



# Analysis of the Bitcoin blockchain: socio-economic factors behind the adoption

Francesco Parino<sup>1</sup>, Mariano G. Beiró<sup>2</sup> and Laetitia Gauvin<sup>1\*</sup>

\*Correspondence:  
laetitia.gauvin@isi.it

<sup>1</sup>Data Science Laboratory ISI Foundation, Torino, Italy  
Full list of author information is available at the end of the article

## Abstract

As the first decentralized digital currency introduced in 2009 together with the blockchain, Bitcoin offers new opportunities both for developed and developing countries. Bitcoin peer-to-peer transactions are independent of the banking system, facilitating foreign exchanges with low transaction fees, such as remittances, and offering a high degree of anonymity. These opportunities together with other key factors led the Bitcoin to become extremely popular and caused its price to skyrocket during 2017 (Henry et al. in J Digit Bank 2(4):311–337, 2018).

However, while the Bitcoin blockchain attracts a lot of attention, it remains difficult to investigate where this attention comes from, due to the pseudo-anonymity of the system, and consequently to appreciate its social impact. Here we make an attempt to characterize the adoption of the Bitcoin blockchain by country. In the first part of the work we show that information about the number of Bitcoin software client downloads, the IP addresses that act as relays for the transactions, and the Internet searches about Bitcoin provide together a coherent picture of the system evolution in different countries. Using these quantities as a proxy for user adoption, we identify several socio-economic indexes such as the GDP per capita, freedom of trade and the Internet penetration as key variables correlated with the degree of user adoption.

In the second part of the work, we build a network of Bitcoin transactions between countries using the IP addresses of nodes relaying transactions and we develop an augmented version of the gravity model of trade in order to identify socio-economic factors linked to the flow of Bitcoin between countries. In a nutshell our study provides a new insight on Bitcoin adoption by country and on the potential socio-economic drivers of the international Bitcoin flow.

**Keywords:** Bitcoin blockchain; Transaction network; Bitcoin adoption

## 1 Introduction

Bitcoin is a digital currency created in 2009 as an alternative to the banking system. Not only does it offer a payment mechanism without any centralized control (i.e. by institutions, governments, or banks), but it has also introduced the revolutionary concept of the blockchain. All the Bitcoin transactions are indeed recorded in the blockchain, a shared public ledger, organized in chronological order within blocks and made secure thanks to strong cryptography. Using end-user softwares called Bitcoin clients, users can easily interact with the blockchain, performing Bitcoin transactions (i.e. transfer of Bitcoin value between identifiers of alphanumeric characters string called *Bitcoin addresses*). Bitcoin users typically use services called *Bitcoin exchangers*, to buy and sell Bitcoin in exchange

for other currencies at a price fluctuating freely. Thanks to a growing number of merchants and services that let people use Bitcoin all over the world, especially during the last years, Bitcoin has become a solid reality and a fascinating object of study. The possible future applications of the blockchain and of cryptocurrencies in general appear as very promising, even if this technology is relatively new and at the first stage of its evolution.

Studying Bitcoin adoption is an important challenge to understand the possible social impact of a decentralized cryptocurrency, based on the blockchain technology. In fact, recent literature abounds with different lines of research linked to the Bitcoin blockchain. A large part of the effort is devoted to the study of the blockchain technology itself, in particular to its development [2–5] and to its application to other domains [6]. Another important line of research concerns the financial and economic aspects, where one of the main questions is related to the evolution of prices [7–10]. Others concern issues with regulatory institutions and policy [2, 11]. From a social point of view, the study of the uptake of the Bitcoin proves to be a challenging task due to the pseudo-anonymity of the system. Digital cryptocurrencies such as Bitcoin can have a significant social impact, as they allow for fast transactions at low costs, offering a solution for tips, donations, and micro-payments without the need of a banking system, paving the way for their wide adoption. However, as users can generate as many Bitcoin addresses as they want, this impact is difficult to quantify. In the direction of investigating the social impact of Bitcoin, previous studies have used either external data such as the number of Bitcoin client software downloads by country, the amount of each fiat currency involved in Bitcoin transactions on exchange [12], and Bitcoin transaction data [13, 14]. To exploit Bitcoin transactions data, a crucial step is the process of deanonymization that consists in grouping the Bitcoin addresses belonging to the same users. This technique was mainly used in literature to characterize the type of usage [14–16]. Such possibility to reveal the Bitcoin addresses belonging to the same user also led to raise questions on the level of privacy offered by the Bitcoin [17].

Here we propose to combine both Bitcoin transactions data and external data sources to quantify the Bitcoin adoption by country; underlining the main factors that might represent a motivation or a deterrent for the Bitcoin adoption, and we explore its evolution in time. Moreover, with the introduction of specific metrics, we build and model an international Bitcoin flow network, and from this model we extract the socio-economic indexes associated to the dynamic of transactions.

We organize the rest of this paper as follows: Sect. 2 provides an overview of the datasets that we used and a description of the pre-processing stage. We analyze three data sources to investigate how relevant they are as proxies to evaluate the Bitcoin adoption. In Sect. 3 we characterize the Bitcoin adoption per country, underlying the relevance of various socio-economic factors and analyzing the adoption trends. In Sect. 4, as a first step we group Bitcoin addresses into users relying on a deanonymization technique. Secondly, using the Internet addresses (IPs) of the nodes that relay transactions, we assign to each user a country. These steps allow us to build the international transaction network, in which we estimate the Bitcoin flows between countries. Finally, we model the international Bitcoin flow network using an augmented version of the gravity model of trade, and we explore the socio-economic variables that are correlated to these flows. Section 5 summarizes and discusses our results.

## 2 Data collection and pre-processing

As our goal is to investigate Bitcoin adoption by country, we gathered three additional sources of information besides the Bitcoin transactional data. First, we obtained the IP address of the first node that has relayed each transaction from *blockchain.info*, a Bitcoin block explorer service. Then, we collected the number of downloads for Bitcoin Core, one of the major Bitcoin clients. Finally, we used information from Google Trends to quantify the collective attention towards Bitcoin. Some details about these datasets are reported in Table 1 and each of them is explained in details in the following sections.

### 2.1 Bitcoin blockchain

The full Bitcoin blockchain database is publicly accessible; we collected the list of transactions using the API from *blockchain.info* [18] over a period from 2009-01-09 to 2016-02-25.<sup>a</sup>

The blockchain database contains all the Bitcoin transactions between Bitcoin addresses. For each transaction we gathered the input and output Bitcoin addresses. Moreover, as the blockchain database grows by appending groups of transactions organized in blocks, we collected for each transaction its position inside the block and the block height (i.e. the number of blocks preceding a particular block). Some general information about the Bitcoin blockchain dataset we collected is reported in Table 2. We have used as timestamp for each transaction the Unix timestamp of the creation of the block in which it is contained. The blockchain does not provide any time information for the transactions, but it contains the timestamp of block creation [20]. Considering that several blocks are created each hour, the block timestamp is a good proxy for our study.

**Table 1** Summary of datasets collected from various sources

Data	Coverage period	Characteristic	Source
Blockchain	2009/01/09 - 2016/02/25	For each transaction: input and output Bitcoin addresses, amounts, fee, transaction timestamp <sup>*</sup>	<i>blockchain.info</i> [18]
IP	2009/01/09 - 2016/02/25 <sup>**</sup>	For each transaction: IP address of the first node that relays the transaction	<i>blockchain.info</i> [18]
Bitcoin client (Bitcoin Core [19])	2009/01/09 - 2015/06/05 <sup>**</sup>	Number of Bitcoin Core downloads per country, daily aggregated	<i>sourceforge.net</i>
Google Trends time series	2009/01/09 - 2017/02/01	Evolution of the number of queries, per country with a week resolution. The data are normalized for each country between 0 and 100.	Google
Google Trends country interest	2009/01/09 - 2017/02/01	Assigns a score to countries based on relative in-country queries. The data are normalized between 0 and 100.	Google

<sup>\*</sup>Timestamp of the block creation.

<sup>\*\*</sup>Effective coverage period shorter (see Bitcoin Client and Internet Protocol addresses subsections).

**Table 2** General statistics about the blockchain dataset collected

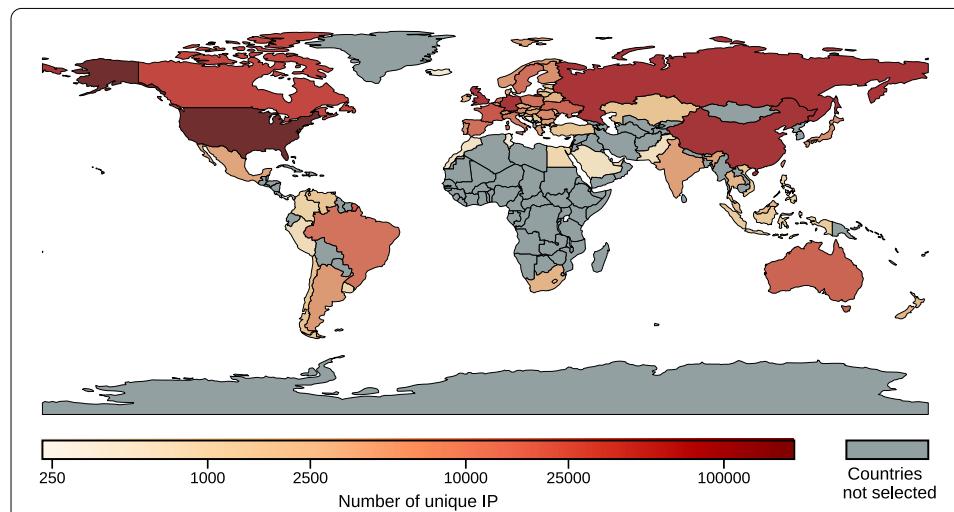
Number of blocks	400,000
Number of transactions	111,793,127
Number of unique Bitcoin addresses	128,894,781

Regarding the transaction amounts, we converted them from BTC (Bitcoin currency) to USD, using a daily exchange rate, as the Bitcoin price has drastically changed over the years (see Appendix A.1).

