Physica A 527 (2019) 121339

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Cluster analysis on the structure of the cryptocurrency market via Bitcoin–Ethereum filtering

Jung Yoon Song [a,b], Woojin Chang [a,c], Jae Wook Song [d,*]

[a] Department of Industrial Engineering, Seoul National University, Seoul 08826, Republic of Korea
[b] Institute for Industrial Systems Innovation, Seoul National University, Seoul 08826, Republic of Korea
[c] SNU Institute for Research in Finance and Economics, Seoul National University, Seoul 08826, Republic of Korea
[d] Department of Data Science, Sejong University, Seoul 05006, Republic of Korea

HIGHLIGHTS

- Structure of the cryptocurrency market is analyzed via Bitcoin–Ethereum filtering.
- Agglomerative hierarchical clustering and minimum spanning tree analyses are applied.
- Market leadership of the Bitcoin and Ethereum are established.
- Six homogeneous clusters composed of less traded cryptocurrencies are detected.
- Market structure is transformed after the announcements of various regulations.

ARTICLE INFO

Article history:
Received 13 September 2018
Received in revised form 13 February 2019
Available online 30 April 2019

Keywords:
Cryptocurrency
Market structure
Bitcoin–Ethereum filtering
Clustering
Minimum spanning tree

ABSTRACT

The purpose of this research is to analyze the structure of the cryptocurrency market based on the correlation-based agglomerative hierarchical clustering and minimum spanning tree. In order to detect a reasonable and distinct collective behavior among the market entities, we propose a filtering mechanism, called Bitcoin–Ethereum filtering, to exclude their linear influences to other cryptocurrencies. In this regard, we carefully examine the market structures for the cases of before and after filtering in terms of the Total, Pre-, and Post-regulation periods. Based on the results, we discover the leadership of the Bitcoin and Ethereum in the market, six homogeneous clusters composed of relatively less-traded cryptocurrencies, and transformation of the market structure after the announcement of regulations from various countries.

1. Introduction

In 2008, the Bitcoin, the first cryptocurrency that aims to decentralize the central banking system with an implemented blockchain, was developed and introduced by Satoshi Nakamoto [1]. Since then, the cryptocurrency market has received profound attention from investors, researchers, and policy-makers in many countries. After the massive success of Bitcoin in trading volumes, various alternative cryptocurrencies have developed to improve certain features of the Bitcoin. In this regard, the related studies in the initial stage have focused on the evolution and development of the Bitcoin to analyze the cryptocurrency market. For instance, Ciaian (2018) discovered the strong long-run price interdependence between the Bitcoin and its alternative cryptocurrencies (altcoin) [2]. In recent years, there have been approximately 1900 altcoins

* Corresponding author.
E-mail addresses: jaewook.song@sejong.ac.kr, songjw.kr@gmail.com (J.W. Song).

https://doi.org/10.1016/j.physa.2019.121339
0378-4371/

created which have shown a considerable volume of trading. Thus, research in analyzing the cryptocurrency market as a single system could produce more valuable insights.

In perspective of Econophysics, a large portion of previous literature has focused on examining the efficiency of the cryptocurrency market based on its multifractal behavior. Urquart et al. (2016) insisted that the Bitcoin is inefficient and similar to the emerging financial market [3]; F. Bariviera (2017) observed the existence of long memory before 2014 based on the Hurst exponent and the Detrended Fluctuation Analysis [4]; Takaishi (2018) and da Silva Filho et al. (2018) identified the existence of multifractality in the Bitcoin return series based on the multifractal detrended fluctuation analysis [5,6]; and Zhang et al. (2018) discovered the inefficiency in the Ripple, Ethereum, NEM, Stellar, Litecoin, Dash, Monero, and Verge [7]. In contrast, Kristoufek (2018) applied the long-range dependence, fractal dimension, and entropy and claimed that the Bitcoin is efficient during the cooling period immediately after the strong bull market [8]; and Tiwari et al. (2018) explored the efficiency in cryptocurrency market using the applied rolling estimates [9].

From the work of [10], the complex system also possesses a large portion of Econophysics by analyzing the structure of various financial markets. Especially, many correlation-based approaches have discovered the collective market behavior [11–15], clustering phenomenon [16–21], topology of financial markets based on the minimum spanning tree(MST) [18,22–26], and its application to the portfolio management [27–30]. In particular, the studies on the crisis-driven structural changes in the financial markets have presented high correlations during the crisis [31]. Furthermore, many pieces of literature have discovered the collective behavior and the change of the market structure during the crisis in terms of the less robust network [32–41]. In this respect, similar approaches can be applied to the cryptocurrency market. For example, Stosic et al. (2018) revealed the structure and collective behavior of the cryptocurrency market from Aug-2016 to Jan-2018 based on the cross-correlation minimum spanning tree [42]. However, the number of similar attempts is fairly limited.

Our paper aims to analyze the structure of the cryptocurrency market and the collective behavior of its entities based on the correlation-based clustering and MST. After 2018, the cryptocurrency market has experienced a massive bull market. Such a phenomenon yields the high correlations among the cryptocurrencies, which incurs difficulties in constructing a meaningful correlation-based market structure. Notably, most of the cryptocurrencies exhibit extremely high correlations with the Bitcoin and Ethereum. Therefore, in this study, we attempt to utilize Bitcoin–Ethereum filtering to eliminate their linear influences to other cryptocurrencies. In this circumstance, we analyze the market structure for the before and after filtering cases and also for the Pre- and Post-regulation periods.

This paper is organized as follows: Section 2 explains the data and mathematical backgrounds for the structural break, filtering mechanism, clustering, and minimum spanning tree; Section 3 presents the experimental results and discusses the implications in terms of the collective behaviors and market structure; Section 4 concludes.

2. Data and methods

2.1. Data and structural break

The data used in this research, obtained from the API of the Binance Exchange (http://www.binance.com/), are hourly prices of 76 cryptocurrencies with 24/7 trading time. The list of cryptocurrencies is described in Table 1. The duration of the experiment is from Dec-2017 to Mar-2018, which yields 2860 observations for each cryptocurrency. Note that we select all the cryptocurrencies whose records exist for the duration of the experiment except the Tether(Dollar-backed) and DigixDAO(Gold-backed). Besides, the Binance Exchange experienced technical difficulty, which causes the missing data from 2018-02-08 9 AM to 2018-02-09 6 PM and the mismatched data by 28 min-based hourly prices from 2018-02-09 6 PM to 2018-02-11 12 AM. However, we decide to utilize the data as provided since the time series of entire cryptocurrencies are aligned in the same time intervals.

