0.008216651	Acceptable	−1.721726	2016-06-02	2018-08-12
Iconomi (ICN)	658	10.17673	0.2528404	0.0008252317	0.5810012	0.0104604	Acceptable	0.8235436	2016-09-30	2018-08-12
NEO (NEO)	665	3.823118	0.8727192	0.0008927334	0.5522979	0.006478303	Acceptable	1.316035	2016-09-09	2018-08-12
Lisk (LSK)	811	11.45478	0.1772377	0.001803885	0.231552	0.009606102	Acceptable	3.072645	2016-04-06	2018-08-12
Stellar (XLM)	1009	9.622045	0.2925614	0.002200075	0.1086226	0.007992198	Acceptable	1.221268	2014-08-05	2018-08-12
Verge (XVG)	1387	8.300241	0.4047048	0.002115592	0.05316656	0.007575786	Acceptable	−2.84182	2014-10-09	2018-08-12
MaidSafeCoin (MAID)	1560	10.43771	0.2356377	0.003279288	0.006001835	0.007513696	Acceptable	3.73407	2014-04-22	2018-08-12
Dash (DASH)	1641	5.958045	0.6519316	0.001418531	0.09750916	0.00621291	Acceptable	0.8615983	2014-01-19	2018-08-12
DigiByte (DGB)	1649	25.9	0.00111	0.003.21	0.005	0.01088511	Acceptable	−2.4136	2014-01-10	2018-08-12
Bitcoin (BTC)	1933	30.8193	0.0001512958	0.0006696828	0.2740357	0.01158613	Acceptable	5.881506	2013-04-28	2018-08-12
Gnosis (GNO)	468	8.754344	0.3634412	0.006937894	0.03889326	0.01312756	Marginally acc.	1.135551	2017-04-18	2018-08-12
Golem (GLM)	633	11.07461	0.1975074	0.003690431	0.0967096	0.0129236	Marginally acc.	6.131378	2016-11-11	2018-08-12
Zcash (ZEC)	653	20.82315	0.007632357	0.001029657	0.5104994	0.01293599	Marginally acc.	−0.9372237	2016-10-28	2018-08-12
Decred (DCR)	915	17.6832	0.02373108	0.0005975181	0.5788401	0.01375337	Marginally acc.	−1.586765	2016-02-08	2018-08-12
Ethereum (ETH)	1102	25.77399	0.00115	0.000378	0.658996	0.01482756	Marginally acc.	−0.08431323	2015-08-07	2018-08-12
NEM (XEM)	1230	27.13364	0.0006703807	0.008295528	0.000037	0.01417723	Marginally acc.	3.19854	2015-03-29	2018-08-12
Tether (USDT)	1258	34.91683	0.0000277	0.0138	0.00000003	0.01391653	Marginally acc.	−5.969747	2014-10-06	2018-08-12
EOS (EOS)	401	15.36398	0.05244271	0.003494984	0.2462301	0.0200535	Nonconformity	−2.819878	2017-06-20	2018-08-12
TENX token (TENX)	402	10.5	0.234	0.00808	0.0389	0.01539412	Nonconformity	−7.119347	2017-06-27	2018-08-12
Veritaseum (VERI)	431	11.32151	0.1841391	0.01211339	0.005402612	0.01726905	Nonconformity	−1.603899	2017-04-25	2018-08-12
Basic Atention T. (BAT)	438	19.05523	0.01456707	0.01293943	0.003456598	0.02196946	Nonconformity	0.2319942	2017-05-29	2018-08-12
PIVX (PIVX)	903	28.08438	0.0004584671	0.01199764	0.0000197	0.01890993	Nonconformity	−7.031687	2016-01-30	2018-08-12
Dogecoin (DOGE)	1702	83.1755	1.1210^{-14}	0.02422157	1.2510^{-18}	0.0214206	Nonconformity	−9.527495	2013-12-08	2018-08-12