## 2.2 Internet Protocol addresses

To get an insight about users and their geo-localization we considered the IP of the nodes which relay the transactions in the Bitcoin network. Bitcoin indeed uses a *gossip protocol* in which users communicate their new transactions to all their connected peers across the network. Some studies have shown that the first node/IP which communicates the transaction to a node—such as blockchain.info—which is connected to a large part of the network, is likely to be its creator [21, 22]. We thus downloaded the IP addresses of the first nodes that act as relay in each transaction from blockchain.info. As our goal is to perform a socio-economic analysis at the country level, we mapped the IPs into their corresponding countries. This process is described in A.3. Moreover, we are aware that some users use TOR in order to increase their anonymity in the network. TOR is an Internet protocol which reroutes connections through a virtual circuit so that the IP address is hidden for the rest of the network. During the geo-localization process we thus filtered those transactions relayed by TOR exit nodes (see Appendix A.2), which represent less than 0.001% over the total number of transactions.

One quantity of interest for studying the Bitcoin adoption, is the number of IP addresses that appear for the first time in the Bitcoin network. We call this quantity *unique IP* and, as explained in the following section, this can be used as a proxy to study the adoption in different countries. In Fig. 1 we report the total number of unique IP by countries over our period of study for a selection of countries. We limit the interval of analysis regarding relay node IP to the period beginning from March 2012 to May 2014, because there is some uncertainty on the reliability of the data outside this interval. Indeed, the number of new IP appearing in the system (as IP of relay nodes) shows a sharp drop after May 2014 (Fig. 4), following the collapse of Bitcoin exchanger MtGox. MtGox was the dominant



**Figure 1** Map representing the number of new IP addresses appeared in the Bitcoin system by country in the time interval from 2012-09 to 2014-05, for the countries selected for our study (see the list in Table 9). Countries have been selected based on the activity (IP and clients) as explained in Sect. 3

player then, handling up to 70% of all Bitcoin transactions worldwide [23]. The level of the signal after May 2014 becomes too low and there is little information to extract.

### 2.3 Bitcoin client

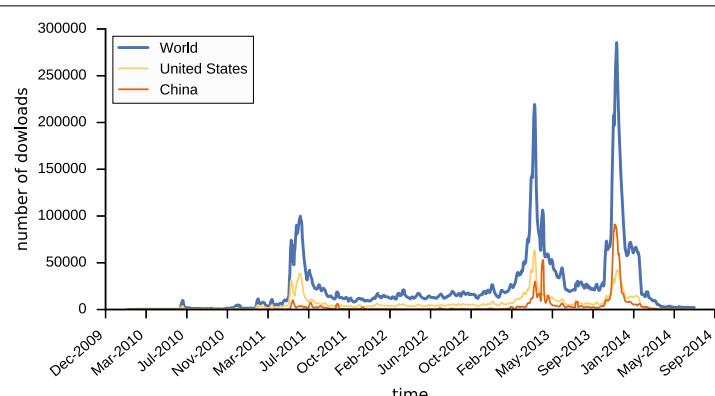
To better assess the Bitcoin uptake we also consider the number of Bitcoin Client downloads. Generally speaking, a Bitcoin client is a software used to manage and store Bitcoin addresses and make transactions on the Bitcoin network. The official Bitcoin client is called *Bitcoin Core*, and it was available from sourceforge.net [24]. SourceForge provides some statistics about the downloads, including the total number of downloads, daily aggregated by country, as shown in Fig. 2. As other clients exist and some users perform transactions through web-based services, the data from Bitcoin Core does not involve all the Bitcoin users. However, as explained in Sect. 3, we assume that it gives a reasonable insight on the general distribution and trends of users. We limit the interval of analysis on the number of client downloads to the period from the beginning of 2011 up to May 2014. Similarly to the IP address data, the Bitcoin client data suffers a sharp drop after May 2014 (Fig. 4) and the data becomes unexploitable.

### 2.4 Google Trends

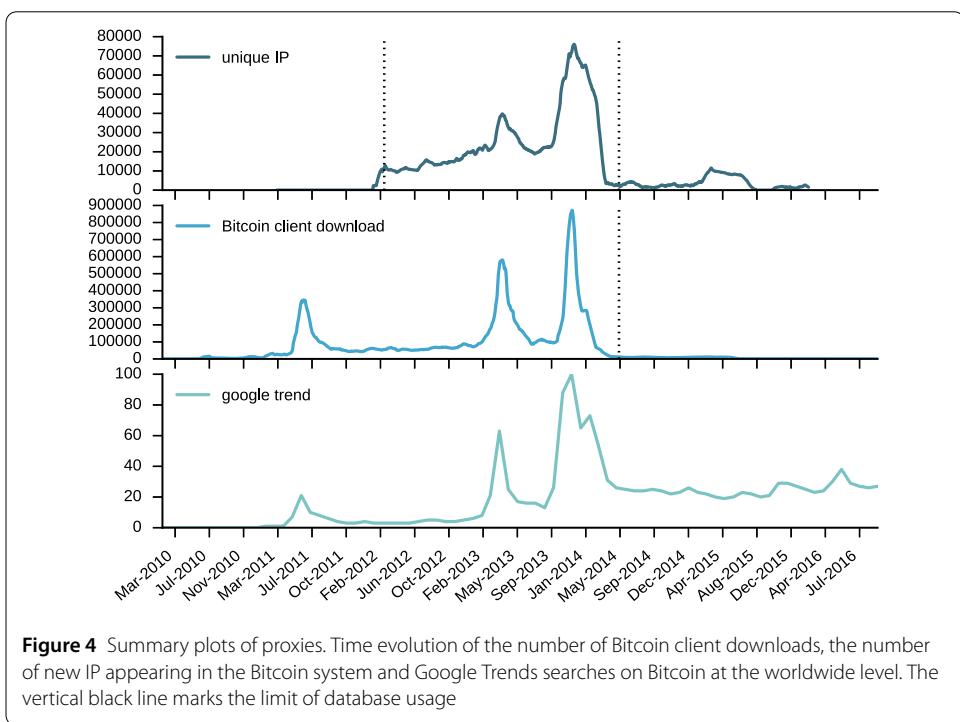
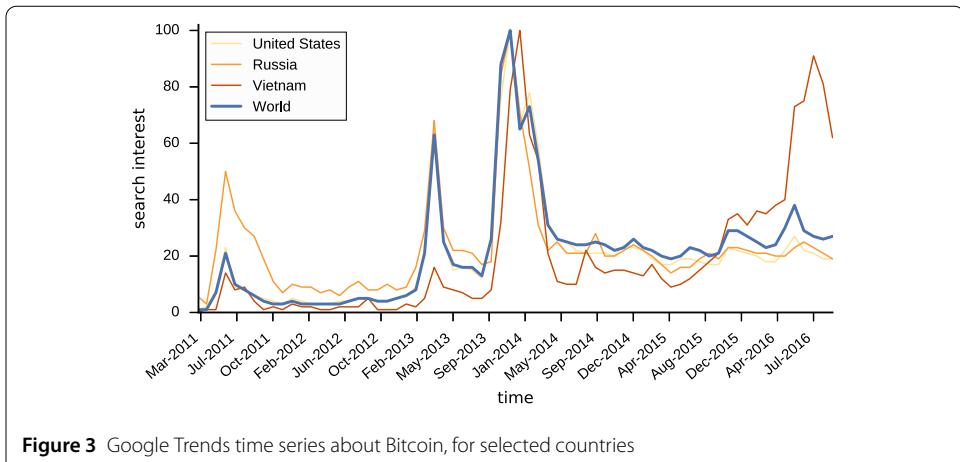
Here we use Google Trends as a proxy for the collective attention on the subject, a method already proposed in [25]. Figure 3 provides, for each country and with a weekly resolution, the evolution of the number of queries for the specific keyword “Bitcoin”, relative to the total number of queries. Besides, we extracted the Google’s interest by region, using the country’s relative number of queries, the scale goes from 0 to 100, 100 being assigned to the country with the highest number of searches on Bitcoin. Although the time series experience a similar drop as the other two (Fig. 4), the level of the signal after May 2014 remains high enough for the data to be exploitable.

### 2.5 Country socio-economic indexes

With the aim of exploring the relationship between some socio-economic indexes and the Bitcoin adoption, we gathered some datasets at country level as summarized in Table 3. We mainly focused on indexes that can distinguish the most developed, richest and wealthiest countries from developing countries. We want to underline that the country development



**Figure 2** Evolution of the number of Bitcoin clients downloaded. World trends, and country-level trends for 2 of the main countries in term of number of downloads



cannot be summarized into a one dimensional economic indicator (indeed there is no criterion that is generally accepted [26]).

