Do connections pay off in the bitcoin market?

Kwok Ping Tsang [a], Zichao Yang [b,*]

[a] *Department of Economics, Virginia Tech, Pamplin Hall, Blacksburg, VA 24061, United States of America*
[b] *Wenlan School of Business, Zhongnan University of Economics and Law, Wuhan, China*

ARTICLE INFO

ABSTRACT

This paper identifies the bitcoin investor network and studies the relationship between connections and returns. Using transaction data recorded in the bitcoin blockchain from 2015 to 2020, we reach three conclusions. First, connectedness is not strongly correlated with higher returns in the first four years. However, the correlation becomes strong and significant in 2019 and 2020. Second, returns also differ among those connected addresses. By dividing the connected addresses into ten decile groups based on their centrality, we find that the top 20% most-connected addresses earn higher returns than their peers during most of our sample period. Third, eigenvector centrality is more related to higher returns than degree centrality for the top 20% most-connected addresses, implying that the quality of connections may matter more than quantity among those highly connected addresses.

1. Introduction

Bitcoin was first introduced in 2009 as a decentralized alternative to the fiat money used in our society (Nakamoto, 2008). Unfortunately, due to the lack of application scenarios in the real world, bitcoin, and other alter-coins, till today, are still largely used and treated as a new asset to diversify the investment portfolio. Driven by crypto-enthusiasts and speculators, the market value of bitcoin increased from 3.85 billion US dollars at the beginning of 2015 to 546 billion US dollars at the end of 2020.[1] Bitcoin eventually became a household name in the second half of 2017 when its price increased from less than 3000 US dollars to 19,140 US dollars in less than six months. Since then, bitcoin and other alter-coins have gained substantial attention from regulation institutions, companies, researchers, and investors in the past few years.

In this paper, we do not take a stand on the controversial question of what the fundamentals are behind bitcoin. Instead, we only treat bitcoin as an asset that can potentially be widely adopted for transactions and other purposes. Hence, its price is sensitive to news related to its usage, technological updates, regulations, restrictions, and other factors that influence its popularity. For example, the shutdown of the bitcoin exchange operator BTCC in China on September 14, 2017, caused a 16.2% market plunge (Vigna and Deng, 2017). In an asset market, well-informed investors who obtain news on the asset earlier will earn a higher return, and other less-informed investors will follow as the information is passed to the periphery and earn less (Grossman and Stiglitz, 1980; Kyle, 1985). While it has been shown that such an information network is an accurate description of the stock market (Ozsoylev et al., 2014), our paper attempts to check if the same patterns are present in a new, quickly expanding, and largely unregulated market like that of bitcoin. We will show in the following pages that in this aspect, the bitcoin market is largely like the stock market, with more investors who are more centrally placed on average earning higher returns than others.

This paper asks three questions: (1) Comparing with bitcoin addresses that are unconnected in the bitcoin investor network, do connected addresses earn a higher return? (2) Do more connected addresses earn a higher return than less connected ones? (3) When we use centrality to measure an address' connectedness, which centrality is more related to higher returns?

---

* Corresponding author.
E-mail addresses: byront@vt.edu (K.P. Tsang), yang_zichao@outlook.com (Z. Yang).

[1] Source: https://coinmarketcap.com/currencies/bitcoin/

To answer the first question, we construct a bitcoin investor network using bitcoin blockchain transaction data from 2015 to 2020. Based on the network, all the qualified transaction addresses are clustered into a connected group and an unconnected group. We do not find a consistent pattern showing that the connected group has higher returns than unconnected one during the six years. However, the connected group shows significantly higher returns in 2019 and 2020. To answer the second question, we further divide the connected addresses into ten decile groups based on their degree centrality or eigenvector centrality. We find that the addresses in the two most-connected deciles defined using either centrality measure consistently show higher returns than all the other connected addresses. To answer the third question, we evaluate which centrality measure matters more in the two most-connected deciles, and the results consistently show that high eigenvector centrality is correlated with high returns among those most-connected addresses. This result may imply that the quality of connections is more important than the quantity among those highly connected addresses. We also conduct a series of robustness checks. We choose three different thresholds to identify the network connection between addresses, and weekly returns are also used to replace the monthly returns used in the main results. All the results in these robustness checks are consistent with our main results.

A number of studies focus on the characteristics or behaviors of bitcoin investors. Xi et al. (2020) conduct a web-based revealed preference survey to profile Initial Coin Offering (ICO) investors' characteristics in China and Australia, and they find the investing motivation factors are different in these two countries. After analyzing the bitcoin trading data from Mt. Gox, Glaser et al. (2014) conclude that new bitcoin users from 2011 to 2013 tend to treat bitcoin as an alternative investment option rather than a new currency. This result aligns with the survey result conducted by Mahomed et al. (2018) in South Africa. Hung et al. (2021) show that different kinds of investors may exert opposite effects on price discovery in the bitcoin future market. Using the Mt. Gox exchange trading data, Gandal et al. (2018) and Chen et al. (2019) demonstrate that there are market manipulations inside the Mt. Gox exchange which cause unprecedented spikes in bitcoin price. To study the behaviors of bitcoin investors, we think exchange data provides a more accurate and comprehensive picture. However, exchange data are not publicly available. The leaked Mt. Gox exchange trading data has been widely used in several studies, but the drawback is that it only covers the period from April 2011 to November 2013. In this paper, we are using the data obtained directly from the bitcoin blockchain. Bitcoin blockchain raw data is publicly accessible, and it allows us to look at the more recent periods where the market is more established and mature. However, bitcoin blockchain raw data also comes with its own problem: the transaction data are connected to bitcoin addresses, not individual investors. In this paper, we propose a simple parsing procedure to mitigate this drawback.

Most studies on investor networks focus on stock markets. By exploiting a dataset that includes all account-level trading information on the Istanbul Stock Exchange in 2005, Ozsoylev et al. (2014) find that investors' investment returns are positively correlated with their centrality in the network. Rossi et al. (2018) show that investment managers who are better connected tend to have a better portfolio performance. Using data from the Chinese stock market, Li et al. (2017) show that super investors who trade the most are earning higher profits than the other investors. Ahern (2017) analyzes an illegal insider trading network and finds that people in the network earn a 35% return over 21 days, and more central traders in the network earn even higher returns. Walden (2019) introduces a dynamic noisy rational expectations model and finds out that an agent's profitability is determined by Katz centrality. However, stock markets are generally more regulated and mature. In this paper, we look at the bitcoin market, which is relatively new and less regulated. Liu and Tsyvinski (2020) find that coin returns are positively related to cryptocurrency network growth rates, and current cryptocurrency prices include information about the expected network growth. As a complement to their study, we focus on how returns differ among investors inside the bitcoin network using address-level transaction data.

The remainder of the paper is structured as follows: Section 2 introduces the basic technical background of bitcoin. Section 3 explains data sources and the blockchain data parsing procedure. Section 4 constructs the bitcoin investor network. Section 5 shows the empirical results about the effect of connectedness and centrality on bitcoin investment returns. Section 6 concludes the paper.

2. Background

The original idea of Nakamoto (2008) is to build bitcoin as a decentralized payment system that does not need a centralized clearinghouse yet still keeps transaction records secure and immutable. In the following sections, we will explain some technical details about bitcoin that are necessary to understand the rest of this paper.

2.1. Structure of a bitcoin block

Blocks are storage units that contain confirmed transaction data. Each block in the blockchain, which is a sequence of blocks, contains a piece of information about the location of its previous block. The size of each block is set to be less than 1 MB to ensures that each block can only contain a certain number of transaction records. Meanwhile, every block includes a block header that includes six fields: version, previous block hash, merkle root, timestamp, target, and nonce. The complete structure of a bitcoin block is summarized by Antonopoulos (2017) and reproduced in Table 1. In this paper, the relevant data comes from the transactions section in every bitcoin block.

2.2. Bitcoin transactions

Every transaction in the bitcoin system consists of two parts: transaction input and output. Except for coinbase transactions, a transaction input usually should come from an unspent transaction output (UTXO) generated from a previous transaction.[2] In Table 2, we list the major components in a bitcoin transaction.

---

[2] The coinbase transaction in a block sends the mining reward to the miner who successfully mined this block. The transaction input in a coinbase transaction is not linked to a UTXO. We remove all coinbase transactions from our dataset.

Table 1
The structure of a bitcoin block.

Field	Description
Block Size	The size of the block
Block Header: Version	A version number to track software/protocol upgrades
Block Header: Previous Block Hash	A reference to the hash of the previous block in the chain
Block Header: Merkle Root	A hash of the merkle tree root of this block's transactions
Block Header: Timestamp	The approximate creation time of this block
Block Header: Target	The Proof-of-Work algorithm target for this block
Block Header: Nonce	A counter used for the Proof-of-Work algorithm
Transaction Counter	The number of transactions included in this block
Transactions	The transactions data

This table is adopted from Table 9-1 and Table 9-2 in Antonopoulos (2017). Hash function can transfer data of arbitrary size into data of a fixed size. And the output can be called hash. Hash function is preimage resistance, which means as long as the input is the same, we can always get the same output, but it is much more difficult to deduce the input from the output. Merkle root is the root of merkle trees, which are data structures that helps nodes quickly verify transactions and reduce data transaction.
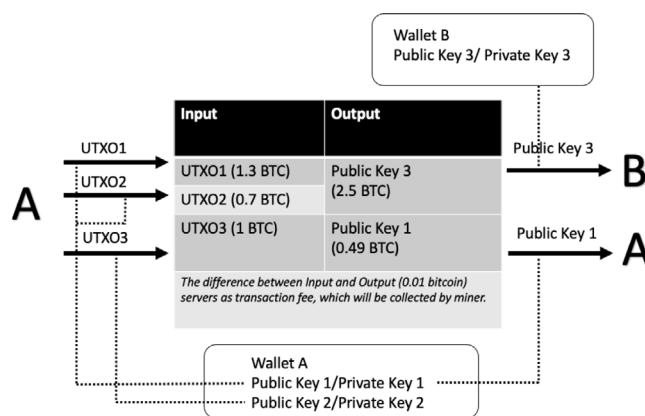

Fig. 1. A simplified transaction framework I. This figure shows a simple transaction. Private key 1 in wallet A controls UTXO1 and UTXO2, private key 2 controls UTXO3. User A takes out 3 bitcoins from her bitcoin wallet, and sends 2.5 bitcoins to user B, and sends back 0.49 bitcoins to herself. The 0.01 bitcoin difference serves as the transaction fee collected by the miner who successfully confirmed this transaction.