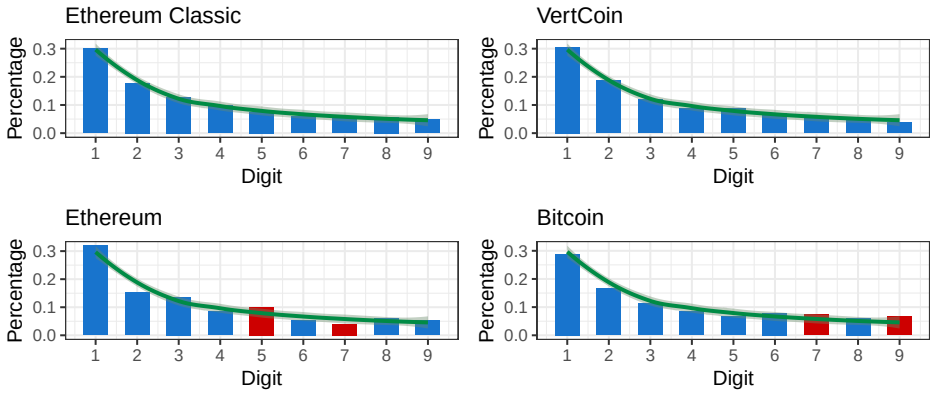

Figure 4. The two best conforming (ETC) and (VTC) currencies with "Close conformity" and the two biggest cryptocurrencies (BTC)—"Acceptable conformity" and (ETH)—"Marginally acceptable conformity" for aggregated value in USD transaction history.

5.1. TENX Token (TENX)

Figure 5 shows the TENX aggregated transactions and the conformance to Benford's law. The MAD value, a well documented Wirecard scandal (Crypto.com, TenX crypto debit cards were frozen following the Wirecard scandal: https://decrypt.co/33695/crypto-debit-cards-frozen-following-wirecard-scandal accessed on 1 March 2021) shows a possible reason for non-conformity.

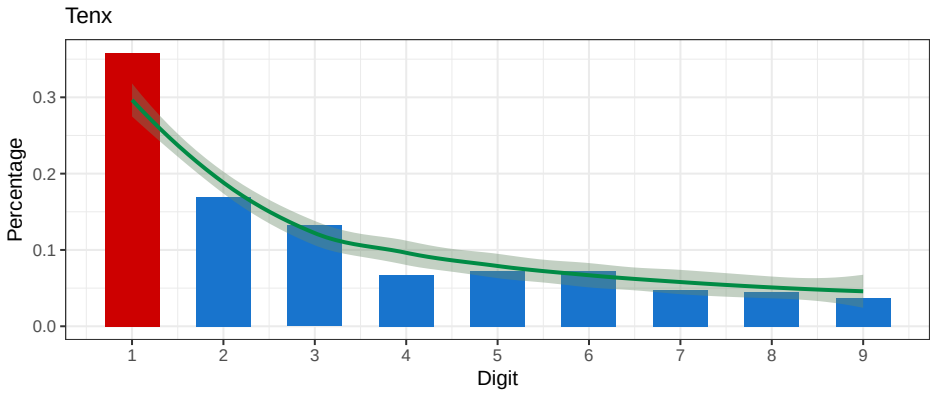

Figure 5. TENX aggregated transactions and the conformance to Benford's law. Digit 1 overflows, digit 4 (almost) underflows. Overall, the daily aggregated transaction values do not conform.

5.2. Veritaseum (VERI)

Figure 6 shows the VERI aggregated transactions and the conformance to Benford's law.

The U.S. Securities and Exchange Commission (SEC) said it has reached a settlement with Reggie Middleton, organizer of the fraught $14.8 million Veritaseum (VERI) initial coin offering (ICO) (Analysis of the Veritaseum Scam: https://steemit.com/money/@financialcritic/analysis-of-the-veritaseum-scam accessed on 1 March 2021). The case was closed on October 2019, but the frauds were committed well within the observation period of our research.