### 3 Bitcoin adoption at the country level

With the goal of appreciating Bitcoin adoption at the country level, we have identified Bitcoin client downloads, IP of relay nodes and Google Trends as possible sources of information. Here, we show that these quantities provide a similar and consistent picture of users. Then, we show how countries with different developing indexes have different trends of adoption and lastly, we explore how country socio-economic indexes are linked to Bitcoin adoption.

**Table 3** Summary of socio-economic indexes used

Socio-economic index	Additional information	Source
Internet penetration	Individuals using the Internet (% of population)	World bank <sup>[1]</sup>
Population		World bank
GDP per capita (PPP)	GDP per capita converted to international dollars using purchasing power parity rates	World bank <sup>[1]</sup>
Inflation	GDP deflator (annual %)	World bank <sup>[1]</sup>
HDI	Human develop index	World bank <sup>[1]</sup>
Developing index	Classification: developing/developed	United Nations <sup>[2]</sup>
World Bank countries classification	Countries are divided into four income groupings: low (L), lower-middle (LM), upper-middle (UM) and high (H). Income is measured using gross national income (GNI) per capita	World bank <sup>[1]</sup>
Overall freedom	Freedom from economical, political and social point of view. Values between 0–100	Heritage Foundation <sup>[3]</sup>
Freedom to trade	Freedom to trade goods. Values between 0–100	Heritage Foundation <sup>[3]</sup>

[1] <https://data.worldbank.org>

[2] <http://unctadstat.unctad.org/EN/Index.html>

[3] <https://www.heritage.org/index/explore>

### 3.1 A coherent picture about the users

The numbers of relay node IP and client downloads are directly related to the blockchain, so both of them give a direct information of Bitcoin usage even if none of them can provide a complete picture of the users. In particular, the number of IP addresses does not consider users that do not run a node, and thus do not appear as an IP in the network. On the other side, the number of client downloads provides only information about users using this specific client. Because of these limitations, we cannot identify the exact number of users per country but a trend of evolution. To compare the information given by the numbers of IP addresses and client downloads, we first select countries whose activity level permits the analysis. For each one year moving windows with one-month step from 2012-03-01 to 2014-05-01, we repeatedly filter out countries for which the number of unique IP addresses or client downloads, is lower than the respective medians. At the end of the filtering process, we select a group of 72 countries, listed in Table 9.

A degree of uncertainty exists about the possibility to obtain information about the users from IP addresses and Bitcoin client downloads. Indeed, the first IP address is a noisy identification of the origin of the transaction, while Bitcoin Core is not the only Bitcoin client in use and might give a partial picture of overall Bitcoin adoption. In order to check if they give a consistent picture of Bitcoin adoption, we study the correlation between the two time series and after removing small fluctuations by applying a moving window average (window length: 1 month, offset: 1 day), we indeed measure a high correlation (Table 4). The fact that they correlate positively even though they potentially concern different users encourages the use of these data sources as proxy for the distribution of users among countries. Additionally, we compute the Spearman correlation coefficient between the ranking of countries given by IP addresses and client downloads in three different years, arriving to the same conclusion.

We also confront the Google Trends time series with the numbers of unique IPs and client downloads computing the pairwise Pearson correlations. Given the high correlations as shown in Table 4, we conclude that the Google Trends time series may also be used as an indicator of the country Bitcoin adoption. We suppose that this assumption

**Table 4** Correlation between Google Trends time series, number of unique IPs and Bitcoin client downloads. Here we report correlations between the time-series at world level, and the average correlation at country level, during the period from March 2012 to May 2014. Moreover, selecting a period of one year we compute the Spearman correlation between the countries, ranked using the three proxies

	World time series	Country time series (mean and $\sigma$ )	Ranking of countries		
			2012	2013	2014
IP-Client downloads	0.78*	0.67 (0.13)	0.96	0.96	0.91
IP-Google	0.86	0.76 (0.06)	0.04	0.85	0.79
Client downloads-Google	0.78	0.61 (0.07)	0.08	0.87	0.74

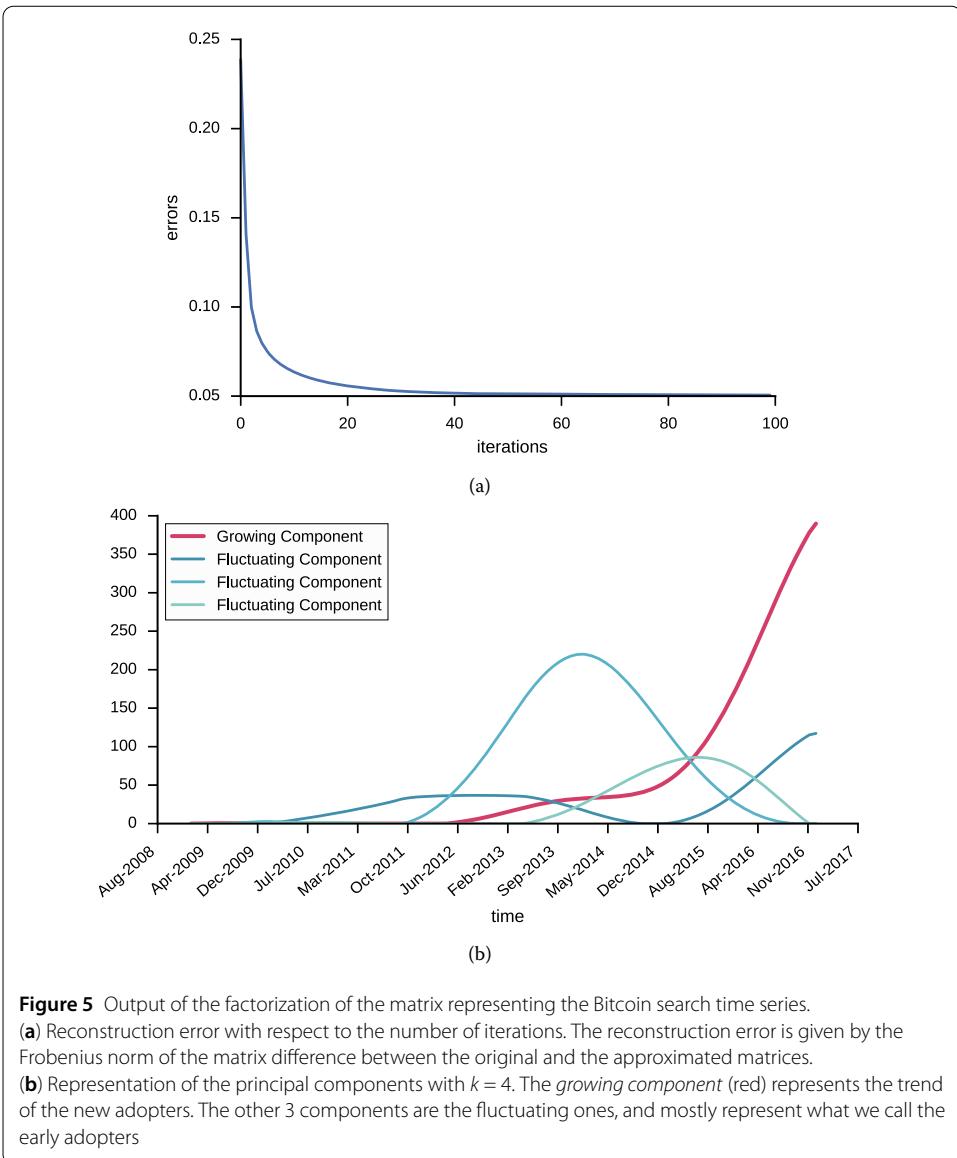
\*World level as sum of activities of countries selected in Sect. 3.

holds for the whole Google Trends data collection period that is longer than for other data sources. This allows us to discuss long term adoption trends of the selected countries.

To assess the relevance of the use of Bitcoin search time series for comparing country adoption, we also measured the Spearman correlation between the pairwise rankings of countries by Bitcoin searches, number of Bitcoin clients downloaded and new IPs appearing. Correlations are also high, apart for the year 2012 where the signal about Bitcoin searches is too low for allowing comparison between countries. Moreover the country ranking based on Google queries heavily depends on Google usage by country, which can be very heterogeneous. As there is no trivial normalization to compensate the heterogeneity of Google usage within countries we will not use the rank provided by sorting Google Trends by countries.

