Anonymity in the Bitcoin Network

Kinga Kądziołka

In the electronic system of payment with the Bitcoin cryptocurrency, entering personal details of the parties in the transaction is not necessary as funds are transferred between addresses[111]. Moreover, the given user may use different addresses for different transactions, which additionally makes it difficult to refer all addresses used by one user with the user himself/herself. High anonymity may attract prospective offenders to this method of payment, e.g. dealing with drugs, money laundering or financing terrorism. Organized crime groups may use virtual currency to hide the actual source of funds. The Silk Road website was an example of these types of activities, where drugs were offered for sale for cryptocurrency.

This paper deals with the issue of identification (based on the Benford's distribution) of non-typical transactions in the Bitcoin network. The Benford's law is used to determine presence the digits (or their sequences) which look unusual in the analysed data set. It is applied, among others, in detecting irregularities in financial operations, tax returns and financial statements. For the sample selected addresses of the Bitcoin wallets, the Benford's distribution will be compared with the distribution of frequencies of occurrence of individual digits in the first most significant position in the amounts of transactions related to these addresses.

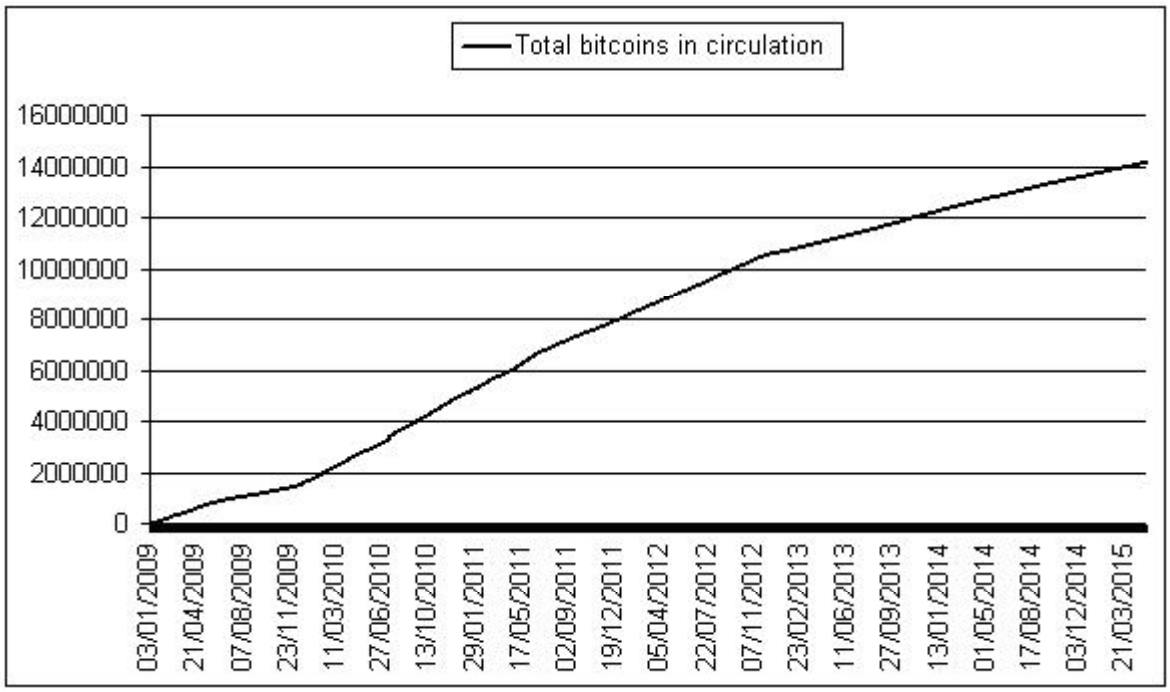
What is Bitcoin?

Bitcoin was described in 2009 by a person (or a group of persons) with the pseudonym Satoshi Nakamoto[112]. It may be understood as a "system of virtual currency functioning within the peer-to-peer payment model, which connects two parties of a transaction without participation of financial institutions[113]". Bitcoin is also understood as a "decentralised digital currency whose units are created in the internet[114]". Bitcoins may be bought directly from the person which has them or in an internet auction or in internet exchange, such as Bitstamp.net, Bitcoin.de or Bitcurex.com. Bitcoins may be

[111] The address in the Bitcoin network is a series of 27-34 alphanumerical characters. It begins with 1 or 3.
[112] S. Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System*, 2009, http://bitcoin.org/bitcoin.pdf, 11.03.2015
[113] Quote: [Perez K., Urbaniak M., *Bitcoin - wirtualny eksperyment czy waluta przyszłości?*, "Ruch prawniczy, ekonomiczny i socjologiczny", 4/2013, p. 164], Peer-to-peer is "wide-area network architecture. In this model, all users are equal and connect directly to other computers in the network" (quote: [Szymankiewicz M., *Bitcoin Wirtualna waluta internetu*, Helion, Gliwice 2014, p. 38]).
[114] Quote: [M. Szymankiewicz, *Bitcoin. Wirtualna waluta Internetu...*, op. cit., p. 21].

also won as awards for computation power made available by users in the so-called bitcoin mining process. This process is described in detail by M. Szymankiewicz (2014). The maximum number of bitcoins allowed in circulation is 21 million. Bitcoin (BTC) is divisible to eighth decimal places. The smallest indivisible part of the bitcoin is called satoshi. 1 satoshi = 0.00000001 BTC. Over 14 million bitcoins are currently in circulation.