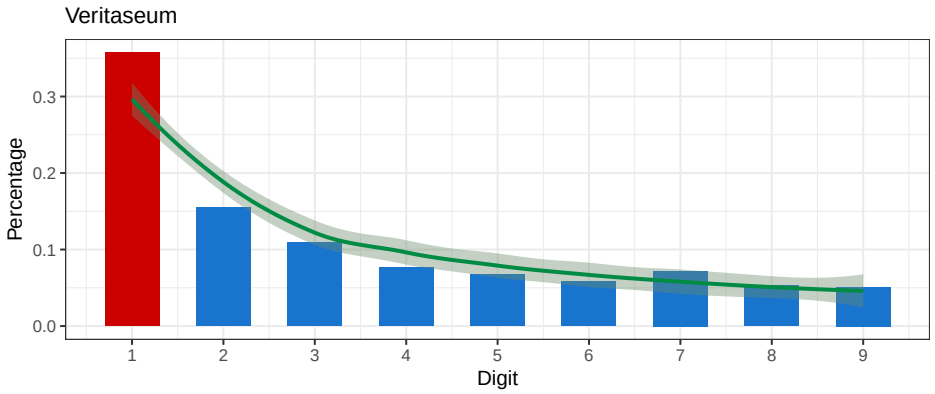

Figure 6. VERI aggregated transactions and the conformance to Benford's law. Digit 1 overflows. Overall the daily aggregated transaction values do not conform.

5.3. Dogecoin (DOGE)

Figure 7 shows the DOGE aggregated transactions and the conformance to Benford's law. The coin was introduced as a satire initially in December 2013 and included an image of the Doge meme as its logo. The author of this coin/crypto currency revealed this motivation publicly. Some properties showing the soundness of our decision are as follows:

- On the 24 September 2018 (a randomly chosen date on a working day at the end of our observation period): the last tweet from the official Tweeter account on 14 July 2018 (80 days) (Dogecoin twitter account: https://twitter.com/Dodgecoin accessed on 1 March 2021);
- Fun and friendly internet currency, the dogecoin logo is a dog from a meme;
- 24 h trading volume on all exchanges according to CoinCodex (Concodex: https://coincodex.com/crypto/XXX/exchanges/ accessed on 1 March 2021) was USD 42.51 million dollars.

In the last years Dogecoin has gained a lot of positive reputation as being a "lost cause" founding platform, and, especially in 2021, the value of the coin has seen a rapid increase in price with the help of celebrity exposure [42]. However, these recent developments were excluded from our analysis as we fixed the observation period from the start of the crypto-assets till the end of 2018.

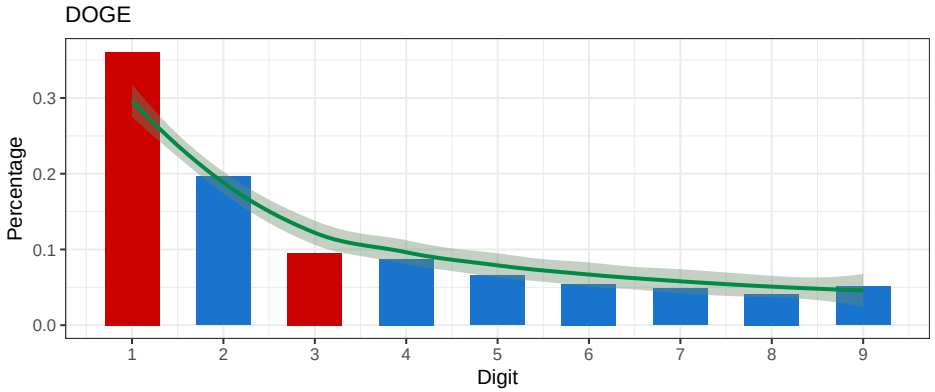

Figure 7. DOGE aggregated transactions and the conformance to Benford's law. Digit 1 overflows, digit 3 underflows. Overall the daily aggregated transaction values do not conform.

5.4. Basic Attention Token (BAT)

Figure 8 shows the BAT aggregated transactions and the conformance to Benford's law. The transactions of the BAT coin are mostly automatically generated as this coin is the basis of a digital marketing platform that periodically rewards users for participation, and as such break Benford's law prerequisites.