### 3.2 Adoption trends: developing versus developed countries

Using the data from Google Trends we studied the evolution of the collective attention by country from 2009 to early 2017. As we are interested in the long term trends, we smoothed the Bitcoin search time series by country using a low-pass filter to focus on variation on a time scale of 3 years. To study the main trends present in the time series, we built a matrix  $A \in \mathbb{R}^{n \times m}$  (where  $n$  represents the number of countries and  $m$  is the number of points in the time-series), and we approximated it through non-negative matrix factorization (NMF) into a product of matrices  $W \cdot H$  with  $W \in \mathbb{R}^{n \times k}$  and  $H \in \mathbb{R}^{k \times m}$ . Applying such approximation, each country Bitcoin search time series can be represented as a linear combination of  $k$  components, stored as the rows of matrix  $H$ , and with the coefficients stored in  $W$ . The number of components has been chosen to be  $k = 4$  using the bi-cross validation method [27]. The left-hand side of Fig. 5 shows that the Frobenius norm of the matrix difference between the original and the approximated matrices tends towards zero. On Fig. 6 we show the approximated trends for the smoothed time series of 6 countries. The shape of the 4 principal components are shown in Fig. 5. We can see a trend of adoption with a high increase only starting from the middle of 2015. The other three components instead fluctuate over time and represent trends of attention that were already notable in the early years of Bitcoin. Looking at the coefficient matrix,  $W$ , we separated the countries in 2 groups. Those having the increasing component as highest coefficient, that we call *growing* countries, and the others whose main components are the fluctuating components, that we call *fluctuating* countries. As shown in Table 5, grouping countries by development indexes we observed that most of the developed countries are among the fluctuating countries. On the other hand, a large part of the developing countries show

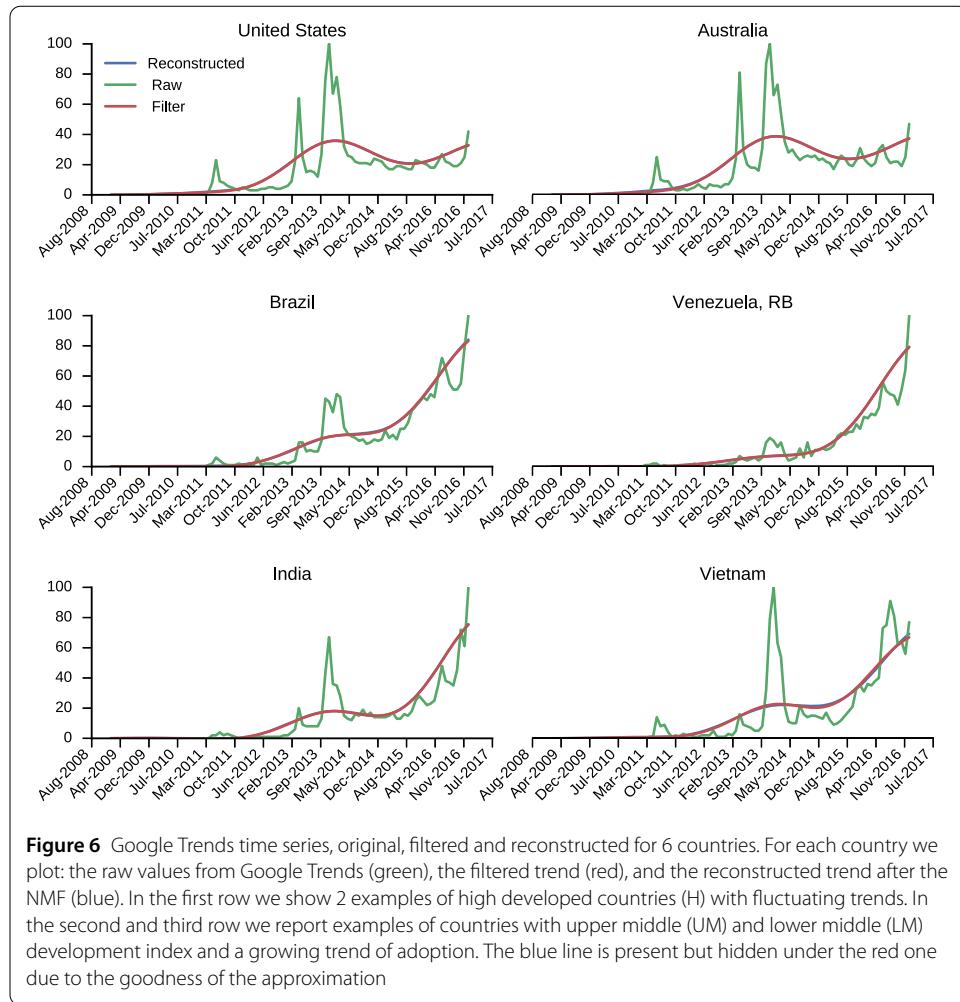


**Figure 5** Output of the factorization of the matrix representing the Bitcoin search time series.  
 (a) Reconstruction error with respect to the number of iterations. The reconstruction error is given by the Frobenius norm of the matrix difference between the original and the approximated matrices.  
 (b) Representation of the principal components with  $k = 4$ . The *growing component* (red) represents the trend of the new adopters. The other 3 components are the fluctuating ones, and mostly represent what we call the early adopters

a recent high interest in Bitcoin. The picture that emerges from this analysis is that at the beginning, attention towards Bitcoin comes only from the developed countries, while starting from 2015 we can see interest picking up in the developing countries.

### 3.3 Socio-economic factors behind the adoption

As measured by the socio-economic indexes, the countries we are analyzing are very heterogeneous. Here we attempt to link the different socio-economic indexes with the different trends of adoption. Focusing on a time interval of one year, we compute the Spearman correlation coefficient between the rank of countries according to the number of client downloads or number of unique IP addresses (normalized by population) and the ranking according to different socio-economic indexes. In the results, reported in Table 6, we observe a high positive correlation both with the Internet penetration, GDP per capita (PPP), HDI, and a small negative correlation with inflation. The general picture that emerges is that socio-economic welfare—as present in most developed countries—appears to have



**Table 5** Repartition of countries between those having mainly an increasing trend, and the others, for different developing index categories (World Bank classification in high (H)/upper middle (UM)/lower middle (LM), and United Nations classification in Developed and Developing)

	Fluctuating	Growing
H	93%	7%
LM	50%	50%
UM	50%	50%
Developed	95%	5%
Developing	46%	54%

stimulated the Bitcoin adoption, at least for the years 2011, 2012, 2013 and 2014 for which we could carry out this analysis.

Beside some expected correlations, like the one regarding the Internet penetration that represents an essential condition to participate in the Bitcoin network, the results obtained for the overall freedom and trade freedom are especially interesting. The two indexes provide a measure of the economic freedom. Trade freedom measures the presence of barriers that affect imports and exports of goods and services, it is measured starting from the average tariff that affect imports and exports of goods and services, and a penalty score that quantifies other type of trade regulation. The overall economic freedom index,

**Table 6** Spearman correlations between the ranks of countries obtained using the number of unique IP addresses and each socio-economic index collected, and the ranks obtained using the number of Bitcoin client downloads and each socio economic index collected

Economic index	Year	Client downloads	p-value	IP	p-value
GDP per capita	2011	0.675	8.25e-11	–	–
	2012	0.638	1.69e-09	0.606	1.70e-08
	2013	0.719	1.18e-12	0.704	5.13e-12
	2014	0.686	2.92e-11	0.696	1.13e-11
HDI	2011	0.806	1.23e-17	–	–
	2012	0.777	1.04e-15	0.733	2.46e-13
	2013	0.791	1.38e-16	0.796	6.77e-17
	2014	0.767	3.76e-15	0.751	3.00e-14
Inflation	2011	-0.409	3.62e-04	–	–
	2012	-0.223	6.00e-02	-0.203	8.70e-02
	2013	-0.317	6.60e-03	-0.277	1.83e-02
	2014	-0.275	1.94e-02	-0.313	7.38e-03
Internet penetration	2011	0.780	6.67e-16	–	–
	2012	0.748	4.51e-14	0.706	4.48e-12
	2013	0.799	3.87e-17	0.794	8.56e-17
	2014	0.780	7.27e-16	0.765	5.37e-15
Overall freedom	2011	0.706	5.26e-13	–	–
	2012	0.678	9.08e-12	0.639	3.19e-10
	2013	0.718	1.43e-13	0.677	9.86e-12
	2014	0.693	2.05e-12	0.659	5.33e-11
Trade freedom	2011	0.814	1.41e-19	–	–
	2012	0.817	7.85e-20	0.789	9.51e-18
	2013	0.864	2.26e-24	0.850	8.10e-23
	2014	0.839	9.26e-22	0.834	2.35e-21

–, IPs data are not available for 2011.

takes a comprehensive view on the country's interactions with the rest of the world and the economic and finance policies within the country. It combines measures for four broad categories: rule of law (property rights, judicial effectiveness, government integrity); Government size (tax burden, government spending, and fiscal health); regulatory efficiency (business freedom, labor freedom, and monetary freedom) and market openness (trade freedom, investment freedom, and financial freedom). The correlations show a positive association between Bitcoin adoption and policies promoting economic freedom, which is somewhat contrary to the common notion that Bitcoin adoption could be driven by overly restrictive legislation.

#### 4 International Bitcoin flow network

In this second section, we attempt to identify the key socio-economic indexes related to the international Bitcoin flow. The process that leads to the estimation of the Bitcoin flow network consists first of all in a clustering of Bitcoin addresses into users, through a deanonymization process, then in a mapping that assigns users to countries.

##### 4.1 Identification of users—clustering of addresses

Bitcoin transactions are made between Bitcoin addresses, which are the result of applying a *hash function* to some input string. Moreover users can create new Bitcoin addresses without limitation in order to hold, receive and send Bitcoin; this is computationally cheap and has no cost for them. This procedure anonymizes the users' activities, as we cannot

know a priori which users are involved in a transaction, nor which set of Bitcoin addresses belongs to the same user.