When a user receives bitcoins from other users, what it means is that this user's bitcoin wallet has detected a new UTXO that can be spent by one of the private keys controlled by this wallet. One feature of UTXO is that it is an indivisible chunk of bitcoin, just like a dollar bill. Users cannot spend a part of a UTXO in one transaction in the same sense that they cannot use a part of a dollar bill in one transaction. However, users only need to tell their bitcoin wallet the desired transaction amount and transaction fee, and then their bitcoin wallet will automatically select from all the UTXOs controlled by the wallet to compose an amount equal or greater than the desired transaction amount.

To verify the ownership of a UTXO, bitcoin relies on the public-key cryptography. A digital wallet can generate almost unlimited pairs of public key and private key, and each pair includes one public key and one private key. A public key is included in every UTXO, and it tells the whole bitcoin network the receiver of that UTXO. On the other hand, a private key is a certification telling miners who is the real owner of that UTXOs as one public key can only be deciphered by its corresponding private key. To simplify the analysis, here we can think of each UTXO as a dollar bill with a face value equal to the amount of bitcoin embedded in this UTXO. Fig. 1 illustrates a bitcoin transaction.[3] In user A's digital wallet, there are 2 pairs of public and private keys (public key 1/private key 1, public key 2/private key 2). Public key 1 is linked to UTXO1 (1.3 bitcoins) and UTXO2 (0.7 bitcoins) respectively, and public key 2 is linked to UTXO3 (1 bitcoin). Now user A wants to send 2.5 bitcoins to user B (public key 3). User A cannot send 1.5 bitcoins from public key 1 and 1 bitcoin from public key 2 to public key 3. What user A can do is send out all 3 bitcoins: 2.5 bitcoins to public key 3 and 0.49 bitcoin to her own public key. The 0.01 bitcoin difference serves as the transaction fee collected by the miner who successfully confirmed this transaction.

In the above transaction, user A sends back the change of 0.49 bitcoins to her own address (public key 1), and such transactions are usually referred as change transactions. Change transactions are easy to be detected if users keep sending back the changes to the addresses that showed up on the input side. However, directly sending back changes to one of the already used addresses may allow other people to trace back all the transactions related to user A. To protect her privacy, user A can use the bitcoin wallet to generate a new pair of public key 4 and private key 4 and send the rest 0.49 bitcoin to public key 4 (see Fig. 2). By doing so, other

---

[3] Here we use public keys instead of bitcoin addresses (public key hashes) to illustrate this simple transaction. In practice, most transactions actually use bitcoin addresses instead of public keys on the output side.
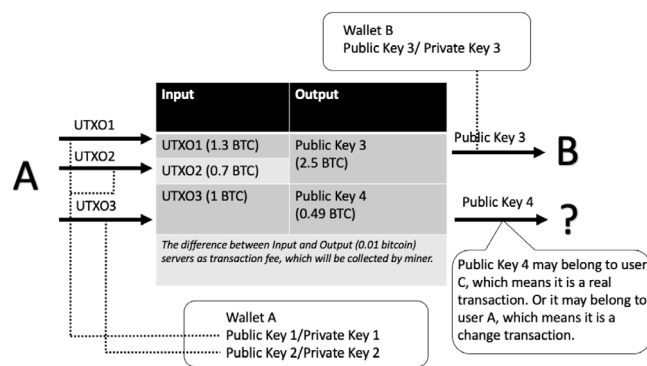
Fig. 2. A simplified transaction framework II. This figure shows a simple transaction. Private key 1 in wallet A controls UTXO1 and UTXO2, private key 2 controls UTXO3. User A takes out 3 bitcoins from her bitcoin wallet, and sends 2.5 bitcoins to user B. In this transaction, We do not know the identity of the person who receive the 0.49 bitcoins. It could be a new address created by user A to receive the change, or it could be owned by a third person. The 0.01 bitcoin difference serves as the transaction fee collected by the miner who successfully confirmed this transaction.

Table 2
The structure of a bitcoin transaction.

Field	Description
Input: Transaction Hash	Pointer to the previous transaction that contains the UTXO
Input: Output Index	The location of the UTXO in the previous transaction
Input: Unlocking Script	The script to fulfill the condition set by the UTXO locking script
Output: Amount	The amount of bitcoin that should be sent to the output address
Output: Locking Script	The script that set the condition to spend the bitcoin

This table is adopted from Table 6-1 and Table 6-2 in Antonopoulos (2017).

people can never tell if the 0.49 bitcoin transaction is a change transaction or not. This practice produces two challenges: (1) It is difficult to cluster the transaction data into individual-investor level because investors can generate new addresses for every new transaction to cover their traces. (2) It is difficult to filter out all the change transactions from the transaction dataset as investors can always generate new addresses to receive changes. Some solutions are proposed in the following section to address these issues.

3. Price and transaction data

3.1. Data source

In the bitcoin transaction raw data, the input side does not include the real transaction amount. Instead, the input side includes the UTXOs that are supposed to be used to fulfill the transaction. It is the miners' responsibility to verify if these UTXOs contain enough bitcoins to fulfill this transaction. The raw transaction data stored in the blockchain is rather hard to interpret. Luckily, Google has pre-parsed the raw data and made it more user-friendly, so we use Google's bitcoin transaction dataset as our main data source.

We also use the minute-level bitcoin price data from Bitstamp and Coinbase to construct a block-level bitcoin price dataset. If the price data is available on both platforms, we use the average value; if the price data is only available on either one platform, we use the available data to represent the block-level bitcoin price data. If the price data is not available from both sources, the price will be imputed based on the average value of neighboring available prices.

Subject to the availability of the minute-level price data, our sample begins from 2015-01-01, and ends on 2020-12-31. Table 3 shows how many unique transaction addresses are recorded into the bitcoin blockchain. Since 2017, there are more than 100 million unique transaction addresses are recorded into the bitcoin blockchain every year. Also, more than 90% of addresses showed up less than three times. In this table, we can see that, even though the absolute number of unique addresses under different repeat time criteria varies, the percentage numbers are relatively stable across different years.

3.2. Data parsing

As mentioned in Section 2, users can generate a new pair of public and private keys for each transaction to hide their identities. One challenge brought up by this privacy feature is that we can never be certain if two addresses are owned by the same person or owned by two different people. Meiklejohn et al. (2013) propose two heuristic rules to parse bitcoin addresses data: (1) the public keys used as inputs in the same transaction should be considered as being owned by the same person; (2) the change address on the output side in a transaction should be recognized as being controlled by the person who initiates the transaction.[4] Although these

---

[4] Change address refers to the output address(es) in a change transaction.

two rules seem plausible, they are not fool-proof in some situations. Kalodner et al. (2020) point out that rule (1) does not apply to CoinJoin transactions.[5] Meanwhile, it is very difficult to detect all the change transactions. One way proposed by Meiklejohn et al. (2013) to detect the change address is if that address has only been used once. However, this method may mislabel some long-term investors and create false super-clusters. In particular, more than 50% of addresses during the sample period only appear once in our dataset. If we follow the above rule, then most of the transactions in our dataset will be falsely labeled as change transactions.

In this paper, we choose to use the following three steps to parse the bitcoin transaction data:

Step 1: Subsample the original transaction dataset to only include addresses that appear at least ten times during a given year.[6]

Step 2: In each transaction, we remove all the output addresses that also show up on the input side as they are likely to be change addresses.

Step 3: In each transaction, we treat all the addresses on the input side of a transaction as being controlled by the same person.[7]

In the first step, we filter out all the users who only use addresses once for any transaction and may care more about transaction privacy than returns, along with the users who do not trade frequently. In this study, we use monthly returns, and weekly returns to measure investors' performance on a short-term horizon. If the address does not trade frequently, we can safely assume this investor does not care that much about the short-term horizon. Meanwhile, removing those less frequent addresses from the dataset can significantly reduce the computing resource requirement and also reduce the number of false super-clusters. In the second step, we did not adopt more aggressive methods to parse out change transactions because we may end up mislabeling some legitimate transactions. Step three also helps us reduce the size of our dataset. These three steps will certainly miss some connections between addresses and mislabel some addresses owned by the same person as they have different owners. However, these steps can significantly reduce computing time by reducing our sample size from more than 700 million addresses to around 10 million addresses. Meanwhile, these steps will not falsely connect some addresses that are not owned by the same person, so we can avoid false super-clusters in our dataset.

In this next section, we will use the parsed dataset to construct the bitcoin investor network.

4. The bitcoin investor network

4.1. Definition

We mostly follow Ozsoylev et al. (2014) to define the bitcoin investor network. However, the bitcoin market is different from the stock market in the following aspects: (1) In the bitcoin market, there is only bitcoin this one asset to trade. (2) Bitcoin transactions are confirmed in batches. When a new block is successfully mined in the bitcoin network, all the transactions inside this block are confirmed at the same time. As a result, we need to modify the approach from Ozsoylev et al. (2014) and define the bitcoin investor network as:

Definition The bitcoin investor network, $\varepsilon^{\Delta b, M}$, is defined such that for each pair of bitcoin addresses, $i, j \neq i$, $\varepsilon^{\Delta b, M} = 1$ if and only if i and j traded in the same direction within Δb blocks at least M times over the period.

To calculate the bitcoin investor network, we choose among 10, 50, and 100 for M. Meanwhile, we set $\Delta b = 1$, which means we keep the time window as 10 min on average. The first reason for keeping the time window Δb short is that the information-driven trading is usually called "fast" trading and happens in a short time window, and we want to separate it from other types of trading. The second reason is to save computing time.