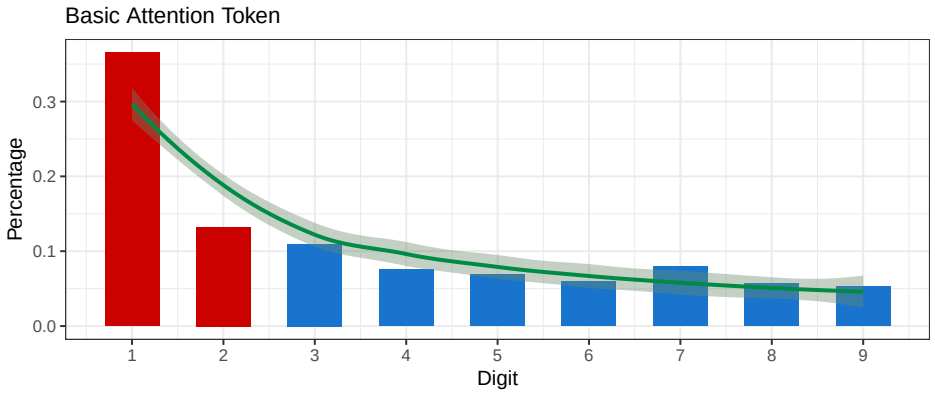

Figure 8. BAT aggregated transactions and the conformance to Benford's law. Digit 1 overflows, digit 2 underflows, digit 7 (almost) overflows. Overall the daily aggregated transaction values do not conform.

5.5. PIVX (PIVX)

Figure 9 shows the PIVX aggregated transactions and the conformance to Benford's law. There was no scandal reported for the PIVX project in the observation period (in fact, the authors could not find any notable anomaly for this cryptocurrency). The only speculation that the authors could give is that the PIVX network relies on anonymous transactions that could be used to hide anomalies.

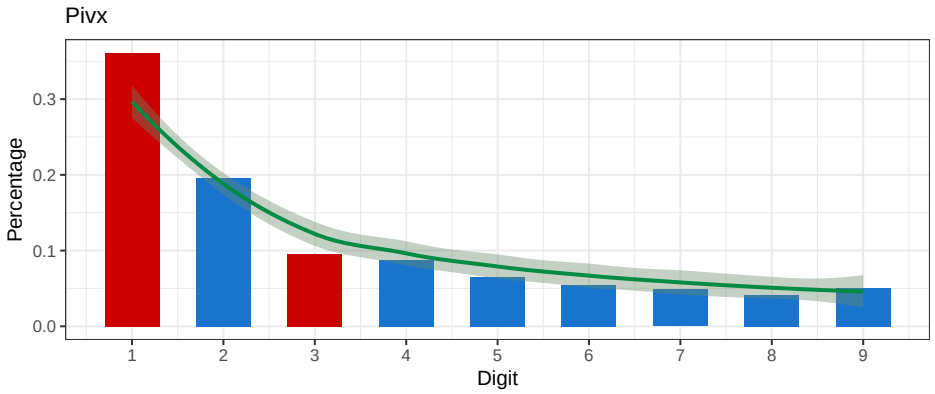

Figure 9. PIVX aggregated transactions and the conformance to Benford's law. Digit 1 overflows, digit 3 underflows. Overall the daily aggregated transaction values do not conform.

5.6. EOS (EOS)

Figure 10 shows the EOS aggregated transactions and the conformance to Benford's law. EOS is regarded as a valid project and survived until 2021. The only drawback is that in 2018, the project was in the starting phase and the backing capital risen by the backers of the project was an order of magnitude bigger than what the proposed project promised to accomplish ("Why EOS Failed to Kill Ethereum: The Fatal Flaw of Centralization in a Decentralized Market": https://coincodex.com/article/10454/why-eos-failed-to-kill-ethereum-the-fatal-flaw-of-centralization-in-a-decentralized-market/ accessed on 1 March 2021).

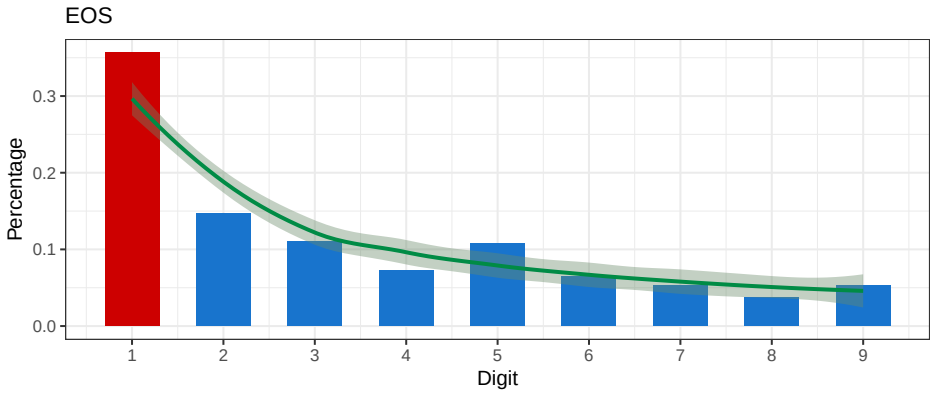

Figure 10. TENX aggregated transactions and the conformance to Benford's law. Digit 1 overflows, digits 2 and 4 (almost) underflow. Overall the daily aggregated transaction values do not conform.