Fig. 1 Total bitcoins in circulation

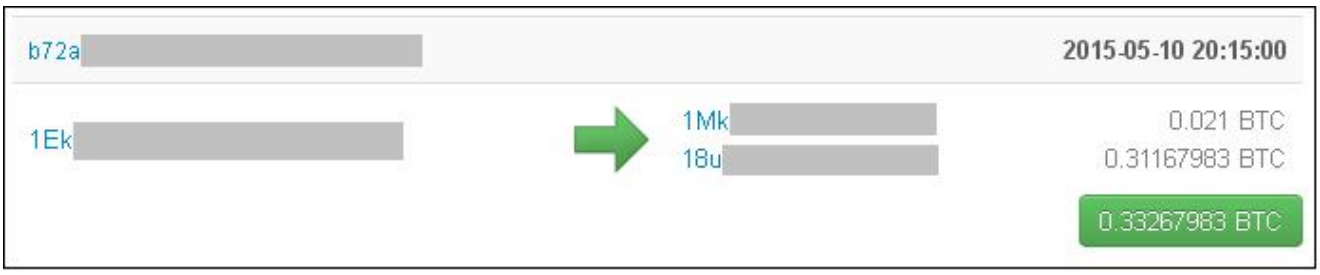


Source: Blockchain.info, 09.05.2015.

To make transactions in the Bitcoin network, the so-called Bitcoin wallet is necessary. To obtain it, the appropriate free software has to be installed. The wallet may be kept on a hard disk or in an internet portal providing such services. Any transaction with the bitcoin cryptocurrency between two users consists in "rewriting the funds from one source address or a larger number of source addresses to the target address or many such addresses[115]". A single transaction may have many inputs and outputs. All transactions in the Bitcoin network are published in open websites, e.g. Blockchain.info. However, in the presented examples some parts of the transacting addresses are masked. Figure 2 presents information on a sample transaction with bitcoins on 10.05.2015. There were sent 0.021 BTC from the address "1Ek…" to the address "1Mk..." and 0.31167983 BTC to the address "18u...". For each transaction there is also provided its identifier. In the presented transaction, the sequence of characters "b72a..." is the identifier.

[115] Quote: [Szymankiewicz M. (2014), op. cit., p. 41].

Fig. 2 A sample transaction in the Bitcoin network



Source: Blockchain.info, 10.05.2015.

Transactions in the Bitcoin network and the Benford's Law

According to the Benford's distribution, probability of occurrence of the digit k ($k = 1, \ldots, 9$) at the first most significant position in the number is expressed with the formula[116]:

$$P(k) = \log_{10}\left(1+\frac{1}{k}\right), \qquad k = 1,\ldots,9$$

However, to make the analysis of compliance of the distribution of the digits with the Benford's distribution reliable, the analysed set of data must meet certain conditions[117]:

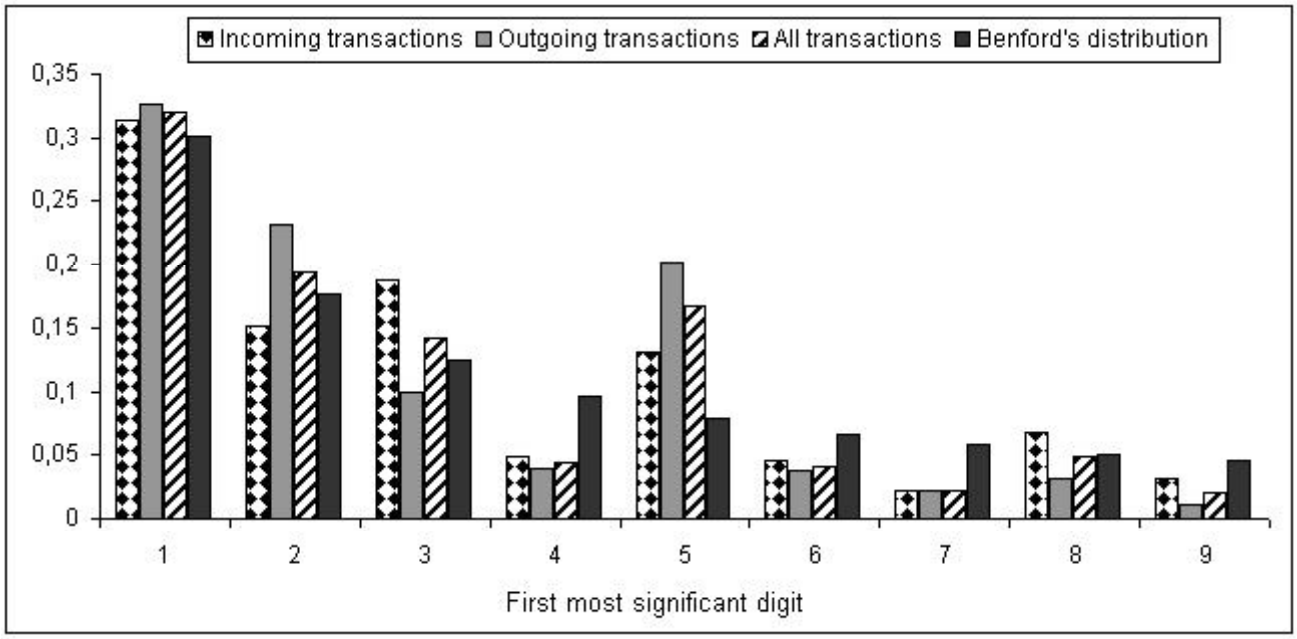
- The data must be presented in the same units,
- The set of data should feature predominance of smaller values,
- The analysed data should not be such data as identifiers, like e.g. phone numbers or bank account numbers,
- Limitations cannot be imposed on the analysed set of data, e.g. with the analysis of transactions which only exceed some amount