Meanwhile, many cryptocurrency exchange platforms yield different prices at the same time. Therefore, we compute the correlations between the prices of the Binance Exchange and those of other 33 exchange platforms to verify whether the data from the Binance Exchange can be a representative. Note that the prices of 33 exchange platforms are extracted from the Cryptocompare (http://www.cryptocompare.com/). The results of the correlations among prices and returns are summarized in Table 2.

The exchange platforms are listed based on the trading volumes of the cryptocurrencies in Table 2 in descending order. The result shows that the numbers of cryptocurrencies listed in Table 2 traded in other 33 platforms during the experiment are remarkably lower than that of Binance Exchange. For the existing cryptocurrencies in each platform, we convert the prices of BTC/USDT cryptocurrencies to USDT and BTC/USD to USD to compute the correlations. The results show the mean price and return correlations of 0.9782 and 0.8859, respectively, with a generally low standard deviation of correlations for the platforms with 10% or more trading volume coverage. In contrast, other platforms with low trading volumes show relatively low correlations with high standard deviations. Since the Binance Exchange exhibits the highest trading volume and the most significant number of cryptocurrencies during the experiment, we decide to consider its price data as a representative.

This research focuses on analyzing the structure of the cryptocurrency market using the correlation-based clustering and MST. In order to construct the correlation matrix, we utilize the stationary return time series of cryptocurrencies in Table 1. Let $r_{i,t}$ be the logarithmic return of cryptocurrency i at time t, then

$$r_{i,t} = \ln(P_{i,t}) - \ln(P_{i,t-1}) \tag{1}$$

Table 1
List of symbols and names of cryptocurrencies.

Cryptocurrency List (76)			
Symbol	Full name	Symbol	Full name
BTC	Bitcoin	ETH	Ethereum
ADA	Cardano	ADX	AdEx
AMB	Ambrosus	ARK	Ark
ARN	Aeron	AST	AirSwap
BAT	Basic Attenti...	BCC	Bitcoin Cash
BCD	Bitcoin Diamond	BCPT	BlockMason Cr...
BNT	Bancor	BQX	BQX
BTG	Bitcoin Gold	BTS	BitShares
CDT	Blox	DASH	Dash
DLT	Agrello	DNT	district0x
ENG	Enigma	ENJ	Enjin Coin
EOS	EOS	ETC	Ethereum Classic
EVX	Everex	FUEL	Etherparty
FUN	FunFair	GAS	Gas
GVT	Genesis Vision	GXS	GXChain
HSR	Hshare	ICN	Iconomi
IOTA	IOTA	KMD	Komodo
KNC	Kyber Network	LINK	ChainLink
LRC	Loopring	LSK	Lisk
LTC	Litecoin	MANA	Decentraland
MCO	Monaco	MDA	Moeda Loyalty...
MOD	Modum	MTH	Monetha
MTL	Metal	NEO	NEO
NULS	Nuls	OAX	OAX
OMG	OmiseGO	POE	Po.et
POWR	Power Ledger	PPT	Populous
QSP	Quantstamp	QTUM	Qtum
RCN	Ripio Credit ...	RDN	Raiden Networ...
REQ	Request Network	SALT	SALT
SNGLS	SingularDTV	SNM	SONM
SNT	Status	STORJ	Storj
STRAT	Stratis	SUB	Substratum
TNT	Tierion	TRX	TRON
VEN	VeChain	VIB	Viberate
WTC	Waltonchain	XMR	Monero
XRP	Ripple	XVG	Verge
XZC	ZCoin	YOYO	YOYOW
ZEC	Zcash	ZRX	0x

where $P_{i,t}$ represents the price of cryptocurrency i at time t. The descriptive statistics of $r_{i,t}$ are summarized in Table 3.

The results in Table 3 shows the almost zero mean and small standard deviation for each return series. Given that the risk-free rate is set to be zero, the Sharpe ratio [43], a risk-adjusted return, is the mean return divided by its standard deviation. In the case of two cryptocurrencies with the same mean returns, a cryptocurrency with a higher Sharpe ratio can be considered as a superior investment asset. Note that the Sharpe ratio is further investigated in Section 3.2. Distribution is approximately symmetric if the skewness is within $\pm 1/2$. In this sense, the return distributions of 48 cryptocurrencies are relatively symmetric, whereas those of 28 cryptocurrencies are right-skewed. Especially, ARN, AST, BCD, NULS, and LSK demonstrate extremely right-skewed distributions. Note that cryptocurrencies with extremely right-skewed distributions possess relatively small market capitalization. Also, the kurtosis suggests that the distributions of all cryptocurrency exhibit fat-tails in comparison to that of the Gaussian distributions whose kurtosis is equal to three. Lastly, the ADF [44] and PP [45] tests suggest that all the return series are stationary based on the strong statistical rejections on unit roots.

The duration of this research includes strong bullish and bearish markets divided by regulations on the cryptocurrency trade. Specifically, regulation and taxation plans carrying strong negative opinions were announced from the Chinese government on 2018-01-10. Thereafter, similar announcements are followed by Korea, France, and Germany. Since then, the direction of the market is changed from a strong bull market to constant bear market. In this context, we assume that the structure of the market is changed due to the regulation. Thus, we select a point of structural break and analyze the cryptocurrency market in terms of Total, Pre-regulation, and Post-regulation periods. At first, we create a simple index of the market with the mean of hourly return series of 76 cryptocurrencies. Based on the simple index, we apply the Chow test statistic [46] to identify the point of structural break. Let CS denotes the Chow statistic for the set of two-time series groups, then

$$CS = \frac{(S_C - (S_1 + S_2))/k}{(S_1 + S_2)/(N_1 + N_2 - 2k)} \sim F_{k,N1+N2-2k} \tag{2}$$

Table 2
Correlations of prices and returns among different cryptocurrency exchange platforms.