This paper constructs the network based on every nature year, and all the following results are reported based on nature year. The reason behind this decision is that we are concerned about the change in the composition of bitcoin investors given the landscape of the bitcoin market has changed a lot from 2015 to 2020. For example, some addresses may be active in 2015 but not in 2019. By estimating the network using the whole sample, we may under-sample the addresses that are only active in a certain period. Meanwhile, if there is no dramatic investor change, estimating the network by year should give us similar results as using the whole data sample. If there is significant investor change, then estimating the network by years can better capture the dynamic change in the bitcoin investor network. Table 4 provides the summary statistics of the bitcoin investor network for each year and values of M. We can see that, firstly, the average number of links is larger in the first two years compared with the number in the rest years, which may imply some structure changes inside the bitcoin network along with time. Secondly, the fraction of links is small in all periods, which means the network is sparse and a small group of addresses are connected in the network. Lastly, the increase of M means the criteria of being connected increase, so we expect fewer addresses will be counted as connected in the dataset. However, more than 70% of addresses in our dataset are labeled as being connected when we choose $M = 10$, which may imply that this threshold is too low.

---

[5] CoinJoin transactions combine transactions from different spenders into one transaction, in this way outsiders cannot tell which spender paid which recipient(s).

[6] In this subsample, each transaction record may no longer preserve all the involved addresses. However, our goal is calculate address level return, so transaction level incompleteness will not affect our final results.

[7] As we mentioned at the beginning of this subsection, CoinJoin transactions can be exceptions. However, comparing with joining in a CoinJoin transaction to protect privacy, a much easier way is to generate different addresses for different transactions. The addresses in this dataset have already been used more than ten times, we think their owners will be less likely to use CoinJion transactions to protect their identities.

Table 3
Number of unique addresses in each period under different criteria.

Times	Year					
	2015	(%)	2016	(%)	2017	(%)
≥ 1	57,242,340	100.000	95,918,573	100.000	146,236,402	100.000
≥ 2	26,316,601	45.974	50,958,896	53.127	78,599,806	53.748
≥ 3	3,662,592	6.398	6,296,190	6.564	9,970,560	6.818
≥ 5	2,247,942	3.927	3,587,896	3.741	5,172,163	3.537
≥ 10	1,180,137	2.062	1,936,159	2.019	2,388,072	1.633
≥ 50	263,505	0.460	417,304	0.435	376,944	0.258
≥ 100	133,154	0.233	195,872	0.204	150,497	0.103
Times	Year					
	2018	(%)	2019	(%)	2020	(%)
≥ 1	116,451,116	100.000	134,589,160	100.000	168,350,219	100.000
≥ 2	62,112,190	53.338	72,928,241	54.186	81,616,565	48.480
≥ 3	8,153,883	7.002	9,446,343	7.019	10,966,593	6.514
≥ 5	4,090,845	3.513	4,989,575	3.707	5,345,686	3.175
≥ 10	1,835,325	1.576	2,437,863	1.811	2,389,075	1.419
≥ 50	296,489	0.255	415,270	0.309	372,375	0.221
≥ 100	119,584	0.103	166,982	0.124	153,683	0.091

This table shows how many unique transaction addresses are recorded in the bitcoin blockchain in each year. Times $\geq n$ means the number of addresses that appear at least n time in a given period. (%) shows the percentage of addresses that appear at least n times in a given period.

4.2. Network stability

As the bitcoin market is still relatively new and the network is still evolving over time, we estimate the bitcoin investor network for each year individually. However, we do hope the connected addresses are relatively stable along with time. A relatively stable connected group in our network reduces the possibility that we are merely picking out some addresses that happen to be frequent traders in certain years. In the following part, we follow Ozsoylev et al. (2014) and consider two different methods to test the stability of the bitcoin investor network.

In the first method, the null hypothesis assumes that the network links are generated randomly. Suppose a network contains N nodes and k_1 links, then the possibility of two different nodes are linked is k_1/K, where $K = N(N-1)/2$ is the total possible links in the network. Suppose the network in two different periods has k_1 and k_2 links, and $k_1 \ll K, k_2 \ll K$, then the expected number of overlap links ($E_{random}[y]$) in these two periods is:[8]

$$E_{random}[y] \approx \frac{k_1 k_2}{K}$$

where k_1 is the number of links in the first period, k_2 is the number of links in the second period, and K is all the possible links among all the addresses in the network.

In the second method, the null hypothesis assumes the bitcoin investor network is not truly randomly generated, and the degree distribution has heavy tails. Hence, we take degree into consideration, and the degree-adjusted expected number of overlap links ($E_{degree-adjusted}$) between period 1 and period 2 is written as:

$$E_{degree-adjusted} = \frac{k_2}{2k_1 N} \sum_{N}^{i=1} (D_i - 1)^2$$

where N is the number of investors in the network, and D is the degree distribution of the network in period 1.

Table 5 shows the stability of the bitcoin investor network for different M. When $M = 100$, the value of $y/E_{random}[y]$ shows that the actual number of overlap links between any two periods is significantly higher than the expected number of overlap links when the network links are randomly generated. For the degree-adjusted method, the actual number of overlap links is still significantly higher than the second null hypothesis predicted (see $y/E_{degree-adjusted}[y]$). Both methods suggest that the connections in this network are relatively stable over years. Hence, the bitcoin investor network does not change dramatically in our sample period.

The table also reports the results for $M = 10$ and $M = 50$. We still can see that, under the random generation hypothesis, the real overlap link data still reject the null hypothesis that the network is not stable over time. However, the value of $y/E_{degree-adjusted}[y]$ becomes relatively small when we use $M = 10$. These results may indicate that: (1) When we set $M = 10$, the threshold of being identified as connected is low, and many false connections are included in the dataset. (2) high-connected, more central nodes stay active in the bitcoin investor network longer than the rest nodes. Hence, when we set a large value of M and rule out more less-connected nodes, we will observe a more stable network. For these reasons, we use $M = 100$ for our main results, but we also show results under $M = 50$ or $M = 10$ are consistent.

---

[8] The condition that $k_1 \ll K, k_2 \ll K$ is satisfied in our dataset as the fraction of links reported in Table 4 is less than 0.02%, which means the bitcoin investor network is very sparse.

Table 4
Summary statistics for the bitcoin investor network in different periods.

Year	Statistics	$M = 10$	$M = 50$	$M = 100$
2015	Number of links	866,491,905	102,742,324	58,853,778
	Average number of links	1468	174	100
	Fraction of links	0.124%	0.015%	0.009%
	Maximum number of links	829,686	172,208	84,355
	Number of addresses	1,180,137	1,180,137	1,180,137
	Fraction of connected addresses	84.514%	17.032%	8.288%
2016	Number of links	880,959,326	132,851,566	48,193,958
	Average number of links	910	137	50
	Fraction of links	0.047%	0.007%	0.003%
	Maximum number of links	1,190,522	229,413	98,438
	Number of addresses	1,936,159	1,936,159	1,936,159
	Fraction of connected addresses	79.879%	14.838%	6.612%
2017	Number of links	783,698,955	100,945,475	23,577,684
	Average number of links	656	86	20
	Fraction of links	0.028%	0.004%	0.001%
	Maximum number of links	1,100,638	136,329	48,300
	Number of addresses	2,388,072	2,388,072	2,388,072
	Fraction of connected addresses	75.261%	10.302%	3.954%
2018	Number of links	695,480,707	101,483,338	24,242,824
	Average number of links	758	111	26
	Fraction of links	0.041%	0.006%	0.001%
	Maximum number of links	1,265,751	172,791	66,035
	Number of addresses	1,835,325	1,835,325	1,835,325
	Fraction of connected addresses	85.298%	12.143%	4.778%
2019	Number of links	898,022,571	148,242,051	43,763,539
	Average number of links	737	122	36
	Fraction of links	0.030%	0.005%	0.002%
	Maximum number of links	1,476,022	211,356	76,289
	Number of addresses	2,437,863	2,437,863	2,437,863
	Fraction of connected addresses	74.178%	10.960%	4.157%
2020	Number of links	624,374,616	67,527,571	20,417,238
	Average number of links	523	57	17
	Fraction of links	0.022%	0.002%	0.001%
	Maximum number of links	1,391,091	201,422	79,156
	Number of addresses	2,389,075	2,389,075	2,389,075
	Fraction of connected addresses	78.877%	11.000%	4.439%

Fraction of links is equal to average number of links divided by the number of all the potential links an address can have, which is 1,180,136 in 2015, 1,936,158 in 2016, 2,388,071 in 2017, 1,835,324 in 2018, 2,437,862 in 2019 and 2,389,074 in 2020. We can see the bitcoin investor network is sparse, and network characteristics are mostly similar across different periods.

4.3. Centrality and returns

After identifying the bitcoin investor network, we can now measure the centrality of every address (node) in it. While many methods have been proposed to measure network centrality, we follow Ozsoylev et al. (2014) and consider degree centrality and eigenvector centrality in this paper as they directly measure how connected an address is in the network.[9] Degree centrality measures how connected an address is in the network by counting how many other addresses are connected with it. Eigenvector centrality measures the connectedness of an address by considering the importance of all the neighbors connected with it. As we can see, these two measures describe different characteristics of the addresses in this bitcoin investor network. The degree centrality measures how widely connected an address is in this network. The eigenvector centrality tells us how important an address's neighbors are in this network.

To measure bitcoin returns, we follow the same approach as in Barber et al. (2009) and Ozsoylev et al. (2014). We choose $\Delta B = 4320$ as the window length, which is roughly a month given that bitcoin blocks are generated every 10 min on average. Defining the window in terms of blocks can help us precisely match bitcoin returns with the transaction data, which is also confirmed in terms of blocks.[10] For each trade, z, of individual i, we define the return as:

$$\mu_{i,z} = sign * \left(\frac{P^{b+\Delta B} - P^b}{P^b} - r_{tbill} \right) \tag{1}$$

where $P^{b+\Delta B}$ is the bitcoin price B blocks later after the trade happened in block b. When we set $B = 4320$, $\frac{P^{b+\Delta B} - P^b}{P^b}$ gives us roughly the monthly return rate of each trade. r_{tbill} is the daily U.S. Treasury 3-month yield rate, and the data for weekends is

---

[9] A thorough discussion about network centrality, please see Jackson (2010).

[10] In the Appendix, for robustness checks, we also consider 1008 blocks, which correspond to weekly returns. And the results are essential the same.

Table 5
Stability of the bitcoin investor network.