However, a partial deanonymization method exists and it permits to reveal the group of Bitcoin addresses likely owned by a single user. This additional step is essential for us to make hypothesis on the destination country of transactions, as the IP proxy gives information only on the sender of a transaction. Moreover this process is useful to remove the self change addresses and the related transactions as explained below. This method is based on two heuristics that take inspiration from the underlying functioning of the Bitcoin transaction [21, 28–32]. In particular, we rely on the definitions reported in “Characterizing Payments Among Men with No Names” [29]. The creator of the Bitcoin suggests, in his original paper, the first heuristic that deals with input addresses [33]. Users who hold more than one Bitcoin address can provide a certain number of input addresses in order to reach the desired amount he wants to spend. Due to this functioning, the same user might hold all the input addresses of a transaction. This observation is used to create the first heuristic. Calling  $t$  a transaction and  $\text{input}(t)$  the set of all the input addresses, we summarize the first heuristic as:

**HEURISTIC 1** If two (or more) Bitcoin addresses are inputs to the same transaction, they are controlled by the same user.

- For a transaction  $t$  all  $\text{input}(t)$  are controlled by the same user.

On the other hand, the second heuristic uses the definition of *shadow/change addresses*. The sum of the Bitcoin contained in the input addresses has to be entirely spent. As a consequence, the part of the amount that exceeds the value that the sender wants to spend is usually sent to a new Bitcoin address. The latter is called a shadow (change) address, it is created by the sender with the only purpose to collect back the change. For each transaction, one of the output addresses might be a shadow address.

Calling  $A_i$  a Bitcoin address we focus on the set of output addresses  $\{A_i\}_{i \in \llbracket 1, n \rrbracket}$  of a transaction  $\text{output}(t)$ . We call the number of times the address  $A_i$  is used as output of a transaction as  $n_{A_i}^o$ . We focus on transactions that have at least  $n$  output addresses,  $n \geq 2$  and adopt the following procedure to identify the shadow addresses:

**HEURISTIC 2** The shadow address  $A_i \in \text{outputs}(t)$ , if it exists, is controlled by the same user that controls the  $\text{inputs}(t)$ .  $A_i$  is classified as a shadow address if both the following conditions are satisfied:

- $n_{A_i}^o = 1$  and  $\forall j \in \llbracket 1, n \rrbracket \setminus i \ n_{A_j}^o \neq 1$

The Bitcoin address  $A_i$  appears only once as output of a transaction, and there is no other output addresses  $A_j$  occurring only once.

- $\forall i \in \llbracket 1, n \rrbracket \ A_i \notin \text{input}(t)$

There is not an explicit self shadow addresses, in the sense that there is no Bitcoin address that is present both as an input and output of the same transaction.

After applying the two heuristics, we do not have directly clusters of users, but we only have a partial aggregation at the transaction level. For instance let us assume transactions involving the addresses A, B, C, D, E that result in three groups {A, B, C}, {A, D}, and {D, E} after the deanonymization process. Then {A, B, C, D, E} should be seen as the same user’s Bitcoin addresses. In other words, groups whose intersection is not  $\emptyset$  should be merged.

This process of grouping turns out to be computationally challenging for our large dataset. The heuristics are applied to each transaction and generate a large number of groups of addresses, of which we have to check all intersections to decide whether to group them. However this problem can be mapped onto the problem of finding the *connected components* in a network. We built a network in which Bitcoin addresses represented the nodes and they were linked together if they belonged to the same partial group. We then extracted the connected components of this network. Each connected component represents the complete group of all the user's addresses.

The whole deanonymization process is highly sensitive to any imperfection of the heuristics. The potential effect of a heuristic error is to infer a wrong grouping from some transaction, it could lead to collapse Bitcoin addresses of different users onto a single entity, with the risk of creating users that seem to control a huge number of Bitcoin addresses. Being aware of this problem, we tried to use the safest heuristics possible, even at the expense of discarding some true linking between Bitcoin addresses. As some false linking could anyway occurs, the timespan we use for the deanonymization starts to play a key role; the longer the period of analysis, the bigger the probability that errors can cause the appearance of big clusters of Bitcoin addresses. Reducing the interval of the analysis might lead to the identification of a large number of small groups of addresses, in other terms the same user might still be split in several group of addresses.

The result shown in this section are based on a deanonymization process that takes into account all the transactions which occurred in the year 2013 (i.e the only year for which we have complete IP information). In order to be confident that the results obtained are relatively independent of the timespan considered for the deanonymization, we carried out the whole modeling analysis—described below—applying deanonymization on different time intervals. In particular, we used the period between block 1 and block 400,000 (the last in our database), and the one between block 180,000 and 300,000 (that corresponds to the period for which we have the IP information). In both the cases the results are similar and lead to identification of the same socio-economic factors that can explain the international Bitcoin flow.

Finally after running the deanonymization, we can build the transaction network between users, identifying in the transactions the shadow addresses.

## 4.2 Country association

Thanks to the deanonymization procedure we can identify transactions in which a specific user appears as sender (creator). Assuming that the first node/IP that relays a transaction is its creator we can associate to each user the list of IP used to send Bitcoins. Using the IP geo-localization, we can now associate countries to users. A quick look at the user's IP addresses reveals that we are far from an ideal situation in which every user operates with a single IP (and can be associated to a single country), that in addition this IP is not used by anyone else. Bitcoin services (i.e., the infrastructures that allow users to transact without being a node of the Bitcoin network) partially creates this problem as users are seen as using the IP address that belongs to the service. Moreover, a user who does not use services might also use several IP addresses. To balance the presence of services in the IP addresses usage we build a metric that has the same form as the *term frequency-inverse document frequency* (TFIDF) metric [34] commonly used to reflect the importance of words inside documents. Indeed, we want to account for how often a user uses one specific IP compared

to his overall activity, and how frequently this IP is used among all the users. This metric respects two main principles that we consider crucial for the discrimination:

1. The score rewards the IP usages that are close to the ideal situation, in which an IP address is used by a single user, who uses only that IP address.
2. Being aware that users can use different IP addresses, this metric takes into consideration the ratio between user IP usage and the overall user activity (measured as number of IP addresses).

The formula used to geo-localize the users is reported in Appendix B.1, together with an alternative version, based on similar principles, created to test the robustness of the assignment. As the metric uses the IP information, we carry out this analysis for the restricted timespan from March 2012 to May 2014 (see Sect. 2). The geo-localization procedure allows the identification of destination and origin for 79% of the transactions in 2013.

In order to test the robustness of the assignment of countries, we compare the result of the 2 versions of the metric, finding that 98% of users received the same association. One of the misclassified users is a very active user in 2013, the TFIDF based method classify it as from United States and the other metric as German. This results in differences in the international flow, but as United States and Germany are both developed countries with similar socio-economic indexes, this will not change the interpretation of the results in the modeling part.

#### 4.3 Flow network

After assigning a country to each user, we created the Bitcoin trade network, in which the nodes represent countries and the weighted links represent the amount of Bitcoins exchanged converted in dollars. From now on, we will focus on transactions achieved in 2013 and work with the restricted group of countries analyzed in the first part of the work. In Fig. 7 a visualization of the international Bitcoin flow network is displayed.