Exchange name	Num	Price mean	Price std	Return mean	Return std	Volume	Type
Binance	76	1	0	1	0	100%	USDT
BitTrex	37	0.9778	0.0702	0.8817	0.0935	48%	USDT
Bitfinex	15	0.9993	0.0005	0.9346	0.0237	41%	USD
Coinbase	3	0.9960	0.0017	0.9168	0.0051	33%	USD
Poloniex	15	0.9973	0.0061	0.9227	0.0325	26%	USDT
Kraken	12	0.8810	0.3803	0.8099	0.1010	20%	USD
Bitstamp	3	0.9987	0.0006	0.9280	0.0020	20%	USD
bitFlyer	2	0.9972	0.0007	0.8076	0.0153	12%	USD
Quoine	2	0.9927	0.0008	0.6045	0.2148	6%	USD
Gemini	2	0.9979	0.0005	0.9125	0.0070	5%	USD
Gateio	23	0.9799	0.0766	0.6401	0.2143	3%	USDT
Cexio	3	0.8988	0.1315	0.8266	0.0884	2%	USD
Yobit	21	0.9494	0.0768	0.4318	0.2668	2%	USD
Exmo	9	0.9975	0.0012	0.7919	0.0218	2%	USDT
Cryptopia	22	0.9729	0.1013	0.6712	0.1638	2%	USDT
LakeBTC	3	0.7242	0.2000	0.4062	0.2874	2%	USD
itBit	1	0.9976	0.0000	0.9222	0.0000	2%	USD
EXX	9	0.8529	0.1208	0.3981	0.1340	2%	USDT
Tidex	21	0.6603	0.3744	0.1751	0.1812	1%	USDT
BitBay	6	0.9757	0.0189	0.3583	0.0834	1%	USD
QuadrigaCX	2	0.9948	0.0008	0.5072	0.0163	1%	USD
LiveCoin	27	0.9478	0.0290	0.2501	0.1275	1%	USD
Coinroom	1	0.9784	0.0000	0.1784	0.0000	0%	USD
LocalBitcoins	1	0.6444	0.0000	0.0049	0.0000	0%	USD
Coinfloor	1	0.2187	0.0000	0.0270	0.0000	0%	USD
Remitano	1	0.8652	0.0000	0.0010	0.0000	0%	USD
Gatecoin	13	0.6649	0.2644	0.0999	0.0873	0%	USD
TheRockTrading	6	0.6858	0.1687	0.3562	0.0817	0%	USD
WavesDEX	6	0.8483	0.1427	0.1198	0.0385	0%	USD
Abucoins	11	0.9885	0.0070	0.4948	0.1519	0%	USD
Coincap	1	0.7870	0.0000	0.0540	0.0000	0%	USD
Lykke	2	0.9404	0.0101	0.3126	0.1164	0%	USD
BTCChina	2	0.8378	0.1515	0.2545	0.0193	0%	USD
BitSquare	9	0.6175	0.2917	0.0648	0.0208	0%	USD

Note: Num, std, USD, and USDT refer to the number of cryptocurrencies listed in Table 2 traded in each platform during the experiment, the standard deviation of correlations, US Dollar, and Tether, respectively.

Fig. 1. Evolutions of simple cryptocurrency index and structural breakpoint.

where S_C, S_1, and S_2 represent the sum of squared residuals from the combined data, the first group, and the second group, respectively. $F_{k,N1+N2-2k}$ is the F-distribution with k and $N_1 + N_2 - 2k$ as its degrees of freedoms. Also, k, $N1$, and $N2$ are the total number of parameters, and the number of observations in the first and second group, respectively. Computing the value of CS for the entire combinations of a simple index, we search for the highest value of CS as the point of the structural break based on F-test. The result of the breakpoint is illustrated in Fig. 1.

In Fig. 1, the evolutions of simple index and the Chow statistic are plotted as blue and gray solid lines, respectively. Then, the structural breakpoint based on the Chow statistics, regulations-related announcements in Korea, Europe, China, and the United States are plotted as the red, navy, black, green, and brown dashed vertical lines, respectively. Note that the headlines of regulations related announcements are summarized in Table 4. Interestingly, the simple index exhibits a bull(bear) market before(after) the breakpoint plotted as the red vertical dashed line. Also, the breakpoint is located in

Table 3
Descriptive statistics for return series.

Symbol	Mean	Std	S/R	Skewness	Kurtosis	ADF	PP
BTC	−1.11E−04	0.018	−0.006	−0.250	7.273	−10.488***	−58.284***
ETH	−1.97E−05	0.020	−0.001	−0.106	9.442	−10.550***	−58.390***
ADA	8.18E−05	0.031	0.003	0.622	9.623	−9.185***	−56.606***
ADX	−1.87E−04	0.031	−0.006	−0.192	9.937	−11.908***	−59.959***
AMB	9.80E−05	0.037	0.003	0.568	9.876	−12.430***	−57.690***
ARK	−1.37E−04	0.029	−0.005	0.343	11.784	−10.661***	−59.955***
ARN	3.12E−04	0.045	0.007	2.817	36.830	−9.776***	−50.695***
AST	9.75E−06	0.039	0.000	1.113	18.751	−10.715***	−59.942***
BAT	7.28E−05	0.030	0.002	0.204	10.518	−17.439***	−58.757***
BCC	−2.13E−04	0.024	−0.009	0.777	21.879	−10.426***	−58.341***
BCD	−3.24E−04	0.043	−0.008	2.067	27.229	−10.869***	−56.632***
BCPT	1.13E−04	0.042	0.003	0.543	9.556	−10.320***	−57.543***
BNT	−2.12E−06	0.022	0.000	−0.025	8.378	−10.318***	−60.321***
BQX	1.12E−04	0.033	0.003	0.192	11.737	−10.658***	−59.536***
BTG	−6.34E−04	0.027	−0.024	0.633	12.793	−10.961***	−61.811***
BTS	−7.22E−06	0.030	0.000	0.319	8.701	−9.749***	−57.788***
CDT	−3.33E−05	0.037	−0.001	0.359	8.777	−10.323***	−59.469***
DASH	−3.05E−04	0.021	−0.014	0.090	10.632	−10.945***	−57.287***
DLT	−1.22E−04	0.041	−0.003	0.870	13.949	−45.770***	−60.524***
DNT	1.75E−04	0.037	0.005	0.953	14.053	−10.540***	−57.032***
ENG	2.76E−04	0.036	0.008	0.411	8.043	−10.115***	−54.998***
ENJ	5.42E−04	0.037	0.015	0.207	9.381	−9.987***	−57.199***
EOS	2.69E−04	0.027	0.010	0.088	8.082	−9.957***	−56.623***
ETC	−1.86E−04	0.026	−0.007	0.288	13.867	−10.541***	−61.087***
EVX	−7.26E−05	0.035	−0.002	0.648	12.181	−9.796***	−61.111***
FUEL	1.89E−04	0.039	0.005	0.440	8.309	−11.194***	−61.135***
FUN	8.90E−05	0.033	0.003	0.100	7.556	−11.129***	−55.931***
GAS	−2.38E−05	0.032	−0.001	0.278	12.738	−12.350***	−59.784***
GVT	4.50E−04	0.035	0.013	0.455	10.016	−10.242***	−57.846***
GXS	−3.52E−06	0.028	0.000	0.435	9.533	−43.030***	−59.987***
HSR	−3.37E−04	0.028	−0.012	0.028	12.955	−10.518***	−59.707***
ICN	−2.04E−04	0.030	−0.007	0.205	9.872	−33.703***	−58.909***
IOTA	−4.67E−05	0.029	−0.002	0.070	10.026	−10.545***	−58.584***
KMD	−4.45E−06	0.030	0.000	0.640	27.052	−9.522***	−59.218***
KNC	−1.47E−06	0.030	0.000	0.114	7.756	−10.719***	−57.915***
LINK	1.92E−04	0.035	0.005	0.175	8.483	−20.264***	−59.628***
LRC	2.27E−04	0.035	0.006	0.867	11.832	−10.096***	−58.533***
LSK	1.58E−05	0.031	0.001	1.821	32.712	−10.301***	−59.399***
LTC	1.25E−04	0.023	0.006	0.431	10.094	−10.384***	−57.786***
MANA	5.64E−04	0.036	0.016	0.555	11.258	−9.811***	−54.922***
MCO	−6.55E−05	0.030	−0.002	−0.048	9.788	−22.889***	−56.318***
MDA	−2.21E−04	0.035	−0.006	0.216	22.184	−21.891***	−62.917***
MOD	−3.63E−05	0.035	−0.001	0.862	12.937	−42.443***	−58.367***
MTH	1.91E−04	0.039	0.005	0.123	8.213	−11.478***	−56.767***
MTL	−9.25E−05	0.031	−0.003	0.081	8.035	−10.783***	−58.087***
NEO	1.47E−04	0.027	0.005	0.314	11.110	−10.907***	−58.959***
NULS	5.56E−04	0.039	0.014	1.197	17.288	−34.379***	−59.881***
OAX	1.11E−04	0.037	0.003	0.924	15.374	−10.993***	−59.399***
OMG	1.87E−05	0.027	0.001	0.231	9.648	−11.135***	−57.561***
POE	4.21E−04	0.042	0.010	0.917	12.229	−9.369***	−55.439***
POWR	−2.86E−04	0.031	−0.009	0.298	9.581	−43.705***	−58.047***
PPT	1.13E−04	0.032	0.003	0.115	10.241	−10.615***	−61.211***
QSP	−2.95E−05	0.034	−0.001	0.206	8.063	−23.466***	−55.487***
QTUM	7.09E−05	0.028	0.002	0.704	11.601	−9.912***	−57.404***
RCN	−1.05E−04	0.032	−0.003	0.098	7.488	−17.768***	−57.297***
RDN	−3.85E−04	0.029	−0.013	−0.207	7.809	−42.515***	−58.345***
REQ	3.58E−04	0.035	0.010	0.679	14.771	−10.242***	−58.385***
SALT	−2.38E−04	0.030	−0.008	0.694	15.192	−11.342***	−60.332***
SNGLS	−2.88E−04	0.037	−0.008	0.815	10.264	−10.865***	−58.535***
SNM	−1.81E−05	0.034	−0.001	0.046	9.063	−10.542***	−58.354***
SNT	1.96E−04	0.031	0.006	0.836	10.457	−41.068***	−57.582***
STORJ	1.37E−05	0.030	0.000	0.408	12.068	−9.840***	−56.474***
STRAT	−1.14E−04	0.029	−0.004	−0.061	7.800	−11.201***	−58.425***
SUB	4.29E−05	0.035	0.001	0.303	9.169	−10.829***	−58.455***
TNT	2.33E−05	0.037	0.001	0.464	9.810	−18.843***	−60.893***
TRX	9.88E−04	0.043	0.023	0.369	13.854	−10.548***	−57.695***