Compared periods	2015–2016	2016–2017	2017–2018	2018–2019	2019–2020
	M=10				
Overlap links, y	88,922,930	76,045,848	47,659,248	141,037,661	111,957,422
$E_{random}[y]$	14200	12860	10184	11650	10477
$y/E_{random}[y]$	6262.002	5913.459	4679.752	12106.090	10686.290
$E_{degree-adjusted}[y]$	1919295	2904363	1894071	2749029	2371696
$y/E_{degree-adjusted}[y]$	46.331	26.183	25.162	51.304	47.207
	M=50				
Overlap links, y	19,849,582	23,140,633	12,688,156	32,286,509	21,635,431
$E_{random}[y]$	294	290	227	323	224
$y/E_{random}[y]$	67325.010	79753.000	55978.990	100002.400	96177.050
$E_{degree-adjusted}[y]$	84,368	90,502	54,490	106,441	77,200
$y/E_{degree-adjusted}[y]$	235.274	255.692	232.853	303.327	280.252
	M=100				
Overlap links, y	13,473,164	14,071,966	10,846,278	13,474,728	12,922,135
$E_{random}[y]$	74	36	21	34	30
$y/E_{random}[y]$	182258.700	387451.400	504552.000	393618.200	424193.800
$E_{degree-adjusted}[y]$	28,690	13,425	6,525	10,835	13,789
$y/E_{degree-adjusted}[y]$	469.607	1048.189	1662.189	1243.585	937.143

This table shows the stability of the bitcoin investor network across the three periods for different M. Overlap links, y, shows the number of intersecting links between two different periods. $E_{random}[y]$ is the expected number of intersecting links between two periods if the network is random. $E_{degree-adjusted}[y]$ is another measure of the expected number of intersection links between two periods that takes the degree distribution of the original network into consideration. Larger $E_{random}[y]$ and $E_{degree-adjusted}[y]$ indicate the network is more stable.

Table 6
Percentage of addresses that showed up multiple years.

Number of years	Number of addresses	Percentage of addresses (%)
1	9,090,646	86.692
2	1,175,234	11.210
3	178,018	1.698
4	30,606	0.292
5	7,864	0.075
6	3,720	0.035

This table display the percentage of addresses that showed up in multiple years in our dataset. Majority of addresses only showed up in one year. And only 0.035% addresses in our dataset are recorded in every year.

imputed based on the neighboring available values.[11] Both rates have been annualized to make sure they are comparable.[12] The symbol *sign* indicates the trade direction, which is negative for input addresses and positive for output addresses. To further reduce the computing time and lower the computing memory requirement, we calculate the average returns for an address if there are multiple trades from this address in a month.

Due to the difficulty of linking all the addresses that are owned by an individual, it is not possible to accurately calculate the trade amount of every investor. As a result, the weighted returns used in Ozsoylev et al. (2014) are highly inaccurate in this study. For example, if there are three addresses, a, b and c, owned by an investor. Suppose c is a safe vault, and the investor usually transfers a large amount of bitcoin into c and moves a small amount of bitcoin out of c when the bitcoin price is increasing rapidly. Without knowing that these addresses are owned by the same person and all the transactions among them are change transactions, the weighted returns will be heavily biased by returns caused by those change addresses. To mitigate the effect of these undetected change transactions, we think the definition of returns in Eq. (1) is a better candidate to measure investors' performance. Eq. (1) measures how often an address can show up at the right side of a trade at the right time and ignores the trading amount due to incomplete trading information on the individual investor level.

After calculating address returns, we first plot out the return difference between the connected and unconnected addresses to see if there is an obvious difference between these two groups. The connected group includes addresses that have at least one connection with other addresses following the definition in Section 4.1. And the unconnected group includes the other addresses in the dataset.

---

[11] The daily U.S. Treasury 3-month yield rate can be found on the U.S. department of the Treasury website. Eq. (1) represents excess returns. However, because the magnitude of r_{bill} is much smaller than the first term in Eq. (1) and r_{bill} also changes less frequently than the first term, adding r_{tbill} or not does not have an observable impact on our results.

[12] To annualize monthly return rates, we multiply the data with $365/30.5$. Following the same logic, we use $365/7$ for weekly return rates. We did not use the formula $(1 + r)^t - 1$ because return rates can be very high in some months, $(1 + r)^t - 1$ may make the performance of an address in a high bitcoin return month dominates its whole year performance.
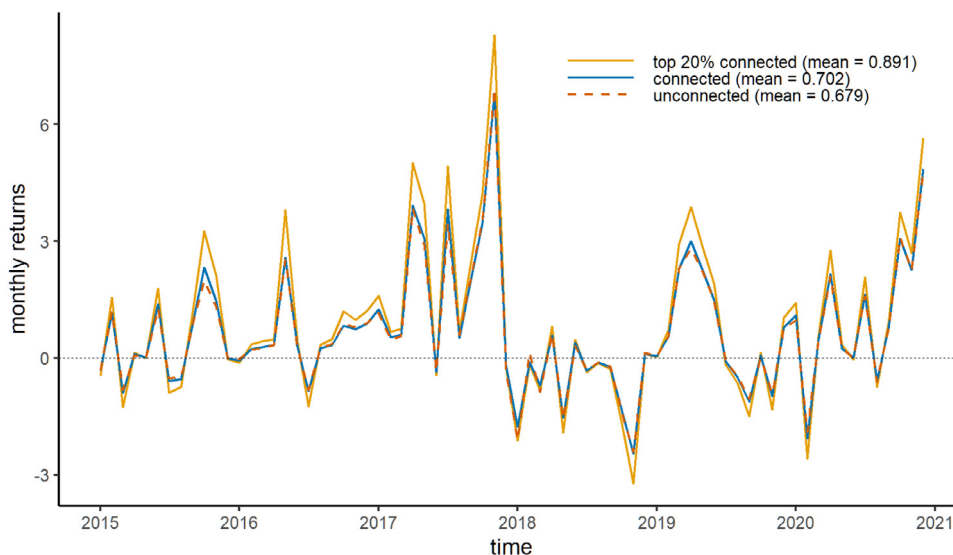
Fig. 3. Average monthly returns of top connected, connected and unconnected groups In this figure, addresses are divided into top 20% most-connected, connected and unconnected group based on the threshold $M = 100$, and then the average monthly returns of all the addresses in each group are calculated for each month.

Table 7
Summary statistics for regression variables, all addresses.

		Pool	2015	2016	2017	2018	2019	2020
Dependent variable								
	Mean	0.768	0.448	0.507	2.157	−0.634	0.666	1.175
Monthly returns	Median	0.181	0.107	0.329	1.203	−0.303	0.121	0.518
	S.D.	2.538	1.545	1.358	3.707	1.754	2.043	2.601
	Mean	0.702	0.466	0.495	2.091	−0.770	0.600	1.087
Weekly returns	Median	0.225	0.193	0.224	1.498	−0.300	0.097	0.450
	S.D.	3.674	2.805	2.140	4.667	3.923	3.480	3.380
Independent variables								
(1) Address level								
	Mean	−2.606	−2.545	−2.350	−2.053	−2.641	−3.091	−2.808
Trade volume	Median	−2.612	−2.301	−1.865	−1.979	−2.784	−3.084	−2.892
	S.D.	3.142	3.940	3.681	2.909	2.793	2.994	2.784
	Mean	1.544	1.765	1.725	1.553	1.387	1.499	1.475
Number of trades	Median	1.386	1.609	1.609	1.386	1.386	1.386	1.386
	S.D.	1.114	1.149	1.190	1.114	1.059	1.072	1.084
(2) Market level								
	Mean	5.391	2.725	3.039	5.667	6.430	6.204	6.471
Price volatility	Median	5.934	2.810	2.767	5.757	6.592	6.572	6.203
	S.D.	1.665	0.729	0.670	1.102	0.736	0.853	0.769
	Mean	17.224	17.487	17.979	17.781	16.879	16.871	16.678
Trade volume	Median	17.087	17.309	17.886	17.832	16.803	16.833	16.655
	S.D.	0.577	0.460	0.457	0.215	0.200	0.236	0.077
	Mean	15.812	15.146	15.734	15.959	15.683	15.962	16.011
Number of trades	Median	15.924	15.078	15.727	15.925	15.659	15.944	16.008
	S.D.	0.278	0.231	0.102	0.123	0.136	0.101	0.051
Observations			50,808,976	4,798,982	7,699,212	9,128,361	8,713,662	10,319,744	10,148,626

This table shows the summary statistics of the control variables in the regression analysis. Definition of these variables can be found in Section 5.1. All the independent variables are in log.

Based on the node centrality, we also plot out the returns of the top 20% most-connected addresses.[13] Fig. 3 shows that returns for the unconnected and connected groups almost overlap. The connected group, on average, merely earns 2% higher annualized

---

[13] Here we plot the results based on degree centrality. Using either degree or eigenvector to pick the top 20% most-connected addresses produces similar results.

Table 8
Summary statistics for regression variables, connected addresses.

		Pool	2015	2016	2017	2018	2019	2020
Dependent variable								
Monthly returns	Mean	0.702	0.446	0.499	2.183	−0.590	0.636	1.156
	Median	0.224	0.087	0.398	1.303	−0.236	0.092	0.580
	S.D.	2.166	1.430	1.262	3.328	1.399	1.808	2.400
Weekly returns	Mean	0.635	0.419	0.483	2.107	−0.742	0.553	1.087
	Median	0.227	0.194	0.284	1.673	−0.350	0.113	0.454
	S.D.	2.692	2.152	1.630	3.585	3.016	2.473	2.528
Independent variables								
(1) Variables of interest								
Degree	Mean	0.088	0.117	0.109	0.111	0.152	0.099	0.085
	Median	0.004	0.003	0.005	0.008	0.014	0.004	0.006
	S.D.	0.173	0.246	0.203	0.197	0.239	0.210	0.173
Eigenvector	Mean	0.034	0.205	0.042	0.037	0.043	0.076	0.066
	Median	0.002	0.137	0.003	0.001	0.007	0.001	0.010
	S.D.	0.133	0.183	0.170	0.154	0.132	0.226	0.167
(2) Control variables								
(a) Address level								
Trade volume	Mean	−0.916	−1.540	−0.975	−0.703	−1.037	−0.848	−0.501
	Median	−1.088	−1.639	−0.790	−0.956	−1.394	−1.209	−0.857
	S.D.	3.152	3.536	3.372	2.946	2.858	2.913	2.985
Number of trades	Mean	3.080	3.053	3.119	3.103	2.990	3.061	3.098
	Median	3.258	3.296	3.401	3.367	3.178	3.178	3.178
	S.D.	1.198	1.117	1.236	1.246	1.098	1.192	1.227
(b) Market Level								
Price volatility	Mean	4.980	2.667	3.041	5.657	6.386	6.179	6.454
	Median	5.470	2.556	2.767	5.757	6.592	6.572	6.203
	S.D.	1.805	0.729	0.669	1.087	0.730	0.851	0.762
Trade volume	Mean	17.277	17.436	17.959	17.781	16.864	16.860	16.678
	Median	17.143	17.309	17.886	17.832	16.803	16.833	16.655
	S.D.	0.588	0.425	0.447	0.214	0.187	0.233	0.077
Number of trades	Mean	15.747	15.124	15.735	15.953	15.682	15.957	16.011
	Median	15.831	15.078	15.727	15.925	15.659	15.944	16.008
	S.D.	0.322	0.216	0.100	0.123	0.130	0.101	0.051
Observations		4,609,688	721,816	970,325	645,406	680,337	786,836	807,097

This table shows the summary statistics of the control variables in the regression analysis. Definition of these variables can be found in Section 5.1. All the independent variables are in log. Numbers of observations in yearly datasets do not exactly sum up to that in the pooled dataset due to the 0.1% cutoff.

returns than the unconnected one over the full sample. However, the top 20% most-connected addresses earn, on average, almost 20% higher returns than the other two groups.