For the transactions incoming to and outgoing from a sample address "1GFJ...", correspondence of the distribution of the digits at the first most significant position in the amounts of the transaction was examined against the Benford's distribution. There were 1821 transactions in total, including 881 incoming transactions and 940 outgoing transactions. Neither in case of transactions incoming to this address nor in case of transactions outgoing from this address nor in case of all the transactions (incoming and outgoing), distribution of digits at the first most significant position did not correspond with the Benford's distribution (Figure 3). The transactions related to the "1GFJ..."

[116] Cf. [R. Sasin, *Prawo Benforda - użyteczność oraz zastosowanie*, "Kontrola państwowa", Warsaw, 5/2013, p. 34].
[117] Ibidem, p. 36.

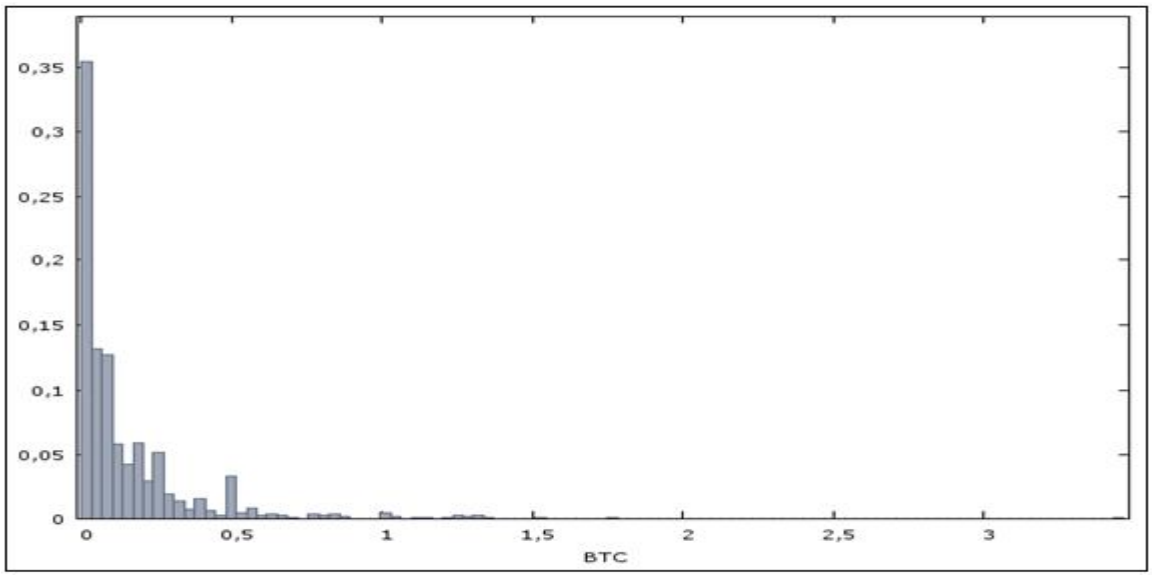
address (both incoming and outgoing) covered small amounts (Figure 4). The average amount in all the transactions was 0.145833 BTC, the median was 0.07787 BTC, and the maximum transfer was 3.42820 BTC.

Fig. 3. Transaction amounts for the "1GFJ..." address and the Benford's distribution



Source: elaboration own.

Fig. 4 Histogram for amounts of all the transactions associated with the "1GFJ..." address