(*continued on next page*)

Table 3 (*continued*).

Symbol	Mean	Std	S/R	Skewness	Kurtosis	ADF	PP
VEN	8.03E−04	0.034	0.024	0.264	7.978	−41.739***	−58.665***
VIB	2.78E−05	0.038	0.001	0.890	13.721	−10.952***	−61.520***
WTC	1.83E−04	0.034	0.005	0.339	8.577	−10.975***	−57.280***
XMR	2.10E−05	0.024	0.001	−0.175	8.891	−11.224***	−57.869***
XRP	2.75E−04	0.030	0.009	0.994	16.286	−9.562***	−60.835***
XVG	6.89E−04	0.043	0.016	0.822	16.278	−10.453***	−58.902***
XZC	−6.24E−05	0.031	−0.002	0.917	43.489	−17.190***	−63.754***
YOYO	−1.11E−04	0.037	−0.003	0.693	12.945	−34.764***	−60.542***
ZEC	−1.50E−04	0.023	−0.007	0.015	8.159	−10.880***	−57.079***
ZRX	3.76E−04	0.034	0.011	0.890	15.535	−12.642***	−58.887***

Note: Std, S/R, ADF, and PP are the abbreviations of the standard deviation, Sharpe ratio, Augmented Dickey Fuller, and Phillips–Perron tests, respectively. Also, the notation *, **, *** indicate 5%, 1%, 0.1% statistical significances.

Table 4
Headlines of regulation-related announcements in various countries.

Date	Country	Headlines	Publisher
2017-12-13	Korea	S Korea considers capital gains tax on cryptocurrencies	Reuters
2017-12-18	Europe	Germany Joins French-led Moves to Regulate Bitcoin at G-20 Level	Bloomberg
2017-12-30	Korea	Bitcoin slips as Korea threatens to shut exchanges	FT
2018-01-10	China	China moves to shutter bitcoin mines	FT
2018-01-11	Korea	Bitcoin dives 14% as South Korea prepares cryptocurrency ban	FT
2018-01-12	Europe	Ukraine to create group to regulate cryptocurrency market	FT
2018-01-15	Europe	Ukraine steps up effort to regulate cryptocurrencies	FT
2018-01-16	China	PBOC official says China's centralized virtual currency trade needs to end: source	Reuters
2018-01-19	Europe	France, Germany to make joint bitcoin regulation proposal at G20 summit	Reuters
2018-01-19	US	Regulator Pours Cold Water on Bitcoin ETFs, US Government Shutdown Looms	WSJ
2018-01-27	Europe	G20 to discuss global regulatory framework for cryptocurrency: ECB's Cœuré	FT
2018-02-06	US	U.S. regulators to back more oversight of virtual currencies	Reuters
2018-02-09	Europe	France and Germany join calls for global bitcoin clampdown	FT
2018-02-13	Europe	UK crypto companies link up for self-regulation	FT
2018-03-03	Europe	BoE calls time on lax regulation of cryptocurrencies	FT
2018-03-09	Europe	EU body strikes back at cryptocurrency regulation	FT

the middle of the pouring of regulation-related announcements from various countries. Therefore, we decide to divide the Pre- and Post-regulation periods based on the breakpoint.