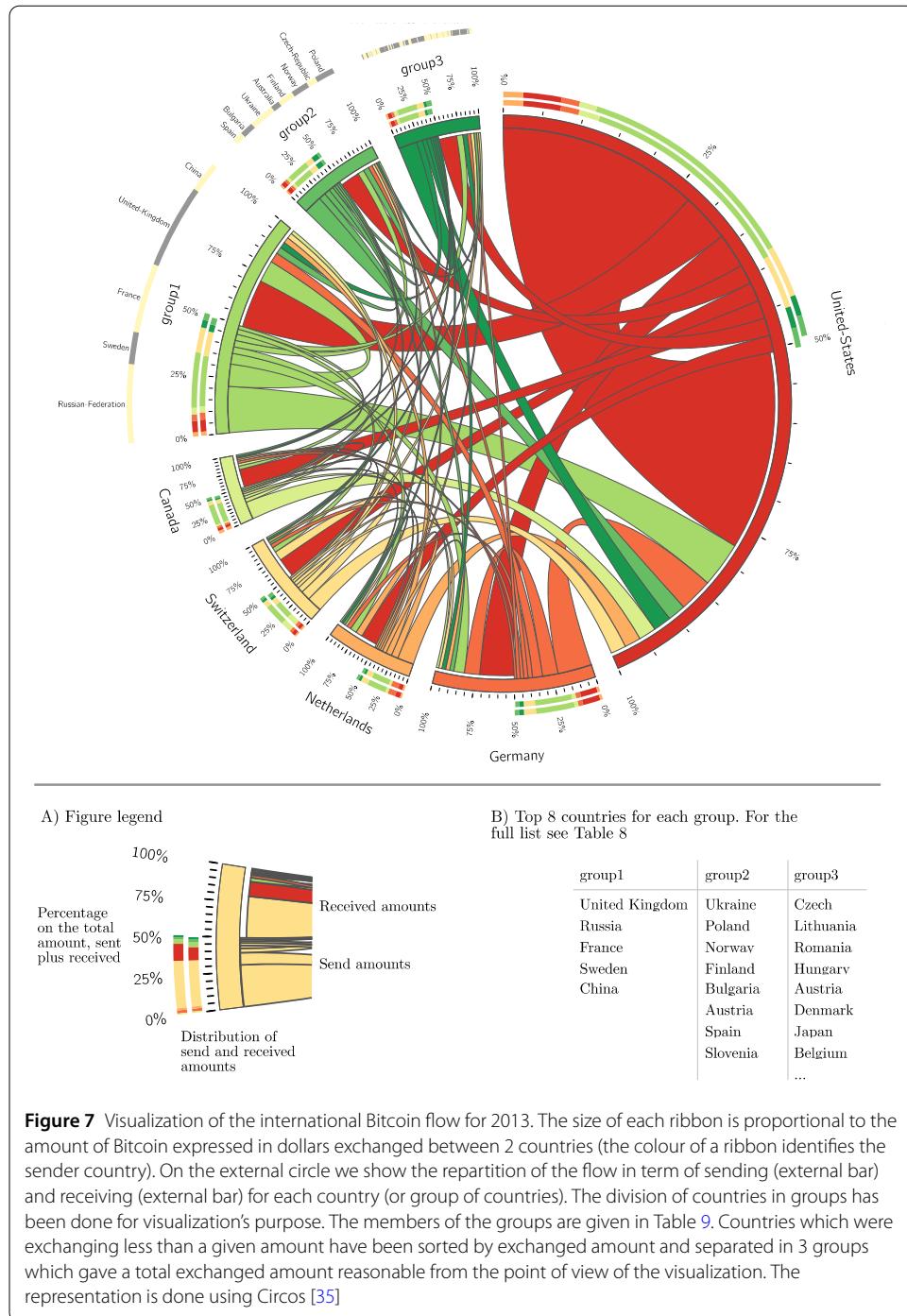
#### 4.4 Flow modeling

To understand which socio-economic indexes are potentially explanatory of the Bitcoin flow, we build a model using as a starting point the *gravity model*, introduced by Jan Tinbergen in 1962 [36] and used to model the bilateral trade flows of different goods and services between countries. The basic form of the model is similar to Newton's law of gravitation: it uses socio-economic indexes that represent the economic mass of the country  $a$ ,  $M_a$ , and which makes the interactions stronger, and a variable representing distance between countries,  $D_{ab}$ , which decreases the strength of the interactions. Adding a constant  $G$ , this model takes the form:

$$F_{ab} = G \frac{M_a^{\beta_1} M_b^{\beta_2}}{D_{ab}^{\beta_3}}, \quad (1)$$

where  $F_{ab}$  represents the flow between countries  $a$  and  $b$  and  $\beta_1, \beta_2, \beta_3$  are coefficients that take real values. The traditional approach for fitting the model consists in taking logarithms of both sides, leading to a log–log model in which it is possible to perform a linear regression [37] ( $\beta_0 = \ln(G)$ ).

$$\ln(F_{ab}) = \beta_1 \ln(M_a) + \beta_2 \ln(M_b) - \beta_3 \ln(D_{ab}) + \beta_0. \quad (2)$$



Here we use an *augmented gravity model* [38–40], which means we are considering additional variables. Calling  $\{X_i^{ab}\}_{i \in \llbracket 1, n \rrbracket}$ , the  $n$  variables that might be either single country quantities (e.g. the masses  $M_a$  and  $M_b$ ) or quantities related to the couple of countries  $(a, b)$  (e.g. the distance  $D_{ab}$ ), the model can now be written as:

$$\ln(F_{ab}) = \sum_{i=1}^n \beta_i \ln X_i^{ab} + \beta_0. \quad (3)$$

Positive  $\beta_i$  are associated to variables  $X_i^{ab}$  that contribute to the mass of countries while negative values instead represent variables that act like distances. However, this approach cannot model the zero observations, and the estimation of the log-linearized equation by least squares (OLS) can lead to significant biases under heteroskedasticity [41]. As an alternative, it is possible to work with its multiplicative form, as shown in Equation 4, replacing the linear regression by a Poisson regression[41].

$$\mathbb{E}(F_{ab}) = \exp \left\{ \sum_{i=1}^n \beta_i X_i^{ab} + \beta_0 \right\}. \quad (4)$$

The vector  $\boldsymbol{\beta} = [\beta_0 \dots \beta_n]$  is estimated maximizing the Poisson log-likelihood:

$$l(\boldsymbol{\beta}|X, F) = \sum_{\forall(a,b)} (F^{ab} \cdot (\boldsymbol{\beta} \cdot \mathbf{x}^{ab}) - e^{\boldsymbol{\beta} \cdot \mathbf{x}^{ab}}), \quad (5)$$

where  $F$  is a vector containing the Bitcoin flow between  $m$  pairs of countries and  $X$  is an  $m \times (n+1)$  matrix, where each column is given by a vector  $\mathbf{x}^{ab}$  whose the values are the variables  $X_i^{ab}, i \in \llbracket 1, n \rrbracket$  concatenated to a 1 that is introduced to take into account the constant term  $\beta_0$ .

Here we use the following group of variables frequently encountered in the literature on trade: population, distance, GDP per capita, and interaction variables that identify countries with a common language or geographic border. Besides, we consider Freedom to Trade, Overall Freedom, Internet Penetration and the World Bank classification of countries by income classes as we observed (see Table 6 and Table 5) that they are linked to Bitcoin adoption. In particular we decide to consider a binary variable that takes the value of 1 for the High-income countries (index H) or 0 otherwise. Additionally to the datasets described before, we downloaded datasets containing information about countries that share a geographic border or the language [42]. Finally, we used a database that reports the distance between each pair of countries, measured using city-level data to account for the geographic distribution of population inside each nation [42].<sup>b</sup>

As a preprocessing step, the variables are standardized, and the Bitcoin flow is estimated in millions of dollars. We then model the flow network maximizing the likelihood introduced below with all the variables mentioned. Despite the heterogeneity of countries in term of trends of adoption, the model achieves a  $R^2$  score of 0.75. This confirms that the socio-economic indexes taken into consideration are good indicators for the international Bitcoin flow.

In order to identify the main drivers of the Bitcoin flow among these socio-economic indexes, we perform a variable selection. To this aim, we introduce  $L_1$  regularization to the model. In practice we estimate the variables which minimize

$$-l(\boldsymbol{\beta}|X, Y) + \lambda \sum_{i=1}^n |\beta_i|, \quad (6)$$

where  $\lambda$  controls the importance of the regularization term. We repeat this process increasing the value of  $\lambda$  from  $10^{-3}$  to  $10^1$ . This leads to the cancellation of the coefficients of the variables that contribute less to the flow. Here we use a 10-folds cross validation in order to set the value of  $\lambda$ , and we use the average *mean squared error* over the different

folds as metric to compare the model's performance. Each of the 10 folds is related to a list of pairs of countries chosen at random. We use as test set the  $k$ th fold that contains  $m_k$  couples of countries  $(ab)_k$ . Calling  $f_\lambda^{-k}$  the model with the regularization term  $\lambda$  trained excluding the  $k$ th fold, we compute the cross-validation error  $CV_k$  as the mean squared error on the test set:

$$CV_k(\lambda) = \frac{1}{m_k} \sum_{\forall(ab)_k} (F^{ab_k} - f_\lambda^{-k}(\mathbf{x}^{ab_k}))^2. \quad (7)$$

Then, we compute the mean of  $CV_k(\lambda)$ , the standard deviation (SD) and the standard error (SE) as:

$$SD(\lambda) = \sqrt{\text{var}(CV_1(\lambda) \dots CV_k(\lambda))}, \quad SE(\lambda) = \frac{SD(\lambda)}{\sqrt{k}}.$$

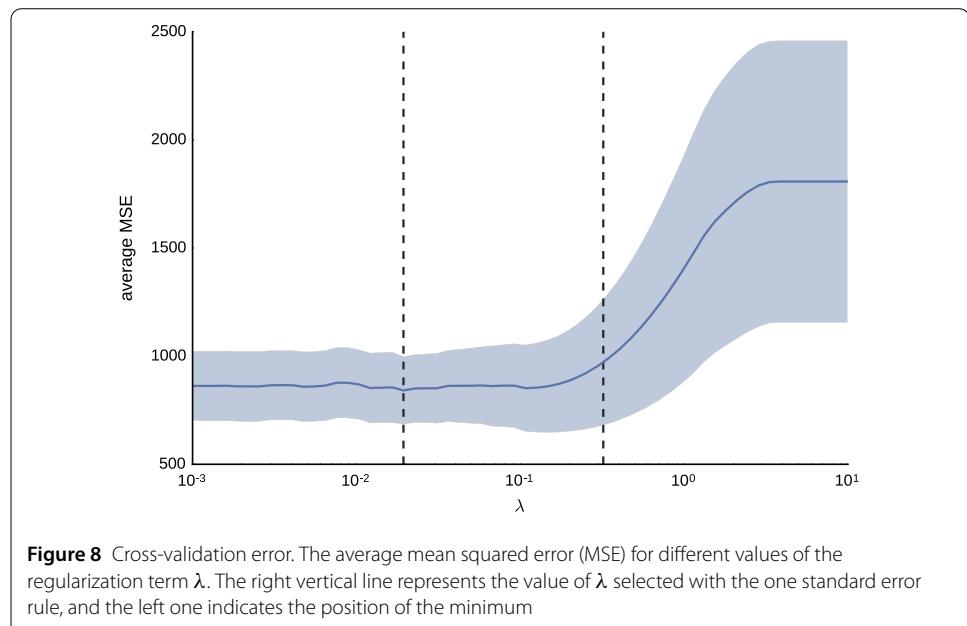
In Fig. 8 for each value of  $\lambda$  tested we show the mean squared error.