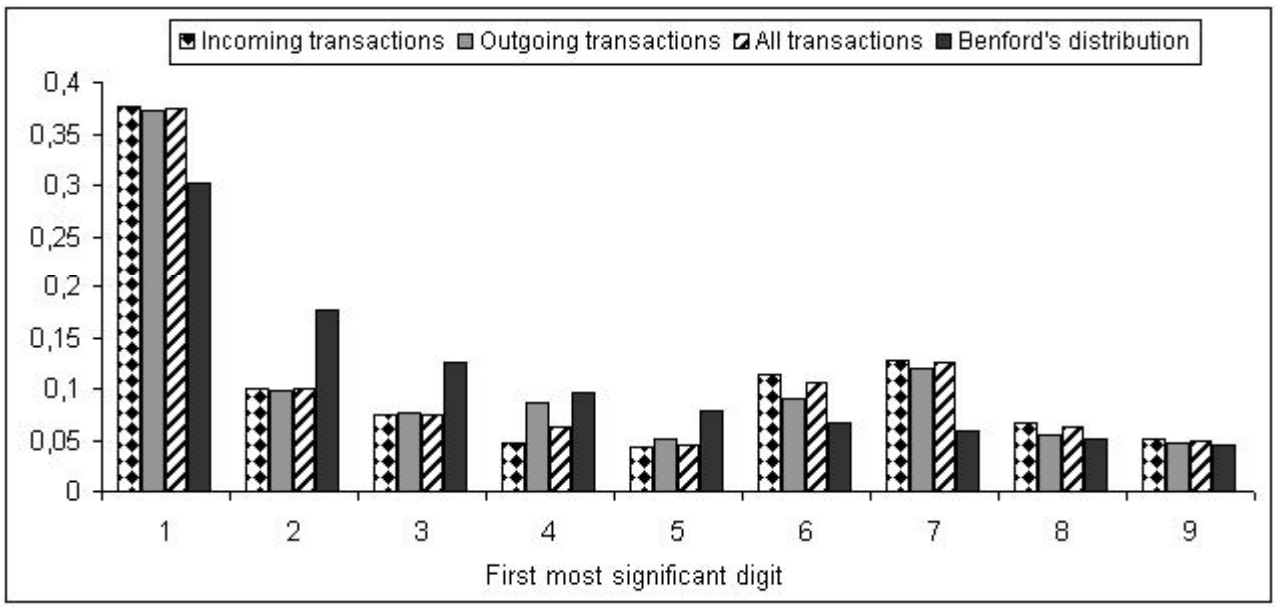


Source: elaboration own.

Correspondence of the Benford's distribution with the distribution of the digits at the first most significant position in case of the transactions related to the "1127..." address was analysed similarly.

There were 610 transactions in total, including 377 incoming transactions and 233 outgoing transactions. Just like in the previous example, for incoming, outgoing and all transactions, this distribution did not correspond with the Benford's distribution (Figure 5).

Fig. 5. Transaction amounts for the "1127" address and the Benford's distribution

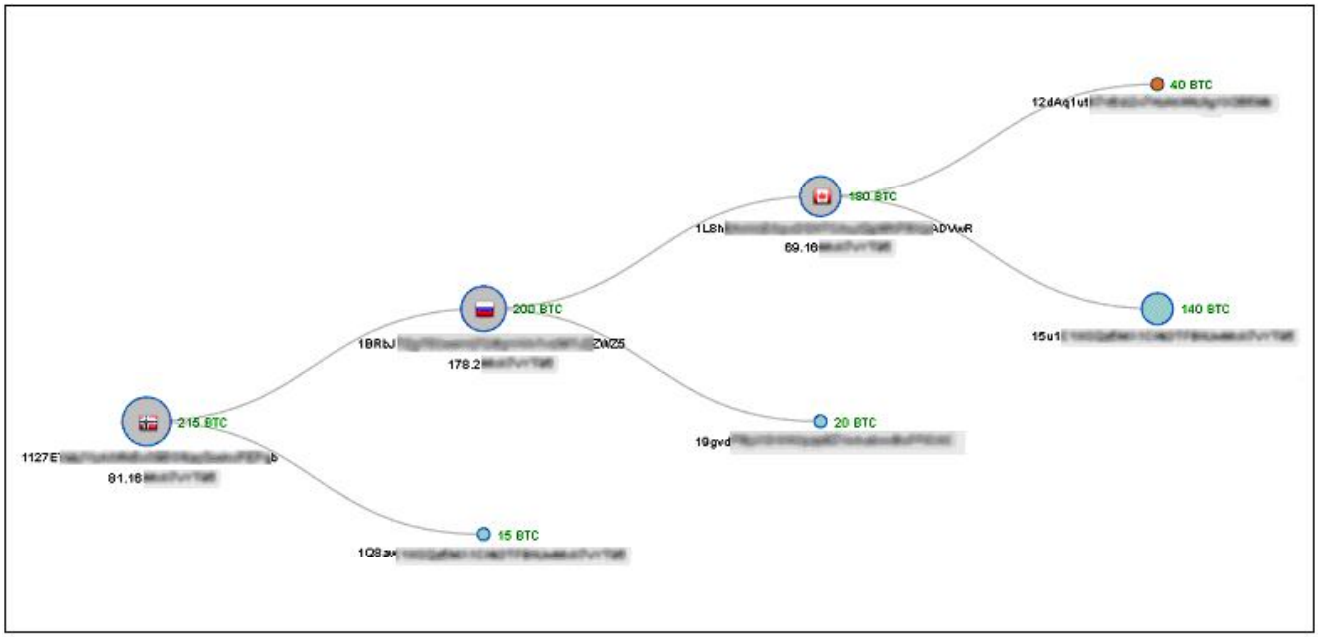


Source: elaboration own.