When we reduce the threshold to $M = 50$ or 10, the pattern remains the same, but the return difference between the connected and unconnected groups increases (see Figure A.1 and Figure A.2). This result may imply that some connected addresses with high returns are categorized as unconnected when we set a high threshold M, so we observe the return difference between the connected and unconnected group shrinks when M increases. However, returns for the top 20% most-connected addresses are always significantly higher than those for the other groups. To check whether the results are sensitive to the time horizon of returns, we also present results using weekly returns under $M = 10$, $M = 50$, or $M = 100$ (see Figure A.3, Figure A.4 and Figure A.5). All results consistently show that the top 20% most-connected addresses are earning higher returns and the return difference between the connected and unconnected groups is relatively small.

While figures are straightforward, they can be misleading as we have not considered other important factors that can affect returns. In the next section, to control for the other factors, we run multivariate regressions at the address level to see how connectedness is related to returns.

5. Results

Based on what we have observed in Fig. 3, in this section, we try to answer three empirical questions in this section: (1) Comparing with addresses that are unconnected in the bitcoin investor network, do connected addresses earn a higher return? (2) Do more connected addresses earn a higher return than less connected ones? (3) When we use centrality to measure an address'

Table 9

Summary statistics for regression variables, top 20% most-connected addresses.

		Pool	2015	2016	2017	2018	2019	2020
Dependent variable								
	Mean	0.679	0.583	0.688	2.622	−0.742	0.823	1.286
Monthly returns	Median	0.359	0.099	0.493	1.765	−0.304	0.186	0.850
	S.D.	2.257	1.481	1.309	3.072	1.410	1.932	2.484
	Mean	0.565	0.529	0.705	2.511	−0.960	0.640	1.250
Weekly returns	Median	0.219	0.305	0.517	2.268	−0.567	0.176	0.826
	S.D.	2.787	2.086	1.526	3.139	3.031	2.466	2.421
Independent variables								
(1) Variables of Interest								
	Mean	0.310	0.587	0.376	0.375	0.438	0.409	0.306
Degree	Median	0.262	0.663	0.309	0.344	0.422	0.374	0.251
	S.D.	0.232	0.320	0.265	0.243	0.256	0.279	0.247
	Mean	0.212	0.280	0.222	0.206	0.166	0.381	0.282
Eigenvector	Median	0.040	0.249	0.006	0.008	0.055	0.253	0.129
	S.D.	0.292	0.205	0.353	0.321	0.260	0.382	0.295
(2) Control Variables								
(a) Address Level								
	Mean	−0.885	−1.057	−1.072	−0.878	−1.363	−1.204	0.044
Trade volume	Median	−1.161	−1.058	−0.968	−1.161	−1.669	−1.495	−0.671
	S.D.	2.611	3.004	2.936	2.862	2.014	2.175	2.848
	Mean	3.331	3.581	3.466	3.385	3.144	3.052	3.628
Number of trades	Median	3.401	3.434	3.466	3.434	3.401	3.332	3.434
	S.D.	1.107	0.833	1.048	1.093	0.767	1.049	1.421
(b) Market Level								
	Mean	5.543	2.655	3.040	5.579	6.366	6.170	6.447
Price Volatility	Median	6.004	2.556	2.767	5.384	6.592	6.572	6.203
	S.D.	1.551	0.729	0.669	1.068	0.727	0.853	0.750
	Mean	17.161	17.427	17.948	17.775	16.856	16.860	16.682
Trade volume	Median	16.884	17.298	17.886	17.832	16.803	16.833	16.659
	S.D.	0.578	0.420	0.443	0.217	0.180	0.232	0.079
	Mean	15.841	15.120	15.737	15.951	15.682	15.956	16.008
Number of trades	Median	15.851	15.078	15.727	15.925	15.659	15.944	16.005
	S.D.	0.173	0.214	0.100	0.121	0.127	0.101	0.051
Observations		653,478	85,733	170,524	110,086	123,908	152,593	144,944

This table shows the summary statistics of the control variables in the regression analysis. Definition of these variables can be found in Section 5.1. All the independent variables are in log. Numbers of observations in yearly datasets do not exactly sum up to that in the pooled dataset due to the 0.1% cutoff.

connectedness, which centrality measure is more correlated to higher returns. For the rest of this paper, our main results are based on $M = 100$, and $M = 50, 10$ are used only for robustness checks.

Given the questions we want to answer, panel regression seems to be the most suitable choice. However, Table 6 shows the percentage of addresses that showed up in multiple years in our dataset. Among the 10,486,088 unique addresses in our dataset, only 0.035% of them (3720 addresses) are active in the whole examined period. The majority of these addresses are only active in a particular year. This phenomenon is related to how bitcoin addresses are generated in the bitcoin network: the cost of generating a new address is so low that users can generate a new address and abandon the old one whenever they want to do so. This disposable address mechanism makes it infeasible to establish panel data to study investors' behavior in the bitcoin market. In the following sections, we will use pooled regressions and regressions on a yearly basis to capture the underlying structural change in the bitcoin market and answer the above three questions.

5.1. Do connected addresses earn higher returns?

To answer the first question, we run multivariate regressions at the address level to see how connectedness is related to returns. We divide the addresses into two groups based on their connectedness. The connected group includes addresses that have at least one connection with other addresses. The unconnected group includes the rest addresses in our dataset. We also truncate the addresses at the top 1% in terms of centrality to reduce the influence of outliers.[14] Meanwhile, the regression also includes some control variables: the monthly market trade volume, the monthly market number of trades, the monthly individual trade volume, the

---

[14] We also truncated our sample at the top 2% or 0.1%, all of them give us consistent final results.

Table 10
Connectedness and returns.

	Monthly returns		
	(1)	(2)	(3)
Variable of interest			
Connectedness	0.121***	0.121***	0.121***
	(0.001)	(0.032)	(0.046)
Other control variables			
Trade volume	−0.047***	−0.047***	−0.047**
	(0.000)	(0.002)	(0.021)
Number of trades	−0.011***	−0.011	−0.011
	(0.000)	(0.022)	(0.038)
Market price volatility	0.264***	0.264***	0.264
	(0.000)	(0.006)	(0.350)
Market trade volume	0.542***	0.542***	0.542
	(0.001)	(0.024)	(0.716)
Market number of trades	−0.235***	−0.235***	−0.235
	(0.003)	(0.077)	(1.182)
Adjusted R^2	0.131		
Observations	50,808,426		

This table display the results from the regression of monthly returns on connectedness. All the control variables are in log. In regression (1), we control the year fixed effect. In regression (2), we control the year fixed effect and cluster s.e. at Connectedness level, and adjust it for the few clusters bias. In regression (3), we control the year fixed effect and cluster s.e. at ×Years level. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Table 11
Connectedness and returns in each year.

	Monthly returns					
	2015	2016	2017	2018	2019	2020
Variable of interest						
Connectedness	0.022***	−0.026***	−0.014***	0.016***	0.269***	0.217***
	(0.002)	(0.002)	(0.005)	(0.002)	(0.002)	(0.003)
Other control variables						
Trade volume	−0.050***	−0.058***	−0.155***	0.037***	−0.007***	−0.073***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of trades	0.032***	0.070***	0.163***	−0.027***	−0.144***	−0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Market price volatility	−0.077***	0.061***	0.014***	−0.652***	−1.187***	1.936***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Market trade volume	0.089***	0.300***	2.859***	1.312***	2.580***	−10.823***
	(0.006)	(0.001)	(0.007)	(0.004)	(0.003)	(0.009)
Market number of trades	0.743***	3.238***	−0.082***	−5.872***	8.375***	−3.001***
	(0.010)	(0.005)	(0.010)	(0.005)	(0.008)	(0.012)
Adjusted R^2	0.026	0.067	0.039	0.176	0.344	0.330
Observations	4,798,934	7,699,178	9,128,361	8,713,631	10,319,744	10,148,578

This table display the results from the regression of monthly returns on connectedness for each year. All the control variables are in log. Robust regression is conducted to account for heteroscedasticity. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

monthly individual number of trades, and the monthly bitcoin price volatility, which is calculated as the standard deviation of bitcoin prices in each month. All variables are in log. The summary statistics of these addresses are reported in Table 7.

We first run the following pooled regression on the whole dataset to study the relationship between return and connectedness in the whole examined period:

$$r_{i,t} = \beta_0 + \beta_1 connectedness_{i,t} + \beta_j ctrl\_vars_{i,t} + \epsilon_{i,t} \qquad (2)$$

where $r_{i,t}$ is the return of address i at time t, $connectedness$ is a dummy variable that denotes the address is connected with other addresses or not, $ctrl\_vars$ are control variables. And the results are reported in Table 10. To address the concern that undetected correlations among some subgroups in our dataset and the standard errors can be biased (Petersen, 2009), we used three different ways to calculate the standard errors to ensure the results presented here are robust. In regression (1), we include the year fixed effect. In regression (2), we include the year fixed effect, and the standard error is clustered at the connectedness level, and the bias-correcting method proposed in Cameron and Miller (2015) has been adopted to address the few clusters issue.[15] In regression

---

[15] Even though we have used the bias-correcting method to address the few clusters issue, it is still unconventional to use a cluster that only has two values. Hence, the standard errors calculated in this regression should be used for reference only.