2.2. Bitcoin-Ethereum filtering

In this research, we utilize the filtered return series instead of the plain return series to construct the correlation matrix among the cryptocurrencies. Note that the extremely high correlations among cryptocurrencies restrict the relevant analysis of market structure. Hence, we apply the Ordinary Least Squares(OLS) method to filter the linear influences from the Bitcoin and Ethereum as follows.

$$r_{i,t} = \alpha_i + \beta_i r_{bitcoin,t} + \gamma_i r_{ethereum,t} + \varepsilon_{i,t} \tag{3}$$

where $r_{i,t}$, $r_{bitcoin,t}$, $r_{ethereum,t}$ represent the return series of cryptocurrency i, Bitcoin, and Ethereum at time t, respectively; α_i, β_i and γ_i are the intercept and coefficients of linear relationship from the Bitcoin and Ethereum filters, respectively; and $\varepsilon_{i,t}$ is the residuals of OLS filtering.

Since the utilization of combined filtering with both α_i and β_i is applicable if there is no multicolinearity, we employ the Variance Inflation Factor (*VIF*) to test its existence. At first, a simple regression is performed with the return series of Ethereum($r_{ethereum}$) to predict the return series of Bitcoin($\hat{r}_{bitcoin}$) as follows.

$$\hat{r}_{bitcoin,t} = \alpha_{bitcoin} + \beta_{bitcoin} r_{ethereum,t} + \varepsilon_{bitcoin,t}. \tag{4}$$

When $\hat{r}_{bitcoin}$ is predicted variable, the *VIF* can be measured as follows.

$$VIF = \frac{SST}{SSE} = \frac{\sum (r_{bitcoin,t} - \bar{r}_{bitcoin,t})^2}{\sum (r_{bitcoin,t} - \hat{r}_{bitcoin,t})^2} \tag{5}$$

where $\bar{r}_{bitcoin,t}$, *SST*, *SSE* indicate the mean of $r_{bitcoin,t}$, sum of squares for treatments and errors, respectively. If $VIF > 5$, the simultaneous estimation of the Bitcoin and Ethereum coefficients is not appropriate. In this case, we suggest to use the stepwise OLS method to filter the linear relationship as follows.

$$r_{i,t} = \alpha_{1i} + \beta_i r_{bitcoin,t} + \epsilon_{i,t} \tag{6}$$

$$\epsilon_{i,t} = \alpha_{2i} + \gamma_i r_{ethereum,t} + \xi_{i,t} \tag{7}$$

where α_{1i}, α_{2i}, β_i, γ_i, $r_{bitcoin,t}$, $r_{ethereum,t}$, $\epsilon_{i,t}$ and $\xi_{i,t}$ are the intercepts, coefficients of linear relationship, returns of Bitcoin and Ethereum, and each residual from the Bitcoin and Ethereum filters of cryptocurrency i at time t, respectively. In conclusion, the filtered return series can be obtained in two different conditions. Let R_i denotes the vector of filtered return series of cryptocurrency i, then

$$R_i = \begin{cases} \varepsilon_i, & \text{when} \quad VIF \leq 5 \\ \xi_i, & \text{when} \quad VIF > 5 \end{cases} \tag{8}$$

where ε_i and ξ_i represent the set of residuals of cryptocurrency i in Eqs. (3) and (7), respectively. Finally, we call this procedure as Bitcoin–Ethereum filtering.

2.3. Clustering and minimum spanning tree

The structure of cryptocurrency market can be analyzed with the correlation-based clustering and MST, which requires the distance among the cryptocurrencies. In this research, we compute the distance by considering the coefficient of the Pearson correlations as the similarity measure. Let R_i and R_j denote the vector of the return series of cryptocurrency i and j for N numbers of cryptocurrencies, respectively, then the correlation coefficient between cryptocurrency i and j is,

$$Correlation(R_i, R_j) = \rho_{ij} = \frac{\langle R_i \cdot R_j \rangle - \langle R_i \rangle \langle R_j \rangle}{\sqrt{\left(\langle R_i^2 \rangle - \langle R_i \rangle^2\right)\left(\langle R_j^2 \rangle - \langle R_j \rangle^2\right)}} \tag{9}$$

where $R_i \cdot R_j$ and $\langle R_i \rangle$ are the inner product of two vectors and the mean of R_i, respectively. Note that ρ_{ij} ranges from -1 to 1 and its calculation for all combinations of i and j yields $N \times N$ symmetric matrix where $\rho_{ij} = \rho_{ji}$ and $\rho_{ii} = 1$.

In this context, the distance using the correlation-based similarity measure can be defined as,

$$Distance(R_i, R_j) = d_{ij} = \sqrt{2(1 - \rho_{ij})} \tag{10}$$

where $d_{ii} = 0$. Note that the distance based on ρ_{ij} is equivalent to the squared Euclidean distance between the z-score normalized time series [21].

At first, based on d_{ij}, we employ the agglomerative hierarchical clustering [17], a bottom-up approach that begins with N numbers of clusters until it becomes one group. Note that the clustering result can be summarized in forms of dendrogram. For the linkage, we use the Ward's linkage [17,47] method. Let C_i and C_j represent the clusters in a Ward's linkage, then

$$Distance(C_i, C_j) = D_{ij} = \sqrt{\frac{2n_i n_j}{n_i + n_j}} \parallel \bar{c}_i - \bar{c}_j \parallel \tag{11}$$

where n and $\bar{c}$ refer to the number of elements and centroid of a cluster, respectively.

Secondly, we utilize the correlation-based MST by applying the Kruskal algorithm [48] into the distance matrix in Eq. (10). We select three measures to analyze the MST. The first measure is the mean of correlation coefficients defined as,

$$\bar{\rho} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{ij} \tag{12}$$

The second measure is the normalized tree length, $L(p)$, which indicates the average distance of edges in MST such that,

$$L(p) = \frac{1}{N-1} \sum_{e_{ij} \in \Theta} d_{ij}^p \tag{13}$$

where e_{ij} and Θ are the edge between i and j and the set of edges.

The last measure is the mean occupation layer [36,49], $\ell(p, v_c)$, which indicates the mass of tree such that,

$$\ell(p, v_c) = \frac{1}{N} \sum_{i=1}^{N} \eta(v_i^p) \tag{14}$$

where v_c and $\eta(v_i^p)$ represent the central node and the level of vertex v_i in relation to v_c, respectively, whose level is taken to be zero. Note that the node with maximum edges can be defined as k_{max}^p. Based on various methods to define the central node of mean occupation layer [36,49], we obtain the central vertex in terms of the node with the highest degree(number of incident edges).

Table 5
Results of the VIF test.

Period	R^2	VIF
Total	0.6242	2.6613
Pre period	0.4078	1.6885
Post period	0.8637	7.3361

Table 6
Results of Bitcoin–Ethereum filtering in Total period.