As with the previous example, the majority of the transactions were about small amounts and the median of the transferred funds for all the transactions was 0.9418 BTC. However, the average amount in all the transactions was 8.20281 BTC, with the maximum transferred amount of 215 BTC. The amount of 215 BTC was transferred on 10.05.2015 from the "34G..." address to the "1127..." address. The history of the transactions published in open websites allows the visualisation of the activities on 10.05.2015 about this amount (Figure 6). The details about the transaction published in the Blockchain.info website include the IP address along with the geolocation determined from it. However, this address does not have to be the address of the computer from which the funds were transferred, as the user may use the TOR network or a proxy server to increase anonymity. For this reason, relating the transaction with the actual location of the person transferring the bitcoins is not clear. In the presented example some parts of the transacting addresses and IP addresses are masked. This information is available in the Blockchain.info website, though. Assuming that in this case the geolocations determined from the IP addresses corresponded with the actual locations of the users of the individual addresses, we can see high speed of transfer of the funds in the Bitcoin network.

Bitcoins were transferred on the same day from Italy[118] to Norway (the geolocation for the "1127…" address), from where the funds were transferred to Russia and then to Canada. With the additional fact that 10.05.2015 was a Sunday, such an international transfer of funds with traditional banking systems would not have been possible.

Fig. 6. Visualisation of the further flow of funds for the selected transaction



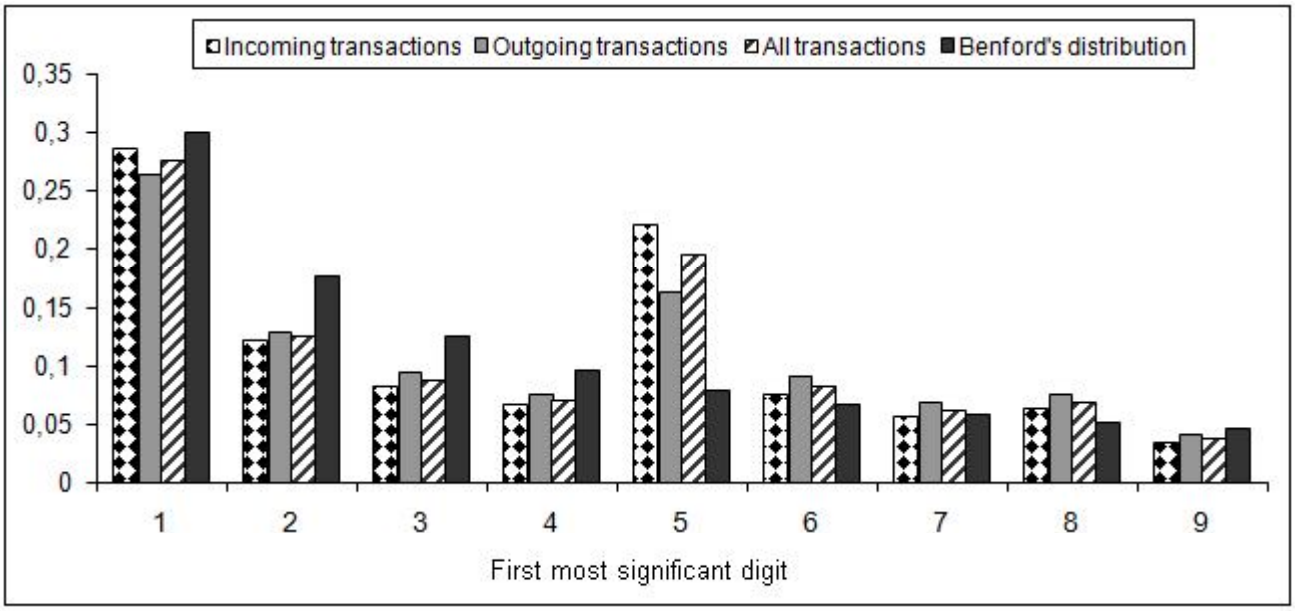
Source: Blockchain.info, 10.05.2015.

Similarly there were analysed distributions of digits at the first most significant position in case of transactions associated with addresses: "31n...", "1N8...", "1Gr...", "19S...". None of them confirmed the Benford's law (Figure 7 – Figure 10). There was used chi-square $\left(\chi^2\right)$ goodness-of-fit test to make this comparisons more precise (Table 2). The test statistic follows a χ^2 distribution with 8 degrees of freedom. Testing on significance level of 5%, value of test statistic should be less than 15,507. If it is greater we should reject assumption that empirical distribution conform to Benford's law[119]. Table 1 presents the details about transactions associated with analyzed addresses. The names of columns of Table 1 represents: "Address" – address of Bitcoin wallet, "All" – total number of transactions associated with the given address, "Incoming" – total number of transactions incoming to the given address, "Outgoing" – total number of transactions outgoing from the given address, "Total Received [BTC]" – total amount received to the given address, "Mean [BTC]" – average amount in all

[118] Italy is the geolocation not included in the figure (Fig. 6) for the "34G…" address, from which 215 BTC were transferred to the "1127..." address.
[119] Cf. [Z. Krakar, M. Zgela, *Application of Benford's Law in Payment Systems Auditing*, Journal of Information and Organizational Sciences, Vol 33, No 1, 2009, p. 42-43].