As the fluctuations of the cross validation error are small on a large range of  $\lambda$  values, instead of choosing the model with the value  $\lambda_{\min}$  that minimizes the error, we apply the *one standard error rule* [43]. This means that we set  $\lambda = \hat{\lambda}$  where  $\hat{\lambda}$  is such that:

$$CV(\hat{\lambda}) = CV(\lambda_{\min}) + SE(\lambda_{\min}). \quad (8)$$

Fitting the flow with the model described with  $\lambda = \hat{\lambda} = 0.08$ , we identify the main variables (among all those selected for the study) that are explanatory of the flow, the coefficients we found for those are reported in Table 7. In that case the  $R^2$  is equal to 0.68 even though some variables have been dropped.

On one hand, the coefficient of the overall economic freedom index drops to 0 due the variable selection meaning that even though this index takes a comprehensive view of the economic freedom of a country it turns out not to be a key factor to describe the flow.



**Table 7** Coefficients for the variables of the augmented gravity model with regularization, applied to the Bitcoin flow. Overall freedom index, common border and common language are excluded as their respective coefficients drop to 0 during the variable selection made through the regularization

Variables	Coefficient
Population In	1.90
Population Out	1.94
Distance	-0.31
Trade Freedom In	1.45
Trade Freedom Out	1.63
High-income countries In	-0.60
High-income countries Out	-1.53
Internet Penetration In	0.55
Internet Penetration Out	0.84
GDP per capita In	1.10
GDP per capita Out	1.29
Constant	-3.94

**Table 8** Coefficients for the variables of the augmented gravity model with regularization applied to the good trade network. Some coefficients drop to 0 during the variable selection made through the regularization

Variables	Coefficient
Population in	1.30
Population out	1.24
Distance	-0.64
GDP per capita in	0.48
GDP per capita out	0.62
Overall freedom out	0.20
Internet penetration in	0.19
Constant	-1.66

On the other hand, the more specific trade freedom index appears to be, after population, one of the most important variable to describe the flow. The geographic distance appears as an impediment for the flow. The interpretation of the negative coefficients obtained from High-income countries index (H) should be understood in the context of partial multicollinearity of the predictors (GDP, High-income,...). The interplay between this index and the other model's predictors results in a negative coefficient.

To put these results into perspective, we compare the results obtained for Bitcoin trade with those obtained for trade of goods in general. We use the international trade network of goods, as reported by the UN Statistics Division in the Comtrade Database and provided by the Atlas Project.<sup>c</sup> We have access to the value of products exchanged between countries classified by commodity class. Summing the values for all the commodity classes we build the exchange network between each pair of countries.

We fit the gravity model using the same predictors as for the Bitcoin network. In this case too, a large part of the flow can be explained using this augmented gravity model. As for Bitcoin trade, we use lasso regression to perform the variable selection analysis. The results are summarized in Table 8. The goods trade network can be roughly explained using population, GDP per capita, and distance. Those parameters appear with the same sign for the Bitcoin trade network, underlining a certain similarity between the two networks. The additional variables essential to describe Bitcoin flow turn out to be the Internet penetration, the World Bank develop index and the freedom to trade. Apart from the internet penetration whose significance for Bitcoin trade is trivial, we discuss in details the results

obtained for the others two variables: both give an insight on the modalities of economic development of a country, but from different perspectives. As reported from the World Bank, the economic development index is closely correlated with non-monetary measures of the quality of life, such as life expectancy at birth, mortality rates of children, and enrollment rates in school. Freedom of trade describes the development of a country from the country economic interactions with the rest of the world and the economic and finance policies. In a nutshell, we have identified some similarities between the international flow of Bitcoin and that of goods. Yet, apart from the obvious influence of internet penetration, the level of development by country appears to play a greater role in Bitcoin exchanges than for the trade of goods.

## 5 Conclusion

The blockchain infrastructure offered by cryptocurrencies like Bitcoin is attracting interest from a variety of areas such as trade, finances, government and policy. However, it turns out to be a challenging task to quantify this attraction and the adoption by countries.

In this work we aimed at understanding which are the main factors associated to the adoption of Bitcoin as the first blockchain technology in many countries. In order to do this, we applied different techniques for deanonymizing and geolocating the users. Due to the partial anonymity offered by the blockchain, discovering the location of Bitcoin users is a challenging task; we tackled this problem by combining a series of proxies with the transactional data coming from the Bitcoin public ledger. In the first part of the work we showed that the number of IP addresses associated to the relay nodes of the transactions, the number of Bitcoin client downloads, and the interest measured by Google Trends, all give a coherent picture about user adoption by country, even though each of them provides only a partial view of the Bitcoin system. Relying on this result, we analyzed the Bitcoin search time series to explore the evolution of the country attention, and we observed the presence of a net increasing trend of attention from 2015 to 2017, coming mostly from developing countries. Besides, considering the Bitcoin client downloads and IP addresses as proxies for user adoption, we have seen that the adoption is highly correlated with the population, the GDP per capita, the freedom of trade and the Internet penetration for the years 2012, 2013 and 2014. Overall we also confirm that Bitcoin adoption trends have not been homogeneous all around the world: since its introduction, Bitcoin has had a fast growth in many developed countries, while its adoption in developing countries increased very slowly.

In the second part of the work, we focused on Bitcoin flow that is still little explored in the literature, in particular due to the issues related to deanonymization. and we observed that freedom of trade, GDP and population appear as key variables to explain Bitcoin flow.

While this work gives a hint on the socio-economic indexes linked with Bitcoin adoption, it relies on to use of the IP addresses of relay nodes, which are available only for a restricted time period. As future work, the exploration of other data sources beside [blockchain.info](#) could provide IP information for a different period. Another interesting path to overcome this problem would be to model the behavior observed in the transactions with respect to the current distribution of the IP usage accessible, in order to infer the international Bitcoin flows for longer periods of time.

Though we consider here the total flow generated by users and business services (i.e. web-based services like gambling, exchanging, market, mining, clients, etc.), a separate

analysis of these types of flows and activities could also help to understand how the Bitcoin is being currently used.

## Appendix 1: Additional information on the datasets

### A.1 Converting Bitcoin to USD

From 2011 it has been possible to exchange Bitcoin with fiat currencies. The law of supply and demand dictates the price. The value of 1 Bitcoin (BTC) is usually given in U.S dollars (USD) and this value has changed drastically over the last years, from cents to thousands of dollars. Because of this, considering the exchange amounts directly in BTC's might not be representative of their real value, and thus we converted BTC into USD using a daily exchange rate obtained from [blockchain.info](https://blockchain.info).

### A.2 TOR IP addresses

The use of Bitcoin offers a good degree of anonymity through the use of pseudonyms but it does not guarantee a complete privacy. For this reason, some users cover their daily activities using TOR [44]. TOR is an Internet protocol which reroutes its users' connections through a virtual circuit so that the user IP address is hidden for the rest of the network, who sees instead the IP address of the last node used by the TOR protocol, also called *Exit Node* and which belongs to TOR [45]. It is possible to get the full historical list of IPs used as *Exit Nodes* by TOR at <https://collector.torproject.org>, including their corresponding timespan of activity. Comparing this list against our dataset we found around 50,000 TOR transactions that we removed from our study, as by definition we cannot geo-localize them.

### A.3 IP geo-localization

IP addresses can be mapped into countries using several online geolocation tools. These tools rely on different sources of information for building their private databases (probably reversal DNS, pings, and the WHOIS protocol, among others), and their results have been validated at the country level [46]. Particularly, we used the <http://freegeoip.net> API to map every IP address that appeared as a node in the Bitcoin transaction network into a country. As historical records are not available for mapping the node's country at the moment of each transaction, we used the information available at January 2017. We assume that only a negligible fraction of IP addresses have changed location during these years.

## Appendix 2: Methodology

### B.1 Country association additional metric

The metric that uses the TFIDF metric assigns to each user  $u$  a country using the following procedure:

1.  $\text{TFIDF}_{\text{IP},u} = tf_{\text{IP},u} \times \log(\frac{N}{df_{\text{IP}}})$ , where:
  - $tf_{\text{IP},u}$  is the number of occurrences of that IP in user  $u$ .
  - $df_{\text{IP}}$  is the number of users that have used that IP.
  - $N$  is total number of users.
2. We choose as user  $u$ 's country the most present country among the top  $\text{TFIDF}_{\text{IP},u}$  measures for the user. We consider as top those values that cover 50% of the cumulative sum of the TFIDF values for the user.

**Table 9** Selected countries for this work. Moreover, as done for the visualization in Fig. 7 the countries are divided in three groups based on the sum of amount sent and received

Group 1	Group 2	Group 3
United States	France	Poland
Germany	Russian Federation	Australia
Netherlands	United Kingdom	Ukraine
Switzerland	Sweden	Bulgaria
Canada	China	Czech Republic
		Finland
		Spain
		Norway
		Venezuela, RB
		Thailand
		Hong Kong
		Moldova
		Luxembourg
		Singapore
		Mexico
		Denmark
		Latvia
		Slovenia
		Uruguay
		Peru
		Panama
		New Zealand
		Iceland
		Israel
		Turkey
		Ireland
		Arab Emirates
		Italy
		Estonia
		Malta
		Macedonia, FYR
		Slovak
		India
		Pakistan
		Morocco
		Japan
		Indonesia
		Malaysia
		Lithuania
		Hungary
		Belarus
		Bosnia and Herzegovina
		Vietnam
		Portugal
		Taiwan
		Greece
		Belgium
		Romania
		Philippines
		Kazakhstan
		Chile
		Brazil
		Tunisia
		Egypt, Arab Rep.
		Korea, Rep.
		Austria
		Saudi Arabia

$$C_i = \arg \max_{c \in \text{Countries}} \left\{ \text{median}_{\text{IP} \in \text{IP}_{i,c}} (\text{Ratio}(\text{IP})) \cdot \frac{\#\text{IP}_{i,c}}{\#\text{IP}_i} \right\}, \quad (9)$$

$\text{IP}_{i,c}$  = IP addresses that belongs to country  $c$  and it is used by user  $i$ ,

$\#\text{IP}_{i,c}$  = number of time user  $i$  use  $\text{IP}_{i,c}$ ,

$\#\text{IP}_i$  = total number of IP usage from user  $i$ ,

$\text{Ratio}(\text{IP}) = \frac{\#\text{IP}_{i,c}}{\#\text{IP}}$  with  $\#\text{IP}$  as overall IP usage in the Bitcoin network.