Symbol	R^2	F_{pvalue}	Symbol	R^2	F_{pvalue}
ADA	0.4919	0***	MANA	0.3174	1.52E−237***
ADX	0.4995	0***	MCO	0.4662	0***
AMB	0.3661	1.81E−283***	MDA	0.3285	9.87E−248***
ARK	0.5126	0***	MOD	0.3826	7.94E−300***
ARN	0.2537	3.23E−182***	MTH	0.3375	4.18E−256***
AST	0.3449	4.87E−263***	MTL	0.4353	0***
BAT	0.4633	0***	NEO	0.6434	0***
BCC	0.4496	0***	NULS	0.3391	1.46E−257***
BCD	0.1863	1.35E−128***	OAX	0.3184	2.07E−238***
BCPT	0.2933	5.08E−216***	OMG	0.6218	0***
BNT	0.7179	0***	POE	0.3369	1.81E−255***
BQX	0.4411	0***	POWR	0.5255	0***
BTG	0.4839	0***	PPT	0.3262	1.37E−245***
BTS	0.5389	0***	QSP	0.4248	0***
CDT	0.3357	2.20E−254***	QTUM	0.5554	0***
DASH	0.6661	0***	RCN	0.4293	0***
DLT	0.2518	1.36E−180***	RDN	0.4821	0***
DNT	0.3370	1.34E−255***	REQ	0.4213	0***
ENG	0.3774	1.29E−294***	SALT	0.4787	0***
ENJ	0.3697	5.32E−287***	SNGLS	0.3086	1.47E−229***
EOS	0.6230	0***	SNM	0.3980	1.89E−315***
ETC	0.6232	0***	SNT	0.4767	0***
EVX	0.3470	5.15E−265***	STORJ	0.4838	0***
FUEL	0.3429	3.47E−261***	STRAT	0.5959	0***
FUN	0.4452	0***	SUB	0.4478	0***
GAS	0.4719	0***	TNT	0.3569	1.68E−274***
GVT	0.3800	3.56E−297***	TRX	0.3418	4.23E−260***
GXS	0.4267	0***	VEN	0.4626	0***
HSR	0.4869	0***	VIB	0.3033	8.17E−225***
ICN	0.3849	3.95E−302***	WTC	0.4618	0***
IOTA	0.5342	0***	XMR	0.6460	0***
KMD	0.4348	0***	XRP	0.4224	0***
KNC	0.4790	0***	XVG	0.3248	2.68E−244***
LINK	0.3990	1.62E−316***	XZC	0.4054	3.46E−323***
LRC	0.3840	3.71E−301***	YOYO	0.3205	2.18E−240***
LSK	0.4658	0***	ZEC	0.6186	0***
LTC	0.6616	0***	ZRX	0.4347	0***

Note: the notation *, **, *** indicate 5%, 1%, 0.1% statistical significances. F_{pvalue} lower than 1.00E−350 is considered as zero.

3. Results and discussion

3.1. Filtering results

In this study, we employ the OLS method to filter the linear influence from the Bitcoin and Ethereum to other cryptocurrencies. At first, the VIF test in Eq. (4) is performed to select the appropriate Bitcoin–Ethereum filtering in Eq. (8).

The results of VIF test are shown in Table 5. Since the values of VIF in the Total($\approx$ 2.66) and Pre-regulation($\approx$ 1.69) periods are smaller than 5, ε_i should be used based on Eq. (3). In contrast, the Post-regulation period should use the stepwise OLS filtering, ξ_i, in Eq. (7) since its VIF is approximately 7.34. The results of OLS filtering in Total, Pre-, and Post-regulation periods are summarized in Tables 6, 7, and 8, respectively.

Consequently, the same filtering is applied to the Total and Pre-regulation periods. Although the values of R^2 for the entire cryptocurrencies in both periods are not significantly high, the OLS filtering is statistically sound due to approximately zero p_{value}s in F tests. In the case of the Post-regulation period, the OLS filtering of Bitcoin is statistically significant for all cryptocurrencies based on their R^2 and p_{value}. However, the OLS filtering of Ethereum, which is performed after the filtering of Bitcoin, shows relatively low values of R^2 and some cases of insignificant p_{value} such as BCD, DNT, and MTL. It refers that a large part of the linear influence is pre-eliminated through the Bitcoin filtering. Therefore, the Ethereum filtering is not applied to BCD, DNT, and MTL.

Table 7
Results of Bitcoin–Ethereum filtering in Pre-regulation period.

Symbol	R^2	F_{pvalue}	Symbol	R^2	F_{pvalue}
ADA	0.2941	8.43E−81***	MANA	0.2160	1.07E−56***
ADX	0.3792	2.25E−110***	MCO	0.3242	8.04E−91***
AMB	0.2544	3.15E−68***	MDA	0.2370	6.41E−63***
ARK	0.3857	8.77E−113***	MOD	0.2560	9.82E−69***
ARN	0.1244	2.75E−31***	MTH	0.2443	3.87E−65***
AST	0.2564	7.40E−69***	MTL	0.3802	1.02E−110***
BAT	0.3570	2.76E−102***	NEO	0.4675	1.17E−145***
BCC	0.2442	4.13E−65***	NULS	0.2249	2.59E−59***
BCD	0.1071	9.17E−27***	OAX	0.2220	1.93E−58***
BCPT	0.1762	2.72E−45***	OMG	0.4727	6.44E−148***
BNT	0.6017	1.94E−212***	POE	0.2059	9.57E−54***
BQX	0.3050	2.21E−84***	POWR	0.3823	1.62E−111***
BTG	0.3090	1.04E−85***	PPT	0.2381	2.92E−63***
BTS	0.3844	2.83E−112***	QSP	0.2464	8.55E−66***
CDT	0.2391	1.48E−63***	QTUM	0.4097	6.24E−122***
DASH	0.5611	4.39E−190***	RCN	0.3562	5.22E−102***
DLT	0.1387	4.49E−35***	RDN	0.3754	5.71E−109***
DNT	0.2605	4.04E−70***	REQ	0.2767	3.33E−75***
ENG	0.2353	1.99E−62***	SALT	0.3529	8.19E−101***
ENJ	0.2209	3.95E−58***	SNGLS	0.2368	7.02E−63***
EOS	0.4685	4.69E−146***	SNM	0.2824	5.02E−77***
ETC	0.5201	1.51E−169***	SNT	0.3286	2.41E−92***
EVX	0.2437	5.74E−65***	STORJ	0.4036	1.45E−119***
FUEL	0.2398	9.19E−64***	STRAT	0.4519	5.10E−139***
FUN	0.3193	3.62E−89***	SUB	0.2775	1.85E−75***
GAS	0.3351	1.42E−94***	TNT	0.2991	1.90E−82***
GVT	0.2263	1.01E−59***	TRX	0.2069	4.82E−54***
GXS	0.2917	4.82E−80***	VEN	0.2698	4.78E−73***
HSR	0.3531	6.82E−101***	VIB	0.1843	1.48E−47***
ICN	0.2741	2.09E−74***	WTC	0.3125	6.64E−87***
IOTA	0.3862	5.89E−113***	XMR	0.5144	7.81E−167***
KMD	0.3042	3.86E−84***	XRP	0.2161	1.02E−56***
KNC	0.3885	7.82E−114***	XVG	0.1875	1.82E−48***
LINK	0.2575	3.44E−69***	XZC	0.2700	4.36E−73***
LRC	0.2770	2.63E−75***	YOYO	0.2329	1.06E−61***
LSK	0.3629	2.07E−104***	ZEC	0.5127	4.93E−166***
LTC	0.5257	3.11E−172***	ZRX	0.3538	3.93E−101***

Note: the notation *, **, *** indicate 5%, 1%, 0.1% statistical significances.