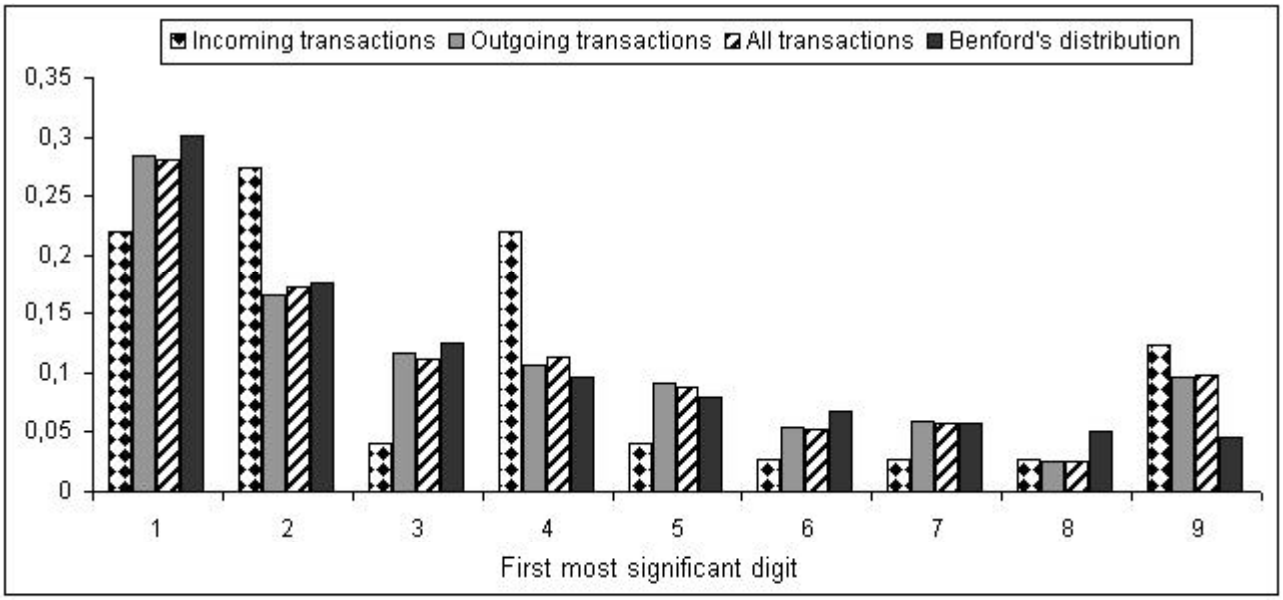
the transaction associated with the given address, "Median [BTC]" – median of amount in all the transaction associated with the given address, "Maximum [BTC]" – maximum amount of transactions associated with the given address.

Fig. 7 Transaction amounts for the "31n..." address and the Benford's distribution



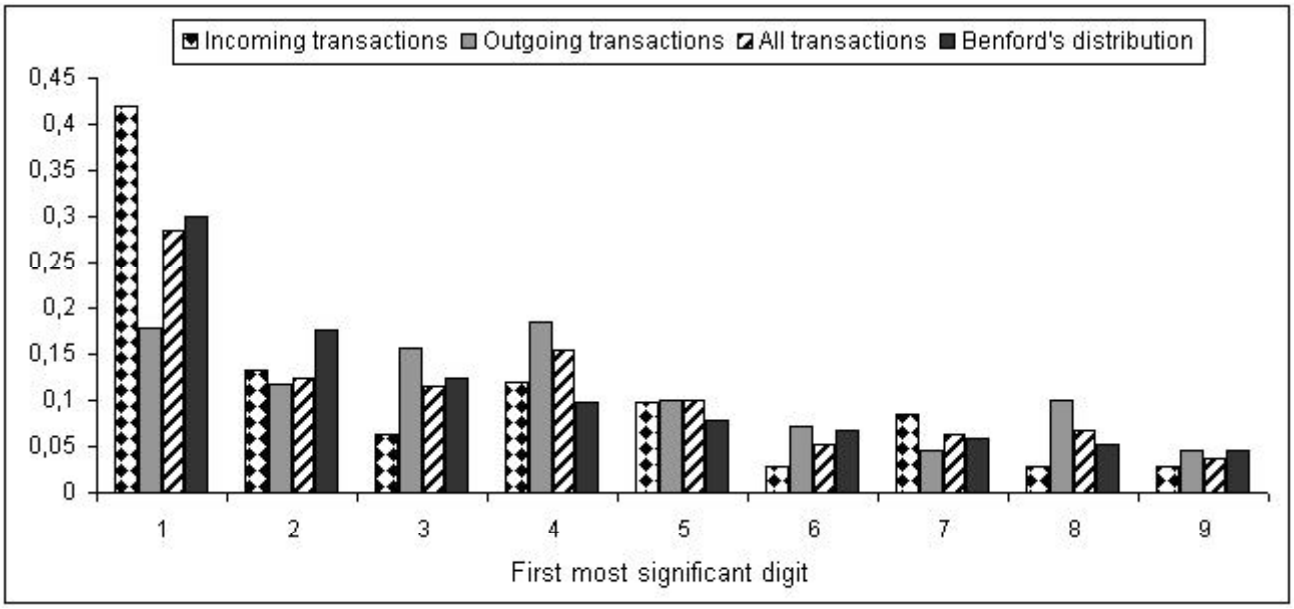
Source: elaboration own.