5.7. Additional Currencies

An examination of all remaining cryptocurrencies that did not meet the criteria presented in Section 4, mainly due to the lack of data, show additional cases that support the validity of the presented method. By lowering the requirement for the minimum number of observations to 300 days, we can observe additional cryptocurrencies that do not conform to Benford's law that have documented scams and scandals attributed to the observation period, such as: the Enigma (ENG) (Enigma Ethereum marketplace was hijacked, its investors duped by phishing scam: https://www.zdnet.com/article/enigma-ethereum-marketplace-hijacked-by-attackers/ accessed on 1 March 2021); SALT (SALT) (SALT COIN EXIT SCAM! Massive selloff predicted by Morgan Stanley: https://www.youtube.com/watch?v=E2iNt3Z6qaY accessed on 1 March 2021); and Waltonchain (WTC) (Monumentall stupid tweet blows up in blockchain company's face: https://mashable.com/2018/02/28/waltonchain-twitter-scam-wtc/?europe=true accessed on 1 March 2021).

6. Discussion and Future Work

The main goal of the presented research was to test the applicability of Benford's law to the cryptocurrency transaction networks as a preliminary screening tool. The research focused on some well-documented anomalies and frauds from the past and compared the proposed metric on proven ecosystems that performed normally in the same time period. We focused on the time period between 2009 (time of the first transaction on the Bitcoin network) and 2018, as there were already enough transactions to meet all of Benford's law prerequisites, but also enough time had passed so that the anomalies and frauds had already emerged to the public.

The results show that the proposed method is suitable for the proposed domain. All the big blockchain platforms by market capitalization that were not biased by any big scandal or lawsuit and that are still functioning three years after the observation timeframe, such as Bitcoin (BTC), Ethereum (ETH), or OmiseGo (OMG), conform to Benford's law. However, failing to comfort to Benfords distribution does not necessarily imply fraud. The method can produce false positives in the form of non-conformity of a cryptocurrency and no particular fraudulent reason can be found. This can result from the nature of the transactions of the observed currency. The method does not find the actual anomaly, but it can be used as a preliminary screening that should always lead into fine-grained methods such as Machine Learning methods and graph-based searching. The inspection of the six cryptocurrencies that were classified as non-conforming to Benford's law revealed three currencies with well-documented anomalies: two (TENX and VERI) were tainted by scandals and lawsuits and one (DOGE) was invented as a joke—and in the first years it was regarded so. As an additional observation, Dogecoin is now a respected cryptocurrency and in the last year grew to USD $50B market capitalization. The method is obviously not

suitable to predict the future of an observed cryptocurrency. The transactions of the BAT coin are mostly automatically generated, as this coin is the basis of a digital marketing platform. The two remaining cryptocurrencies that were identified by the method as possible candidates for anomalous behaviour were EOS and PIVX, and although we could speculate to some extension why these two did not conform to Benford's law, the results are inconclusive.

All major cryptocurrencies that existed in the selected time-frame (2009–2018) were tested for the conformity to Benford's law. The data availability statement is presented in Section 4.2.

Future work, which is already underway, will focus on newer data. One such possible source has already been identified: Kaggle (Cryptocurrency Historical Prices: https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory accessed on 1 March 2021). Another open issue that can be tackled with the same methodology is a comparison of all ERC20 tokens [43]. Ethereum-based cryptocurrencies were selected to ensure a common (thus fair) technical basis—all these cryptocurrencies use the same technological platform, so all possible reasons for differences that arise from basic technology are eliminated.

Author Contributions: Conceptualization, A.T. and J.V.; methodology: A.T. and J.V.; software, A.T.; validation, A.T. and J.V.; formal analysis, A.T. and J.V.; investigation, A.T. and J.V.; funding acquisition and resources, J.V.; data curation, A.T.; writing–original draft preparation, A.T. and J.V.; writing–review and editing, A.T. and J.V.; visualization, A.T. and J.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by H2020 grant number 739574 and by the Slovenian Research Agency (ARRS) grant number J2-2504.