Fig. 2 visualizes the color-map of correlation matrices. Figs. 2a, 2c, and 2e refer to the correlation among the cryptocurrencies for each period before Bitcoin–Ethereum filtering, whereas Figs. 2b, 2d, and 2f are after the filtering. The color-map presents the correlation with gradation color scheme where the blue, white, and red colors indicate negative, zero, and positive correlations, respectively. The color-maps before filtering for all periods are mostly red, which indicates high correlations among the cryptocurrencies. In this milieu, the correlation-based analysis hardly produces a meaningful result. However, the color-map after filtering is mixed with white and red colors, which indicate the reduced correlations. Based on six correlation matrices, we proceed with further analysis in terms of clustering and MST.

3.2. Clustering analysis

As previously described, we analyze the cryptocurrency market using the agglomerative hierarchical clustering of correlation matrices in Fig. 2. The results are plotted in Fig. 3.

According to dendrograms before Bitcoin–Ethereum filtering, most of the cryptocurrencies are assembled in two clusters whose centers are Bitcoin and Ethereum in the Total (Fig. 3a) and Post-regulation period (Fig. 3e), respectively. Although the Bitcoin and Ethereum-based assembling phenomenon in the Pre-regulation period (Fig. 3c) is weak, a significantly smaller number of clusters is observed in the before filtering than in the after filtering. As a result of plain return series, we confirm that the major assets of the cryptocurrency market are the Bitcoin and Ethereum. In other words, the cryptocurrencies that follow the Bitcoin and Ethereum can be identified. However, the influences of both coins are so large that it is difficult to detect homogeneous clusters among minor cryptocurrencies. In contrast, the dendrograms after filtering exclude two strong center cryptocurrencies and form various clusters. Specifically, the Total period after filtering can form from one to 74 clusters through agglomerative hierarchical clustering. The results for nine to 14 clusters are shown in Fig. 4.

According to dendrograms, we discover that six major clusters continuously survive when the number of clusters increases from nine to 14. Note that these clusters belong to the Bitcoin and Ethereum clusters before filtering. Thus, by

Table 8
Results of Bitcoin–Ethereum filtering in Post-regulation period.

Symbol	R^2_{btc}	$F_{pvalue,btc}$	R^2_{eth}	$F_{pvalue,eth}$	Symbol	R^2_{btc}	$F_{pvalue,btc}$	R^2_{eth}	$F_{pvalue,eth}$
ADA	0.772	0***	0.019	3.17E−9***	MANA	0.544	2.62E−308***	0.004	6.92E−3**
ADX	0.654	0***	0.005	3.27E−3**	MCO	0.702	0***	0.003	1.29E−2*
AMB	0.490	8.26E−265***	0.005	3.03E−3**	MDA	0.484	4.85E−260***	0.002	4.66E−2*
ARK	0.665	0***	0.008	1.51E−4***	MOD	0.496	2.65E−269***	0.009	4.67E−5***
ARN	0.539	8.78E−304***	0.006	7.84E−4***	MTH	0.464	2.48E−245***	0.004	1.21E−2*
AST	0.467	1.98E−247***	0.004	1.03E−2*	MTL	0.496	3.47E−269***	0.001	1.47E−01
BAT	0.586	0***	0.007	4.86E−4***	NEO	0.749	0***	0.024	5.27E−11***
BCC	0.737	0***	0.004	5.72E−3**	NULS	0.467	1.68E−247***	0.009	6.99E−5***
BCD	0.278	3.35E−129***	0.001	1.15E−01	OAX	0.430	3.38E−221***	0.004	8.26E−3**
BCPT	0.424	6.00E−217***	0.006	1.05E−3**	OMG	0.765	0***	0.014	3.37E−7***
BNT	0.768	0***	0.046	2.86E−20***	POE	0.550	2.94E−313***	0.012	5.18E−6***
BQX	0.561	1.48E−323***	0.011	1.27E−5***	POWR	0.670	0***	0.014	5.16E−7***
BTG	0.681	0***	0.008	1.44E−4***	PPT	0.419	1.02E−213***	0.004	5.24E−3**
BTS	0.701	0***	0.011	6.12E−6***	QSP	0.610	0***	0.015	2.69E−7***
CDT	0.454	3.25E−238***	0.006	1.42E−3**	QTUM	0.706	0***	0.015	1.35E−7***
DASH	0.759	0***	0.007	3.04E−4***	RCN	0.528	3.56E−295***	0.002	4.17E−2*
DLT	0.398	8.06E−200***	0.004	4.84E−3**	RDN	0.586	0***	0.010	2.88E−5***
DNT	0.421	1.98E−215***	0.001	1.02E−01	REQ	0.560	1.04E−322***	0.016	7.05E−8***
ENG	0.557	1.97E−319***	0.009	6.38E−5***	SALT	0.621	0***	0.008	1.96E−4***
ENJ	0.577	0***	0.008	1.32E−4***	SNGLS	0.374	4.15E−185***	0.005	4.15E−3**
EOS	0.760	0***	0.023	9.32E−11***	SNM	0.562	4.94E−324***	0.007	4.56E−4***
ETC	0.698	0***	0.015	1.89E−7***	SNT	0.674	0***	0.013	1.12E−6***
EVX	0.504	7.97E−276***	0.007	2.39E−4***	STORJ	0.593	0***	0.003	1.55E−2*
FUEL	0.506	6.34E−277***	0.005	3.06E−3**	STRAT	0.748	0***	0.011	7.29E−6***
FUN	0.584	0***	0.009	7.40E−5***	SUB	0.602	0***	0.011	8.61E−6***
GAS	0.581	0***	0.007	4.17E−4***	TNT	0.397	4.08E−199***	0.005	3.67E−3**
GVT	0.562	4.94E−324***	0.004	4.72E−3**	TRX	0.615	0***	0.007	4.21E−4***
GXS	0.640	0***	0.007	5.08E−4***	VEN	0.678	0***	0.014	6.28E−7***
HSR	0.641	0***	0.007	5.82E−4***	VIB	0.475	2.15E−253***	0.003	1.73E−2*
ICN	0.501	6.47E−273***	0.003	1.30E−2*	WTC	0.590	0***	0.007	5.07E−4***
IOTA	0.758	0***	0.010	1.48E−5***	XMR	0.772	0***	0.008	1.66E−4***
KMD	0.634	0***	0.006	1.49E−3**	XRP	0.730	0***	0.027	2.96E−12***
KNC	0.562	4.94E−324***	0.007	4.78E−4***	XVG	0.646	0***	0.009	7.33E−5***
LINK	0.561	5.93E−323***	0.011	9.67E−6***	XZC	0.686	0***	0.005	2.03E−3**
LRC	0.524	1.57E−291***	0.009	5.83E−5***	YOYO	0.482	1.57E−258***	0.004	1.15E−2*
LSK	0.540	1.01E−304***	0.004	5.00E−3**	ZEC	0.746	0***	0.005	1.71E−3**
LTC	0.796	0***	0.012	3.22E−6***	ZRX	0.496	3.55E−269***	0.007	2.59E−4***