Table 12
Different degree centrality groups and returns.

	Monthly returns		
	(1)	(2)	(3)
Groups by degree centrality			
Group 1	0.016***	0.016	0.016
	(0.002)	(0.015)	(0.048)
Group 2	0.052***	0.052***	0.052
	(0.003)	(0.016)	(0.053)
Group 3	0.072***	0.072***	0.072
	(0.002)	(0.018)	(0.059)
Group 4	0.113***	0.113***	0.113**
	(0.002)	(0.021)	(0.056)
Group 5	0.092***	0.092***	0.092
	(0.002)	(0.025)	(0.061)
Group 6	0.074***	0.074**	0.074
	(0.002)	(0.029)	(0.051)
Group 7	0.104***	0.104***	0.104**
	(0.002)	(0.034)	(0.047)
Group 8	0.128***	0.128***	0.128*
	(0.002)	(0.038)	(0.068)
Group 9	0.234***	0.234***	0.234**
	(0.002)	(0.038)	(0.109)
Group 10	0.326***	0.326***	0.326***
	(0.002)	(0.039)	(0.105)
Other control variables			
Trade volume	−0.046***	−0.046***	−0.046**
	(0.000)	(0.001)	(0.018)
Number of trades	−0.026***	−0.026	−0.026
	(0.000)	(0.023)	(0.035)
Market price volatility	0.263***	0.263***	0.263
	(0.001)	(0.010)	(0.297)
Market trade volume	0.544***	0.544***	0.544
	(0.001)	(0.009)	(0.602)
Market number of trades	−0.231***	−0.231**	−0.231
	(0.003)	(0.098)	(1.011)
Adjusted R^2	0.132		
Observations	50,711,571		

This table display the results from the regression of monthly returns on groups with different degree centrality. All the control variables are in log. In regression (1), we control the year fixed effect. In regression (2), we control the year fixed effect and cluster s.e. at Groups level, and adjust it for the few clusters bias. In regression (3), we control the year fixed effect and cluster s.e. at Groups × Years level. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

(3), we include the year fixed effect, and the standard error is clustered at the connectedness-year level with bias-correction. We can see that no matter which method we use, being connected is always related to higher returns in the data we examined.

Furthermore, to study the underlying structural changes in the bitcoin market, we run this regression on a yearly basis, and the results are reported in Table 11. We can see that connectedness is not always associated with higher returns. One noticeable change is that the coefficient becomes much larger in 2019 and 2020, implying that connectedness has become more important recently. Changing the threshold M to 50 or 10 or use weekly returns instead of monthly returns produces roughly the same results and the conclusion still holds (see Online Appendix Table A.1 to Table A.10).

5.2. Do more connected addresses earn higher returns?

We have just shown that the relationship between connectedness and higher returns varies over time. However, remember that Fig. 3 shows a significant return difference between the top 20% most-connected addresses and the other two groups. Naturally, the next question we want to answer is whether being more connected matters in the bitcoin network.

We use centrality to measure the connectedness of addresses. As mentioned before, we use both degree centrality and eigenvector centrality in this paper. Degree centrality measures how connected a node is by counting the number of nodes connected to it. Presumably, a more connected node plays a more critical role in spreading information in the network as an information hub. Being a hub may give this node access to more exclusive information about the market and help the node earn higher returns. However, it is possible that some addresses may not have high degree centrality but become essential information hubs by occupying important positions in the network and being connected to some influential nodes. That is why we also consider the eigenvector centrality, which measures how important a node is by looking at how important its neighbors are in this analysis. Eigenvector centrality is

Table 13
Different eigenvector centrality groups and returns.

	Monthly returns		
	(1)	(2)	(3)
Groups by eigenvector centrality			
Group 1	0.074***	0.074***	0.074
	(0.002)	(0.015)	(0.062)
Group 2	0.017***	0.017	0.017
	(0.002)	(0.016)	(0.054)
Group 3	−0.001	−0.001	−0.001
	(0.002)	(0.017)	(0.050)
Group 4	0.048***	0.048***	0.048
	(0.002)	(0.019)	(0.062)
Group 5	0.080***	0.080***	0.080
	(0.002)	(0.022)	(0.073)
Group 6	0.114***	0.114***	0.114***
	(0.002)	(0.026)	(0.039)
Group 7	0.106***	0.106***	0.106***
	(0.002)	(0.030)	(0.041)
Group 8	0.161***	0.161***	0.161***
	(0.002)	(0.036)	(0.051)
Group 9	0.243***	0.243***	0.243***
	(0.002)	(0.043)	(0.078)
Group 10	0.308***	0.308***	0.308***
	(0.002)	(0.039)	(0.088)
Other control variables			
Trade volume	−0.046***	−0.046***	−0.046***
	(0.000)	(0.001)	(0.018)
Number of trades	−0.027***	−0.027	−0.027
	(0.000)	(0.023)	(0.035)
Market price volatility	0.264***	0.264***	0.264
	(0.000)	(0.005)	(0.299)
Market trade volume	0.544***	0.544***	0.544
	(0.001)	(0.018)	(0.609)
Market number of trades	−0.235***	−0.235***	−0.235
	(0.003)	(0.080)	(1.018)
Adjusted R^2	0.132		
Observations	50,726,228		

This table display the results from the regression of monthly returns on groups with different eigenvector centrality. All the control variables are in log. In regression (1), we control the year fixed effect. In regression (2), we control the year fixed effect and cluster s.e. at Groups level, and adjust it for the few clusters bias. In regression (3), we control the year fixed effect and cluster s.e. on Groups × Years level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

calculated using iGraph package in R, and the value has been normalized to be between 0 and 1. Due to limited computing power, we only report the results for $M = 50,100$.[16]

To carefully examine the relationship between returns and centrality among connected addresses in the whole period, we first run a pooled regression for returns on deciles with different centrality:

$$r_{i,t} = \beta_0 + \beta_1 g_{i,t}^1 + \beta_2 g_{i,t}^2 + \cdots + \beta_{10} g_{i,t}^{10} + \beta_j ctrl\_vars_{i,t} + \epsilon_{i,t} \tag{3}$$

where $r_{i,t}$ is the return of address i at time t, g^n denotes the different connected decile groups, $ctrl\_vars$ are control variables. In this regression, the ten connected decile groups are categorized using either degree centrality or eigenvector centrality and labeled by ten dummies. Hence, the base group in the regression result is the unconnected addresses, and the 1–10 groups in regression results represent the least connected decile to the most-connected decile. In the regressions, we control bitcoin price volatility, market number of trades, market trade volume, individual trade volume and individual number of trades.

Same as the last section, here we use three different regressions to test the robustness of our results. In regression (1), we include the year fixed effect. In regression (2), we include the year fixed effect, and the standard error is clustered at the group level.[17] In regression (3), we include the year fixed effect, and the standard error is clustered at the group-year level. Tables 12 and 13 show

[16] In this study, Eigenvector centrality is calculated on Virgina Tech ARC servers using 20 cores (128 GB shared memory). Given the size of the network, more computing resources are needed to calculate the Eigenvector centrality when M = 10. However, given that the results of $M = 50$ also serves as robustness checks and the network is more stable when $M = 50$ compared with the network when $M = 10$, we do not think that the lack of results of $M = 10$ has any impact on our final conclusion.

[17] In this regression, there are eleven groups after counting unconnected addresses as one group, we still adopted the bias-corrected cluster-robust variance matrix method (Cameron and Miller, 2015) to address the few clusters issue in our regression.

Table 14

Different degree centrality groups and returns in each year.

	Monthly returns					
	2015	2016	2017	2018	2019	2020
Groups by degree centrality						
Group 1	−0.055***	−0.038***	−0.205***	0.045***	0.151***	0.080***
	(0.002)	(0.013)	(0.034)	(0.004)	(0.018)	(0.019)
Group 2	−0.036***	−0.079***	−0.167***	0.062***	0.222***	0.087***
	(0.003)	(0.014)	(0.038)	(0.004)	(0.021)	(0.020)
Group 3	−0.038***	−0.034**	−0.187***	0.101***	0.260***	0.132***
	(0.003)	(0.015)	(0.041)	(0.005)	(0.023)	(0.022)
Group 4	−0.024***	0.005	−0.184***	0.062***	0.296***	0.171***
	(0.004)	(0.017)	(0.045)	(0.006)	(0.026)	(0.025)
Group 5	−0.015***	−0.021	−0.220***	0.048***	0.271***	0.174***
	(0.004)	(0.020)	(0.053)	(0.007)	(0.030)	(0.029)
Group 6	−0.003	−0.047**	−0.170***	0.037***	0.231***	0.102***
	(0.004)	(0.024)	(0.061)	(0.009)	(0.036)	(0.033)
Group 7	−0.010**	−0.039	0.012	−0.003	0.301***	0.092**
	(0.005)	(0.028)	(0.068)	(0.010)	(0.044)	(0.037)
Group 8	0.074***	0.025	0.155**	−0.131***	0.278***	0.180***
	(0.006)	(0.033)	(0.073)	(0.011)	(0.053)	(0.045)
Group 9	0.122***	0.055*	0.521***	−0.135***	0.340***	0.327***
	(0.006)	(0.030)	(0.071)	(0.010)	(0.050)	(0.049)
Group 10	0.163***	0.108***	0.578***	−0.145***	0.474***	0.477***
	(0.007)	(0.032)	(0.078)	(0.011)	(0.047)	(0.049)
Other control variables						
Trade volume	−0.049***	−0.057***	−0.151***	0.036***	−0.007*	−0.072***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.004)	(0.003)
Number of trades	0.022***	0.063***	0.140***	−0.017**	−0.168***	−0.025
	(0.004)	(0.017)	(0.045)	(0.007)	(0.033)	(0.029)
Market price volatility	−0.077***	0.061***	0.007	−0.651***	−1.186***	1.936***
	(0.010)	(0.005)	(0.009)	(0.009)	(0.021)	(0.012)
Market trade volume	0.093**	0.301***	2.856***	1.311***	2.581***	−10.819***
	(0.043)	(0.009)	(0.061)	(0.025)	(0.057)	(0.095)
Market number of trades	0.736***	3.234***	−0.043	−5.874***	8.382***	−2.991***
	(0.033)	(0.055)	(0.057)	(0.074)	(0.128)	(0.058)
Adjusted R^2	0.027	0.067	0.040	0.176	0.345	0.331
Observations	4,786,299	7,682,846	9,111,915	8,698,574	10,300,872	10,131,065

This table display the results from the regression of monthly returns on groups with different degree centrality for each year. All the control variables are in log, and s.e. is clustered at Groups level with bias-correction. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

that no matter these ten decile groups are categorized using degree centrality or eigenvector centrality, addresses in the top decile groups are earning significantly higher returns, and this pattern is most obvious for the addresses in the top two groups.