Institutional Review Board Statement: The data gathering process did not involve the use of human subjects.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are openly available on Zenodo (Zenodo: https://zenodo.org/record/4682976 accessed on 1 January 2022, doi:10.5281/zenodo.4682976).

Acknowledgments: The authors gratefully acknowledge the European Commission for funding the InnoRenew project (Grant Agreement #739574) under the Horizon2020 Widespread-Teaming program and the Republic of Slovenia (Investment funding of the Republic of Slovenia and the European Regional Development Fund). They also acknowledge the Slovenian Research Agency ARRS for funding the project J2-2504.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Benford, F. The law of anomalous numbers. *Proc. Am. Philos. Soc.* **1938**, *78*, 551–572.
2. Lansky, J. Possible state approaches to cryptocurrencies. *J. Syst. Integr.* **2018**, *9*, 19–31. [CrossRef]
3. Noether, S. Ring SIgnature Confidential Transactions for Monero. *IACR Cryptol. ePrint Arch.* **2015**, *2015*, 1098.
4. Mettler, M. Blockchain technology in healthcare: The revolution starts here. In Proceedings of the 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), Munich, Germany, 14–16 September 2016; pp. 1–3.
5. Campbell-Verduyn, M. Bitcoin, crypto-coins, and global anti-money laundering governance. *Crime Law Soc. Chang.* **2018**, *69*, 283–305. [CrossRef]
6. Nakamoto, S. Bitcoin Whitepaper. Technical Report. 2008. Available online: Bitcoin.org (accessed on 1 March 2021).
7. Jakobsson, M.; Juels, A. Proofs of work and bread pudding protocols. In *Secure Information Networks*; Jakobsson, M., Juels, A., Eds.; Springer: Leuven, Belgium, 1999; pp. 258–272.
8. Newcomb, S. Note on the Frequency of Use of the Different Digits in Natural Numbers. *Am. J. Math.* **1881**, *4*, 39–40. [CrossRef]
9. Singleton, T.W. IT Audit Basics: Understanding and Applying Benford's Law. *Isaca J.* **2011**, *3*, 6.
10. Kenny, D.A. Measuring Model Fit. 2015. Available online: http://davidakenny.net/cm/fit.htm (accessed on 1 March 2021).
11. Berger, A.; Hill, T.P. A basic theory of Benford's Law. *Probab. Surv.* **2011**, *8*, 1–126. [CrossRef]
12. Fewster, R.M. A Simple Explanation of Benford's Law. *Am. Stat.* **2009**, *63*, 26–32. [CrossRef]
13. Kumar, K.; Bhattacharya, S. Detecting the dubious digits: Benford's law in forensic accounting. *Significance* **2007**, *4*, 81–83. [CrossRef]