Fig. 8 Transaction amounts for the "1N8..." address and the Benford's distribution



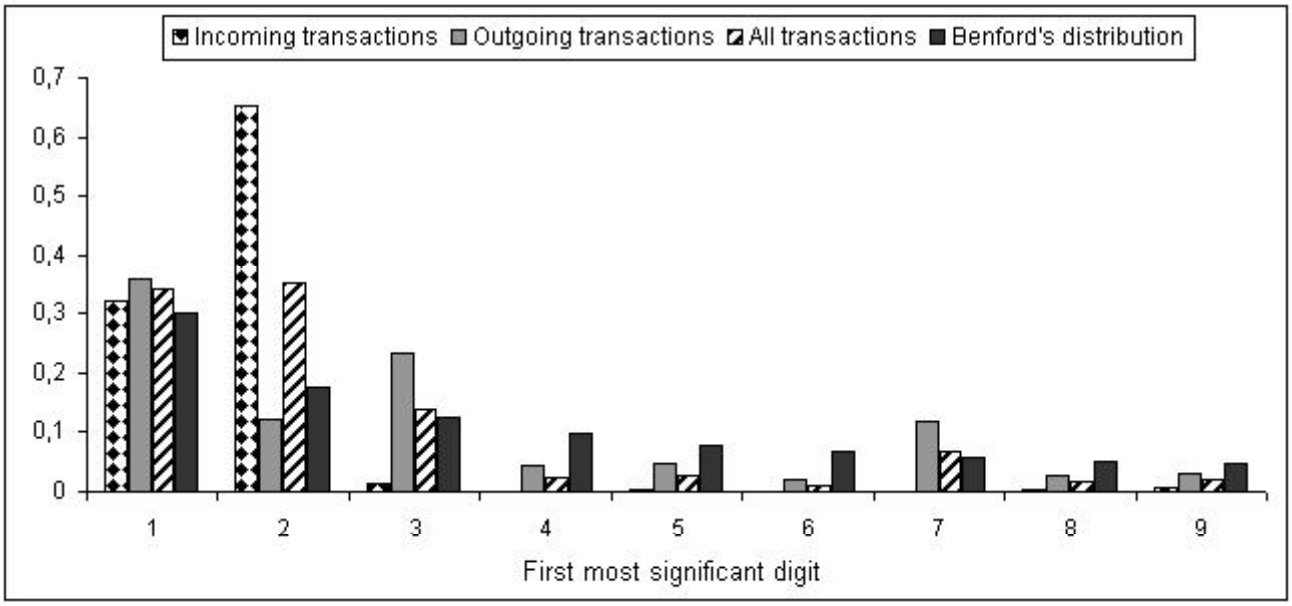
Source: elaboration own.

Fig. 9 Transaction amounts for the "1Gr..." address and the Benford's distribution



Source: elaboration own.

Fig. 10 Transaction amounts for the "19S..." address and the Benford's distribution



Source: elaboration own.

Table 1. The details of the transactions

Address	All	Incoming	Outgoing	Mean [BTC]	Median [BTC]	Maximum [BTC]	Total Received [BTC]
"1GF…"	1821	881	940	0.145833	0.07787	3.42820	132.78068367

Address	All	Incoming	Outgoing	Mean [BTC]	Median [BTC]	Maximum [BTC]	Total Received [BTC]
"112..."	610	377	233	8.20281	0.9418	215	2501.85825198
"31n..."	587	322	265	4.035407	0.120422	999.9937	1206.44134554
"1N8..."	1196	73	1123	3.4646943	0.5139263	115.9415	2079.02185253
"1Gr..."	322	143	179	0.172623	0.075738	1.7534	29.88451061
"19S..."	677	296	381	1.723865	1.2901	50.0001	593.76158347

Source: elaboration own.

Table 2. Chi-square goodness-of-fit test

Address	Incoming transaction	Outgoing transactions	All transactions
"1GF..."	114,84	297,74	321,91
"112..."	88,5	36,28	120,75
"31n..."	97,36	36,81	125,7
"1N8..."	35,58	84,92	97,83
"1Gr..."	21,33	38,16	21,83
"19S..."	520,76	104,32	240,19

Source: elaboration own.