Tables 14 and 15 show the yearly-based results using either degree or eigenvector centrality to divide the connected addresses into ten groups and use the unconnected addresses as the base group. The standard errors are clustered at the group level with bias-correction. From the regression results, we can reach three conclusions: (1) Except in 2018, addresses in the top two decile groups have higher returns than addresses in the rest connected or unconnected groups. In certain years, addresses in these two groups, on average, earn 40% higher returns than the unconnected addresses and 20% higher returns than other connected addresses. (2) In the early years, addresses in the bottom connected groups are earning even lower returns than addresses in the unconnected group. (3) In 2019 and 2020, all connected addresses consistently earn higher returns than unconnected addresses. For example, connected addresses in the bottom 10% earn at least 5.7% higher returns than their unconnected peers, which is larger than the 2% return difference we observed in Fig. 3. Furthermore, among these connected addresses, we also observe the pattern that addresses in higher decile groups generally earn higher returns than the ones in lower decile groups.

These results explain why connectedness is not always associated with higher returns in our sample. The addresses in the bottom connected groups have lower returns than unconnected addresses in the early years, but this pattern goes away in the recent two years. The return difference change between the unconnected addresses and addresses with few connections may imply the demographic change of a certain group of investors in the bitcoin network: the investors who have connections but are unwilling to use blockchain to transfer bitcoins. The identification rule employed in this paper can only detect addresses with connections if those addresses actively trade in the network. Some addresses with connections may be inactive traders, which will be mislabeled as unconnected. High returns associated with these addresses will level up the average return of the unconnected addresses. However, in recent years, mainstream trading platforms begin to provide bitcoin trading, and they are more convenient than transferring bitcoin through blockchain. Those inactive but well-connected traders may have changed their trading platform accordingly. What we observe in the last two years may further confirm the necessity to construct the bitcoin network by each year. Otherwise, we may miss the structural changes that happened inside the network. We also changed M to 50 and use weekly returns instead of monthly returns for robustness checks, and all the conclusions still hold (see Online Appendix Table A.11 to Table A.22).

Table 15

Different eigenvector centrality groups and returns in each year.

	Monthly returns					
	2015	2016	2017	2018	2019	2020
Groups by eigenvector centrality						
Group 1	0.049***	−0.007	0.265***	−0.133***	0.140***	0.057***
	(0.002)	(0.013)	(0.040)	(0.009)	(0.018)	(0.018)
Group 2	−0.060***	−0.032**	−0.254***	−0.005	0.160***	0.133***
	(0.003)	(0.014)	(0.036)	(0.008)	(0.021)	(0.020)
Group 3	−0.121***	−0.111***	−0.165***	0.047***	0.183***	0.055***
	(0.003)	(0.015)	(0.041)	(0.007)	(0.024)	(0.020)
Group 4	−0.023***	−0.031*	−0.273***	0.088***	0.215***	0.121***
	(0.003)	(0.016)	(0.045)	(0.007)	(0.028)	(0.022)
Group 5	−0.007*	−0.005	−0.355***	0.094***	0.270***	0.161***
	(0.004)	(0.019)	(0.049)	(0.009)	(0.031)	(0.026)
Group 6	−0.018***	0.001	−0.017	0.050***	0.234***	0.167***
	(0.004)	(0.022)	(0.053)	(0.012)	(0.036)	(0.030)
Group 7	0.003	0.028	−0.016	0.026*	0.231***	0.103***
	(0.005)	(0.025)	(0.061)	(0.014)	(0.042)	(0.033)
Group 8	0.030***	−0.042	0.062	0.011	0.346***	0.190***
	(0.005)	(0.030)	(0.072)	(0.016)	(0.053)	(0.041)
Group 9	0.105***	0.018	0.332***	−0.081***	0.381***	0.350***
	(0.006)	(0.038)	(0.089)	(0.017)	(0.059)	(0.052)
Group 10	0.191***	0.091***	0.434***	−0.114***	0.463***	0.467***
	(0.008)	(0.033)	(0.084)	(0.017)	(0.043)	(0.048)
Other control variables						
Trade volume	−0.050***	−0.057***	−0.153***	0.036***	−0.008***	−0.072***
	(0.002)	(0.001)	(0.004)	(0.002)	(0.003)	(0.003)
Number of trades	0.021***	0.063***	0.144***	−0.021**	−0.165***	−0.026
	(0.003)	(0.018)	(0.048)	(0.011)	(0.033)	(0.028)
Market price volatility	−0.077***	0.061***	0.009	−0.652***	−1.186***	1.936***
	(0.010)	(0.005)	(0.007)	(0.008)	(0.014)	(0.011)
Market trade volume	0.095**	0.301***	2.855***	1.311***	2.580***	−10.813***
	(0.045)	(0.007)	(0.053)	(0.025)	(0.036)	(0.082)
Market number of trades	0.730***	3.232***	−0.053	−5.870***	8.377***	−2.989***
	(0.044)	(0.047)	(0.046)	(0.078)	(0.085)	(0.053)
Adjusted R^2	0.027	0.067	0.040	0.176	0.345	0.331
Observations	4,790,487	7,682,300	9,115,062	8,700,653	10,301,864	10,135,862

This table display the results from the regression of monthly returns on groups with different eigenvector centrality for each year. All the control variables are in log, and s.e. is clustered at Groups level with bias-correction. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

The results so far suggest that the top 20% most-connected addresses have significantly higher returns than the rest addresses for the most of sample period. In addition, the relationship between connectedness and higher returns is becoming more obvious in 2019 and 2020: all connected groups show higher returns than the unconnected group. And among the connected addresses, more connections are also related to higher returns.

5.3. Which centrality measure matters more?

So far, we have not distinguished the measures of centrality. As mentioned in Section 4.3, the degree centrality tends to measure the quantity of connections and the eigenvector centrality tends to measure the quality of connections an address has. In this subsection, we try to figure out which centrality measure matters more.

First, we run regressions for returns on both centrality measures for all the connected addresses:

$$r_{i,t} = \beta_0 + \beta_1 eigenvector_{i,t} + \beta_2 degree_{i,t} + \beta_j ctrl\_vars_{i,t} + \epsilon_{i,t} \tag{4}$$

where $r_{i,t}$ is the return of address i at time t, $eigenvector_{i,t}$ denotes address i's eigenvector centrality in year t, and similarly, $degree_{i,t}$ denotes address i's degree centrality in year t, $ctrl\_vars$ are control variables. To make these two measures comparable, both centrality measures have been rescaled to be between 0 and 1. Following the previous regressions, we also include bitcoin price volatility, market number of trades, market trade volume, individual trade volume, and individual number of trades as control variables. The summary statistics for these connected addresses can be found in Table 8.

Regression results are reported in Table 16. The first two columns show the relationship between return and both centrality measures across the whole examined period. In column one, we control the year fixed effect, and in column two, we also clustered the standard error at year level with bias-correction. Both results show that the two measures used in this paper are positively correlated with returns in the bitcoin market. In the following columns, the relationship is examined on a yearly basis. Our results show that the coefficients on both centrality measures are positive and large in most years. For example, in 2020, when eigenvector or degree centrality increases by 0.1, we see a 2.33% or 4.20% increase in returns respectively. The results show that having

Table 16

Centrality measures and monthly returns.

	Monthly returns							
	$Pool_f$	$Pool_{fc}$	2015	2016	2017	2018	2019	2020
Centrality measures								
Eigenvector	0.298***	0.298**	0.207***	0.168***	0.449***	−0.055***	0.128***	0.233***
	(0.009)	(0.153)	(0.012)	(0.011)	(0.031)	(0.013)	(0.013)	(0.021)
Degree	0.294***	0.294***	0.203***	0.246***	0.919***	−0.325***	0.314***	0.420***
	(0.006)	(0.114)	(0.008)	(0.009)	(0.025)	(0.007)	(0.015)	(0.021)
Other control variables								
Trade volume	−0.057***	−0.057***	−0.057***	−0.057***	−0.176***	0.032***	−0.015***	−0.082***
	(0.000)	(0.022)	(0.001)	(0.000)	(0.002)	(0.001)	(0.001)	(0.001)
Number of trades	0.069***	0.069*	0.038***	0.124***	0.311***	−0.024***	−0.041***	0.094***
	(0.001)	(0.040)	(0.002)	(0.001)	(0.005)	(0.002)	(0.002)	(0.003)
Market price volatility	0.245***	0.245	−0.110***	0.086***	−0.079***	−0.599***	−1.205***	1.940***
	(0.002)	(0.356)	(0.003)	(0.001)	(0.005)	(0.003)	(0.003)	(0.004)
Market trade volume	0.409***	0.409	0.218***	0.329***	3.163***	1.195***	2.641***	−10.904***
	(0.002)	(0.591)	(0.013)	(0.002)	(0.024)	(0.011)	(0.007)	(0.026)
Market number of trades	0.149***	0.149	0.729***	3.128***	0.385***	−5.536***	8.455***	−3.222***
	(0.007)	(1.051)	(0.023)	(0.012)	(0.032)	(0.016)	(0.023)	(0.039)
Adjusted R^2	0.154	0.154	0.042	0.079	0.059	0.225	0.452	0.381
Observations	4,613,081	4,613,081	721,816	970,623	645,993	682,814	786,836	807,102

This table displays the results from the regression of monthly returns on addresses with different centrality, controlling for other variables. Centrality is measured using either degree centrality or eigenvector centrality. $Pool_f$ denotes the pooled regression with year fixed effect being controlled. $Pool_{fc}$ denotes the pooled regression with year fixed effect being controlled and is clustered at the year level with bias-correction. All the control variables are in log. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 17

Centrality measures and monthly returns among the top 20% most-connected addresses.