14. Nigrini, M.J. Audit sampling using Benford's law: A review of the literature with some new perspectives. *J. Emerg. Technol. Account.* **2017**, *14*, 29–46. [CrossRef]
15. Drake, P.D.; Nigrini, M.J. Computer assisted analytical procedures using Benford's Law. *J. Account. Educ.* **2000**, *18*, 127–146. [CrossRef]
16. Durtschi, C.; Hillison, W.; Pacini, C. The effective use of Benford's law to assist in detecting fraud in accounting data. *J. Forensic Account.* **2004**, *5*, 17–34.
17. Cleary, R.; Thibodeau, J.C. Applying Digital Analysis Using Benford's Law to Detect Fraud: The Dangers of Type I Errors. *Audit. J. Pract. Theory* **2005**, *24*, 77–81. [CrossRef]
18. Hickman, M.J.; Rice, S.K. Digital Analysis of Crime Statistics: Does Crime Conform to Benford's Law? *J. Quant. Criminol.* **2010**, *26*, 333–349. [CrossRef]
19. Burke, J.; Kincanon, E. Benford's law and physical constants: The distribution of initial digits. *Am. J. Phys.* **1991**, *59*, 952. [CrossRef]
20. Zhang, J. Testing Case Number of Coronavirus Disease 2019 in China with Newcomb-Benford Law. *arXiv* **2020**, arXiv:2002.05695.
21. Baum, S.C. Cryptocurrency Fraud: A Look into the Frontier of Fraud. Ph.D. Thesis, Georgia Southern University, Statesboro, GA, USA, 2018.
22. Zuckoff, M. *Ponzi's Scheme: The True Story of a Financial Legend*; Random House Incorporated: New York, NY, USA, 2006.
23. Twomey, D.; Mann, A. Fraud and manipulation within cryptocurrency markets. In *Corruption and Fraud in Financial Markets: Malpractice, Misconduct and Manipulation*; Alexander, C., Cumming, D., Eds.; Wiley: Hoboken, NJ, USA, 2020; pp. 205–250.
24. Gao, B.; Wang, H.; Xia, P.; Wu, S.; Zhou, Y.; Luo, X.; Tyson, G. Tracking Counterfeit Cryptocurrency End-to-end. *Proc. ACM Meas. Anal. Comput. Syst.* **2020**, *4*, 1–28. [CrossRef]
25. Brown, S.D. Cryptocurrency and criminality: The Bitcoin opportunity. *Police J.* **2016**, *89*, 327–339. [CrossRef]
26. Brenig, C.; Müller, G. *Economic Analysis of Cryptocurrency Backed Money Laundering*; ECIS 2015 Completed Research Papers; Association for Information Systems: Atlanta, GA, USA, 2015; pp. 1–18.
27. Farrugia, S.; Ellul, J.; Azzopardi, G. Detection of illicit accounts over the Ethereum blockchain. *Expert Syst. Appl.* **2020**, *150*, 113318. [CrossRef]
28. Sayadi, S.; ben Rejeb, S.; Choukair, Z. Anomaly Detection Model Over Blockchain Electronic Transactions. In Proceedings of the 2019 15th International Wireless Communications Mobile Computing Conference (IWCMC), Tangier, Morocco, 24–28 June 2019; pp. 895–900. [CrossRef]
29. Baek, H.; Oh, J.; Kim, C.Y.; Lee, K. A Model for Detecting Cryptocurrency Transactions with Discernible Purpose. In Proceedings of the 2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), Zagreb, Croatia, 2–5 July 2019; pp. 713–717. [CrossRef]
30. Kamišalić, A.; Kramberger, R.; Fister, I. Synergy of Blockchain Technology and Data Mining Techniques for Anomaly Detection. *Appl. Sci.* **2021**, *11*, 7987. [CrossRef]
31. Shi, J.; Ausloos, M.; Zhu, T. Benford's law first significant digit and distribution distances for testing the reliability of financial reports in developing countries. *Phys. A Stat. Mech. Its Appl.* **2018**, *492*, 878–888. [CrossRef]
32. Fischer, R.A. *Statistical Methods for Research Workers*; Oliver and Boyd: Edinburgh, UK, 1925.
33. Carslaw, C.A. Anomalies in income numbers: Evidence of goal oriented behavior. *Account. Rev.* **1988**, *63*, 321–327.
34. Gorard, S. Revisiting a 90-year-old debate: The advantages of the mean deviation. *Br. J. Educ. Stud.* **2005**, *53*, 417–430. [CrossRef]
35. Nigrini, M.J.M.J. *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*; Wiley: Hoboken, NJ, USA, 2012; p. 352.
36. Pearson, K.X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1900**, *50*, 157–175. [CrossRef]
37. Berger, V.W.; Zhou, Y. *Kolmogorov–Smirnov Test: Overview*; Wiley Statsref: Statistics Reference Online: Hoboken, NJ, USA, 2014.
38. Freedman, L.S. Watson's UN2 statistic for a discrete distribution. *Biometrika* **1981**, *68*, 708–711. [CrossRef]
39. Nigrini, M. Digital Analysis Using Benford's Law: Tests and Statistics for Auditors. *EDPACS* **2001**, *28*, 1–2. [CrossRef]
40. Druică, E.; Oancea, B.; Vâlsan, C. Benford's law and the limits of digit analysis. *Int. J. Account. Inf. Syst.* **2018**, *31*, 75–82. [CrossRef]
41. Alexander, J.C. Remarks on the Use of Benford's Law. 2009. Available online: http://dx.doi.org/10.2139/ssrn.1505147 (accessed on 1 March 2021).
42. Livni, E. Serious money is flowing to the joke cryptocurrency Dogecoin. *New York Times*, 2 August 2021; pp. 1–2.
43. Somin, S.; Gordon, G.; Altshuler, Y. Network analysis of erc20 tokens trading on ethereum blockchain. In *International Conference on Complex Systems*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 439–450.