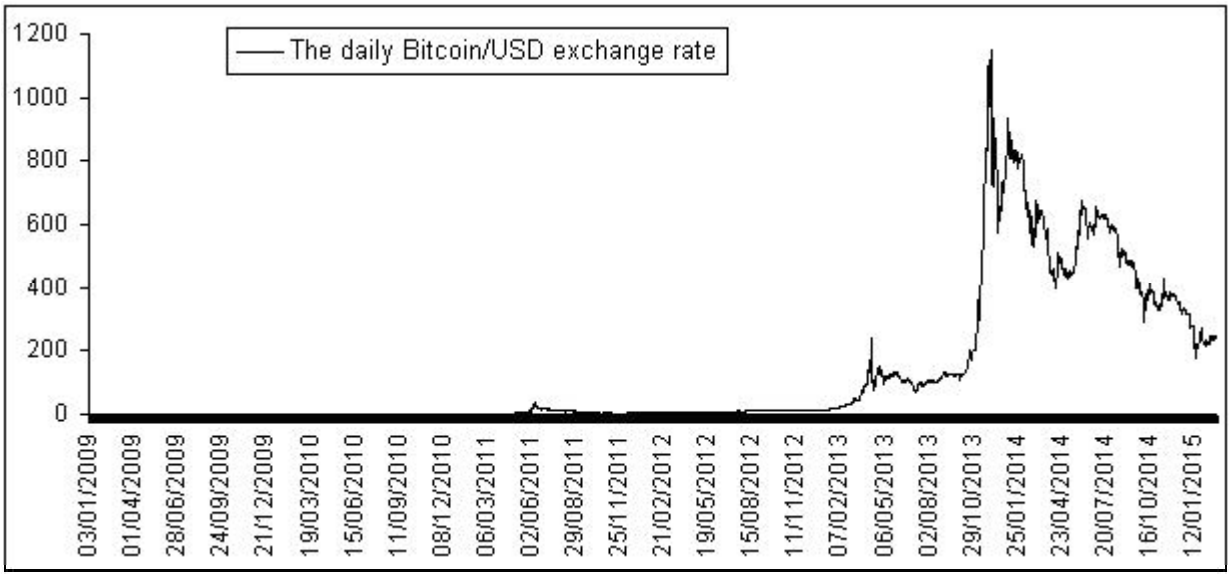
For the sample six selected addresses of Bitcoin wallets, it was found out that the distribution of the digits at the first most significant position in the amounts of transactions related to these addresses did not correspond with the Benford's distribution. One has to remember, however, that lack of correspondence with the Benford's law does not prove any abuse or illegal activity. It is only an indicator of increased risk of irregularities[120]. Similarly, correspondence with the Benford's distribution does not guarantee absence of irregularities. The more so in the case when illegal activities are combined with legal activities and additionally illegal transactions constitute a small part

[120] Cf. [S. Skrzyszowski, *Analiza danych w wykrywaniu wewnątrzkorporacyjnych przestępstw i nadużyć gospodarczych*, [in:] J. Konieczny (ed.) *Analiza informacji w służbach policyjnych i specjalnych*. Warsaw 2012, p. 165].

of the total number of the transactions. For this reason, the results of the analyses conducted with such tools as the Benford's distribution should be viewed cautiously. Lack of correspondence with the Benford's distribution requires further analyses, such as analysis of transactions with the digits at the most significant position for which the highest discrepancies against the Benford's distribution were observed. However, identification of the transaction (e.g. its subject matter) and of the identity of the persons using specific addresses in the Bitcoin network is very difficult due to the anonymity offered by this system of fund transfer. Apart from the test of the digit at the first most significant position, tests of the second digit and testing the digits at the first two most significant positions may also be useful. Sometimes the perpetrator may purposefully use digits like 49, 99 (instead of 50, 100) or round off amounts[121]. Clearly, such amounts may also come from, for example, shop promotions, and not from illegal activities.

The number of people using Bitcoin's wallet and the number of companies that take payment in cryptocurrency are increasing. Bitcoin is accepted as payment for internet services as well as for real goods. It also creates investment opportunities. Figure 11 presents the dynamics of the daily Bitcoin/USD exchange rate. It is important to note that Bitcoin's price is not associated with any "real value" or economic activity, but it is governed only by the forces of demand and supply. This situation may lead to bubbles and crashes on the Bitcoin market.

Fig. 11 The daily BTC/USD exchange rate



Source: Blockchain.info.

[121] Cf. [S. Skrzyszowski, *Analiza danych…*, op. cit., p. 167]. Formulas for the probability of occurrence of particular digits at the second most meaningful position and combinations of digits at the two first most significant positions are presented by, among others, R. Sasin (2013).

O autorze

Kinga Kądziołka ukończyła studia magisterskie na Uniwersytecie Śląskim (matematyka i informatyka). Była słuchaczką Studium Doktoranckiego Uniwersytetu Ekonomicznego w Katowicach. Obecnie pracuje jako analityk kryminalny.

Streszczenie

W artykule poruszono problem identyfikacji (w oparciu o rozkład Benforda) nietypowych transakcji w sieci Bitcoin. Dla przykładowo wybranych adresów portfeli Bitcoin porównano rozkład Benforda z rozkładem częstotliwości występowania poszczególnych cyfr na pierwszej najbardziej znaczącej pozycji w kwotach transakcji związanych z tymi adresami. Rozkłady te nie były zgodne z rozkładem Benforda. Zwrócono uwagę na konieczność zachowania dużej ostrożności przy analizowaniu transakcji za pomocą narzędzi statystycznych takich jak rozkład Benforda. Brak zgodności z rozkładem Benforda w żadnym wypadku nie jest równoznaczny z prowadzeniem działalności niezgodnej z prawem. Z drugiej strony, zgodność z rozkładem Benforda nie stanowi gwarancji tego, że nie występują nieprawidłowości.

Summary

This paper deals with the issue of identification (based on the Benford's distribution) of non-typical transactions in the Bitcoin network. For the sample selected addresses of the Bitcoin wallets, the Benford's distribution was compared with the distribution of frequencies of occurrence of individual digits in the first most significant position in the amounts of transactions associated with these addresses. These distributions did not correspond with the Benford’s distribution. Attention was also paid to the necessity of maintaining high caution in analysing transactions with statistical tools such as the Benford’s distribution. Lack of correspondence with the Benford’s distribution is in no case equivalent with illegal activity. On the other hand, correspondence with the Benford’s distribution does not guarantee absence of irregularities.