	Monthly returns							
	$Pool_f$	$Pool_{fc}$	2015	2016	2017	2018	2019	2020
Centrality measures								
Eigenvector	0.365***	0.365**	0.077***	0.195***	0.690***	−0.046***	0.172***	0.260***
	(0.011)	(0.170)	(0.027)	(0.014)	(0.034)	(0.013)	(0.015)	(0.027)
Degree	−0.035**	−0.035	0.083***	0.009	−0.022	−0.064***	−0.058***	0.169***
	(0.014)	(0.114)	(0.016)	(0.017)	(0.043)	(0.013)	(0.019)	(0.030)
Other control variables								
Trade volume	−0.071***	−0.071**	−0.025***	−0.035***	−0.169***	0.035***	−0.033***	−0.109***
	(0.001)	(0.029)	(0.002)	(0.001)	(0.004)	(0.002)	(0.002)	(0.003)
Number of trades	0.051***	0.051	−0.012	0.097***	0.077***	−0.038***	0.031***	0.074***
	(0.003)	(0.043)	(0.009)	(0.004)	(0.015)	(0.009)	(0.006)	(0.007)
Market price volatility	0.353***	0.353	−0.183***	0.140***	−0.122***	−0.771***	−1.615***	2.248***
	(0.004)	(0.490)	(0.009)	(0.003)	(0.011)	(0.006)	(0.005)	(0.010)
Market trade volume	0.431***	0.431	0.528***	0.470***	3.210***	1.565***	3.405***	−12.938***
	(0.008)	(1.085)	(0.039)	(0.006)	(0.051)	(0.023)	(0.016)	(0.059)
Market number of trades	−0.471***	−0.471	0.664***	4.267***	0.740***	−7.335***	11.329***	−4.293***
	(0.023)	(2.815)	(0.068)	(0.026)	(0.070)	(0.034)	(0.043)	(0.093)
Adjusted R^2	0.249	0.249	0.047	0.102	0.068	0.357	0.687	0.457
Observations	654,088	654,088	85,733	170,541	110,094	124,760	152,593	145,014

This table displays the results from the regression of monthly returns on addresses with different centrality among the top 20% most-connected addresses, controlling for other variables. Centrality is measured using either degree centrality or eigenvector centrality. $Pool_f$ denotes the pooled regression with year fixed effect being controlled. $Pool_{fc}$ denotes the pooled regression with year fixed effect being controlled and is clustered at the year level with bias-correction. All the control variables are in log. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

more connection or improving connection quality are both beneficial in most years. The only exception is 2018, during which the coefficient on degree centrality is negative. Another observation is that the coefficient on degree centrality is usually larger than that on eigenvector centrality. The results are similar when we change the time horizon of returns to weekly or change the threshold to $M = 50$ (see Online Appendix Table A.23 to Table A.25). All these results show that there is no clear evidence that one centrality measure matters more than the other one among the connected addresses.

In Section 5.2, we have shown that the top 20% most-connected addresses are earning significantly higher returns than their peers. In the next regression, we try to find out if both centrality measures also matter equally among these highly connected addresses. Same as above, both centrality measures have been rescaled to be between 0 and 1 to make sure coefficients are comparable, and all the control variables are also included in the following regressions. The summary statistics of these highly connected addresses can be found in Table 9.

The results from pooled regressions in Table 17 clearly show that eigenvector centrality is positively associated with higher returns among the top 20% most-connected addresses, but degree centrality is not. When we run the regression on a yearly basis, the picture is even more clear: except in 2018, high eigenvector centrality is always associated with high returns among the top 20% most-connected addresses. However, there is no clear pattern between return and degree centrality. Changing the time horizon of returns to weekly returns makes the results even stronger (see Online Appendix Table A.26). The coefficient on eigenvector centrality is positive and large, and the coefficient on degree centrality is either on the wrong side or statistically insignificant. For robustness checks, we also report the results when $M = 50$ (see Online Appendix Table A.27 and Table A.28), and the conclusion is mostly similar.

We have shown that, in the bitcoin network, both degree centrality and eigenvector centrality are important to the connected addresses. However, for the top 20% most-connected address in the network, eigenvector centrality seems to be more important, implying that the quality of connections may be more crucial than quantity for them.

6. Conclusion

This paper reaches three conclusions. First, compared with their unconnected peers, connected addresses do not always earn higher returns in the early years. However, our results also show that connected addresses earn higher returns than their unconnected peers in 2019 and 2020. Second, returns also differ among the connected addresses. We divide the connected addresses into ten decile groups based on their centrality, and the top 20% most-connected addresses always earn higher returns than other addresses. Third, we also find that eigenvector centrality is more related to returns than degree centrality among the top 20% most-connected addresses, implying the quality of connections may be more important than quantity to those highly connected addresses.

In this paper, our data parsing rules almost certainly mislabel some addresses that are controlled by one person as owned by different people. With the development of new data parsing technology, such as BlockSci (Kalodner et al., 2020), we may be able to cluster bitcoin addresses into the individual level with higher accuracy in the future. Also, in this paper, we do not model interactions among addresses. A possible future direction is to look at the interactions in order to understand why eigenvector matters more for those highly connected addresses.

CRediT authorship contribution statement

Kwok Ping Tsang: Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing. **Zichao Yang:** Conceptualization, Formal analysis, Software, Writing – original draft, Data curation, Visualization.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jempfin.2022.02.001.

References

Ahern, Kenneth R., 2017. Information networks: Evidence from illegal insider trading tips. J. Financ. Econ. 125 (1), 26–47.

Antonopoulos, Andreas M., 2017. Mastering Bitcoin: Programming the Open Blockchain. O'Reilly Media, Inc.

Barber, Brad M, Lee, Yi-Tsung, Liu, Yu-Jane, Odean, Terrance, 2009. Just how much do individual investors lose by trading? Rev. Financ. Stud. 22 (2), 609–632.

Cameron, A. Colin, Miller, Douglas L., 2015. A practitioner's guide to cluster-robust inference. J. Hum. Resour. 50 (2), 317–372.

Chen, Weili, Wu, Jun, Zheng, Zibin, Chen, Chuan, Zhou, Yuren, 2019. Market manipulation of bitcoin: evidence from mining the Mt. Gox transaction network. In: IEEE INFOCOM 2019-IEEE Conference on Computer Communications. IEEE, pp. 964–972.

Gandal, Neil, Hamrick, J.T., Moore, Tyler, Oberman, Tali, 2018. Price manipulation in the Bitcoin ecosystem. J. Monetary Econ. 95, 86–96.

Glaser, Florian, Zimmermann, Kai, Haferkorn, Martin, Weber, Moritz Christian, Siering, Michael, 2014. Bitcoin-asset or currency? revealing users' hidden intentions. In: Revealing Users' Hidden Intentions (April 15, 2014). ECIS.

Grossman, Sanford J., Stiglitz, Joseph E., 1980. On the impossibility of informationally efficient markets. Am. Econ. Rev. 70 (3), 393–408.

Hung, Jui-Cheng, Liu, Hung-Chun, Yang, J. Jimmy, 2021. Trading activity and price discovery in Bitcoin futures markets. J. Empir. Financ. 62, 107–120.

Jackson, Matthew O., 2010. Social and Economic Networks. Princeton University Press.

Kalodner, Harry, Möser, Malte, Lee, Kevin, Goldfeder, Steven, Plattner, Martin, Chator, Alishah, Narayanan, Arvind, 2020. Blocksci: Design and applications of a blockchain analysis platform. In: 29th {USENIX} Security Symposium ({USENIX} Security 20). pp. 2721–2738.

Kyle, Albert S., 1985. Continuous auctions and insider trading. Econometrica 1315–1335.

Li, Xindan, Geng, Ziyang, Subrahmanyam, Avanidhar, Yu, Honghai, 2017. Do wealthy investors have an informational advantage? Evidence based on account classifications of individual investors. J. Empir. Financ. 44, 1–18.

Liu, Yukun, Tsyvinski, Aleh, 2020. Risks and returns of Cryptocurrency. Rev. Financ. Stud. (ISSN: 0893-9454) URL http://dx.doi.org/10.1093/rfs/hhaa113.

Mahomed, Nadim, et al., 2018. Understanding Consumer Adoption of Cryptocurrencies (Ph.D. thesis). University of Pretoria.

Meiklejohn, Sarah, Pomarole, Marjori, Jordan, Grant, Levchenko, Kirill, McCoy, Damon, Voelker, Geoffrey M, Savage, Stefan, 2013. A fistful of bitcoins: characterizing payments among men with no names. In: Proceedings of the 2013 Conference on Internet Measurement Conference. pp. 127–140.

Nakamoto, Satoshi, 2008. Bitcoin: A peer-to-peer electronic cash system. URL https://bitcoin.org/bitcoin.pdf.

Ozsoylev, Han N, Walden, Johan, Yavuz, M Deniz, Bildik, Recep, 2014. Investor networks in the stock market. Rev. Financ. Stud. 27 (5), 1323–1366.

Petersen, Mitchell A., 2009. Estimating standard errors in finance panel data sets: Comparing approaches. Rev. Financ. Stud. 22 (1), 435–480.

Rossi, Alberto G, Blake, David, Timmermann, Allan, Tonks, Ian, Wermers, Russ, 2018. Network centrality and delegated investment performance. J. Financ. Econ. 128 (1), 183–206.

Vigna, Paul, Deng, Chao, 2017. Bitcoin losses worsen as Chinese exchange closes. Wall Street J. URL https://www.wsj.com/articles/bitcoin-losses-worsen-as-chinese-exchange-closes-1505404288.

Walden, Johan, 2019. Trading, profits, and volatility in a dynamic information network model. Rev. Econom. Stud. 86 (5), 2248–2283.

Xi, Dingli, O'Brien, Timothy Ian, Irannezhad, Elnaz, 2020. Investigating the investment behaviors in cryptocurrency. J. Altern. Invest.