## Acknowledgements

The authors thank M. Tizzoni for helpful comments.

## Funding

The authors acknowledge support from the Lagrange Project of the ISI Foundation funded by the CRT Foundation. The funding bodies had no role in study design, data collection and analysis, preparation of the manuscript, or the decision to publish.

## Abbreviations

BTC, stands for Bitcoin; HDI, is Human development index; GDP, is gross domestic product; TFIDF, stands for term frequency-inverse document frequency.

## Availability of data and materials

All data used for this study was publicly available.

## Competing interests

The authors declare that they have no competing interests.

**Authors' contributions**

LG designed the research question, FP retrieved, processed, analyzed data, performed statistical analysis. LG and FP wrote the final manuscript. MB provided guidance over the project and feedback on the manuscript. All authors read and approved the final manuscript.

**Author details**

<sup>1</sup>Data Science Laboratory ISI Foundation, Torino, Italy. <sup>2</sup>Facultad de Ingeniería, Universidad de Buenos Aires, INTECIN (CONICET), Buenos Aires, Argentina.

**Endnotes**

- <sup>a</sup> I.e., up to the block of *height* 400,000.
- <sup>b</sup> In this last dataset we miss information about Serbia, so we did not consider it in the model.
- <sup>c</sup> <https://comtrade.un.org/data/>.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 23 April 2018 Accepted: 7 October 2018 Published online: 16 October 2018

**References**

1. Henry CS, Huynh KP, Nicholls G (2018) Bitcoin awareness and usage in Canada. *J Digit Bank* 2(4):311–337
2. Böhme R, Christin N, Edelman B, Moore T (2015) Bitcoin: economics, technology, and governance. *J Econ Perspect* 29(2):213–238
3. Antonopoulos AM (2014) Mastering Bitcoin: unlocking digital cryptocurrencies. O'Reilly Media, Inc., Newton
4. Kroll JA, Davey IC, Felten EW (2013) The economics of Bitcoin mining, or Bitcoin in the presence of adversaries. In: Proceedings of WEIS, vol 2013, p 11
5. Javarone MA, Wright CS (2018) From Bitcoin to Bitcoin cash: a network analysis. In: ACM proceeding CryBlock'18 1st workshop on cryptocurrencies and blockchains for distributed systems 2018
6. Swan M (2015) Blockchain: blueprint for a new economy. O'Reilly Media, Inc., Newton
7. Kristoufek L (2015) What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PLoS ONE* 10(4):e0123923
8. Cialan P, Rajcaniova M, Kancs D (2016) The economics of BitCoin price formation. *Appl Econ* 48(19):1799–1815
9. ElBahrawy A, Alessandretti L, Kandler A, Pastor-Satorras R, Baronchelli A (2017) Evolutionary dynamics of the cryptocurrency market. *R Soc Open Sci* 4(11):170623
10. Guo T, Antulov-Fantulin N (2018) Predicting short-term Bitcoin price fluctuations from buy and sell orders. Preprint. [arXiv:1802.04065](https://arxiv.org/abs/1802.04065)
11. Evans D (2014) Economic aspects of Bitcoin and other decentralized public-ledger currency platforms. Coase-Sandor Institute for Law & Economics Working Paper No. 685
12. Krause M (2016) Bitcoin: implications for the developing world. CMC Senior Theses, Paper 1261, [http://scholarship.claremont.edu/cmc\\_theses/1261](http://scholarship.claremont.edu/cmc_theses/1261). Accessed 2018-10-10
13. Reid F, Harrigan M (2013) An analysis of anonymity in the Bitcoin system. In: Security and privacy in social networks. Springer, Berlin, pp 197–223
14. Tasca P, Hayes A, Liu S (2018) The evolution of the Bitcoin economy: extracting and analyzing the network of payment relationships. *J Risk Finance* 19(2):94–126
15. Foley S, Karlseñ J, Putniñš TJ (2018) Sex, drugs, and Bitcoin: how much illegal activity is financed through cryptocurrencies? Available at SSRN: <https://ssrn.com/abstract=3102645> or <http://dx.doi.org/10.2139/ssrn.3102645>
16. Lischke M, Fabian B (2016) Analyzing the Bitcoin network: the first four years. *Future Internet* 8(1):7
17. Androulaki E, Karame GO, Roeschlin M, Scherer T, Capkun S (2013) Evaluating user privacy in Bitcoin. In: International conference on financial cryptography and data security. Springer, Berlin, pp 34–51
18. <https://blockchain.info>. Accessed 2016-11-01
19. <https://bitcoin.org/en/bitcoin-core/>. Accessed 2018-01-17
20. <https://en.bitcoin.it/wiki/>. Accessed 2017-05-27
21. Koshy D, Koshy P, McDaniel P (2014) An analysis of anonymity in Bitcoin using p2p network traffic. In: International conference on financial cryptography and data security. Springer Berlin, pp 469–485.
22. Biryukov A, Khovratovich D, Pustogarov I (2014) Deanonymisation of clients in Bitcoin P2P network. In: Proceedings of the 2014 ACM SIGSAC conference on computer and communications security. ACM, New York, pp 15–29.
23. Decker C, Wattenhofer R (2014) Bitcoin transaction malleability and MtGox. In: Kutylowski M, Vaidya J (eds) Computer security—ESORICS 2014: 19th European symposium on research in computer security. Springer, Wroclaw, pp 313–326
24. <https://sourceforge.net/projects/bitcoin/>. Accessed 2017-01-10
25. Puri V (2016) Decrypting Bitcoin prices and adoption rates using Google search. CMC Senior Theses, Paper 1418, [http://scholarship.claremont.edu/cmc\\_theses/1418](http://scholarship.claremont.edu/cmc_theses/1418). Accessed 2018-10-10
26. Vaggi G (2017) The rich and the poor: a note on countries' classification. *PSL Q Rev* 70:279. <https://ssrn.com/abstract=3100742>
27. Owen AB, Perry PO (2009) Bi-cross-validation of the SVD and the nonnegative matrix factorization. *Ann Appl Stat* 3(2):564–594
28. Nick JD (2015) Data-driven de-anonymization in Bitcoin. PhD thesis
29. Meiklejohn S, Pomarole M, Jordan G, Levchenko K, McCoy D, Voelker GM, Savage S (2013) A fistful of Bitcoins: characterizing payments among men with no names. In: Proceedings of the 2013 conference on Internet measurement conference. ACM, New York, pp 127–140

30. Neudecker T, Hartenstein H (2017) Could network information facilitate address clustering in Bitcoin? In: International conference on financial cryptography and data security. Springer, Cham, pp 155–169.
31. Doll A, Chagani S, Kranich M, Murti V (2014) Btctrackr: finding and displaying clusters in Bitcoin. Princeton University, USA
32. Remy C, Rym B, Matthieu L (2017) Tracking Bitcoin users activity using community detection on a network of weak signals. In: International workshop on complex networks and their applications. Springer, Berlin, pp 166–177.
33. Nakamoto S (2008) Bitcoin: a peer-to-peer electronic cash system
34. Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval. Cambridge University Press, New York
35. Krzywinski M, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.* <http://genome.cshlp.org/content/early/2009/06/15/gr.092759.109.abstract>. <https://doi.org/10.1101/gr.092759.109>
36. Tinbergen J (1962) Shaping the world economy; suggestions for an international economic policy. Books (Jan Tinbergen)
37. Goldberger AS (1968) The interpretation and estimation of Cobb-Douglas functions. *Econometrica* 36:464–472
38. Bergstrand JH (1985) The gravity equation in international trade: some microeconomic foundations and empirical evidence. *Rev Econ Stat* 67:474–481
39. Lewer JJ, Van den Berg H (2008) A gravity model of immigration. *Econ Lett* 99(1):164–167
40. Carrere C (2006) Revisiting the effects of regional trade agreements on trade flows with proper specification of the gravity model. *Eur Econ Rev* 50(2):223–247
41. Silva JS, Tenreyro S (2006) The log of gravity. *Rev Econ Stat* 88(4):641–658
42. Mayer TZ, Zignago S (2011) Notes on CEPPI's distances measures: the GeoDist database. CEPPI Working Paper 2011-25
43. Friedman J, Hastie T, Tibshirani R (2001) The elements of statistical learning, vol 1. Springer, Berlin
44. <https://www.torproject.org>. Accessed 2017-07-01
45. Schaap A (2013) Characterization of tor exit-nodes. In: Proc. 18th Twente student conf.
46. Poese I, Uhlig S, Kaafar MA, Donnet B, Gueye B (2011) IP geolocation databases: unreliable? *Comput Commun Rev* 41(2):53–56

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)