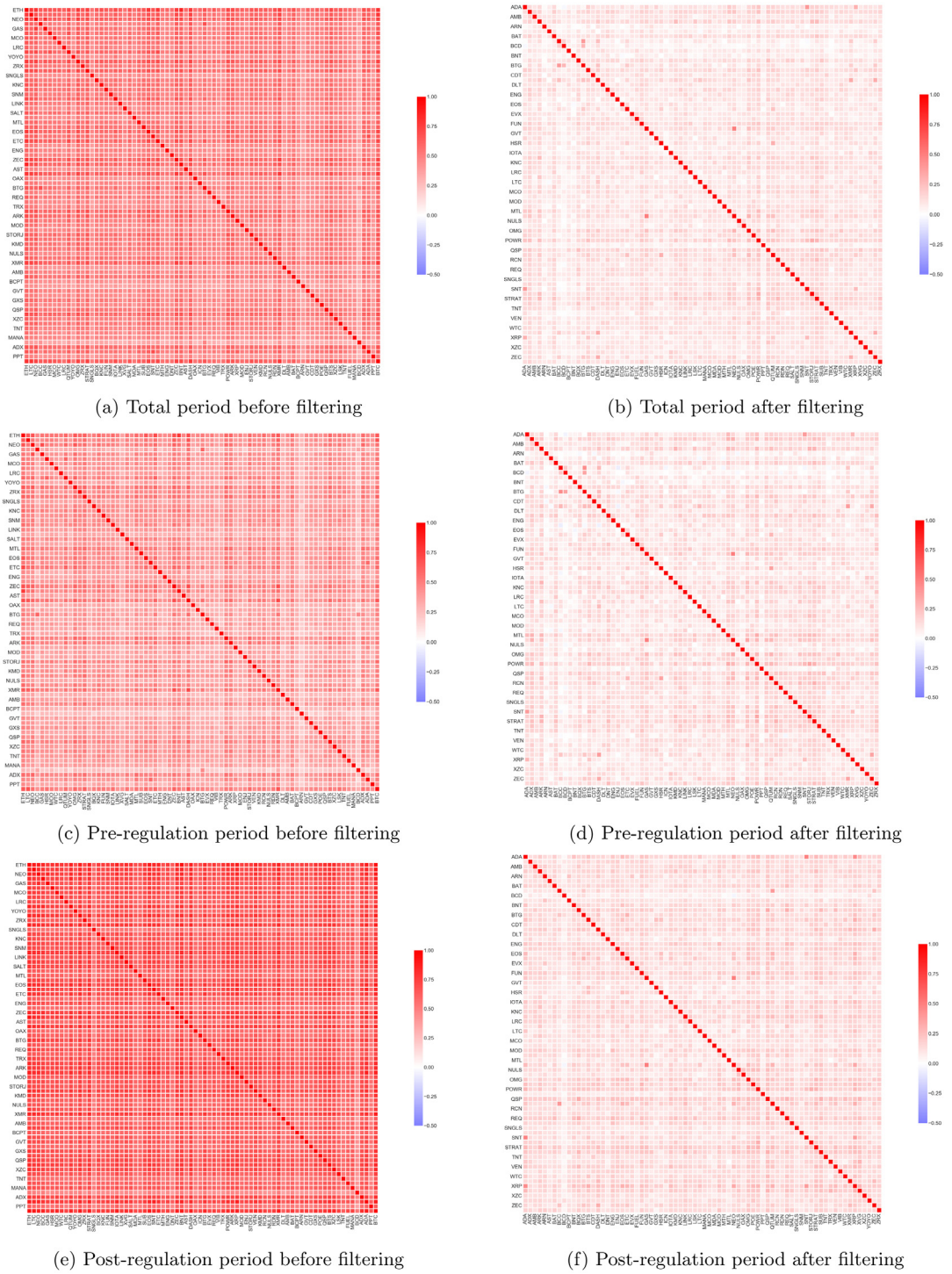
Note: the notation *, **, *** indicate 5%, 1%, 0.1% statistical significances. F_{pvalue} lower than 1.00E−350 is considered as zero.

Table 9
List of major clusters in Total period after Bitcoin–Ethereum filtering.

Cluster#1 (16)		Cluster#2 (2)		Cluster#3 (3)		Cluster#4 (3)		Cluster#5 (4)		Cluster#6 (3)	
ADX	AdEx	GAS	Gas	ADA	Cardano	BCC	Bitcoin Cash	DASH	Dash	ETC	Ethereum Classic
ARK	Ark	NEO	NEO	SNT	Status	BCD	Bitcoin Diamond	LTC	Litecoin	HSR	Hshare
BTS	BitShares			XRP	Ripple	BTG	Bitcoin Gold	XMR	Monero	QTUM	Qtum
KMD	Komodo							ZEC	Zcash		
LSK	Lisk										
MCO	Monaco										
MTL	Metal										
OMG	OmiseGO										
POWR	Power Ledger										
RCN	Ripio Credit ...										
SALT	SALT										
SNM	SONM										
STORJ	Storj										
STRAT	Stratis										
VIB	Viberate										
XZC	ZCoin										

filtering, the influences of Bitcoin and Ethereum are successfully removed, and the homogeneous clusters among minor clusters are obtained. The list of major clusters is summarized in Table 9.

Each homogeneity in six clusters can be interpreted as follows. Cluster#1 in green color includes the token-based cryptocurrencies with the low Sharpe ratio in Table 3. Cluster#2 in brown color refers to the Neo-cryptocurrencies that provide gas to owners as an interest rate income. Cluster#3 in gold color includes the cryptocurrencies with



(a) Total period before filtering

(b) Total period after filtering

(c) Pre-regulation period before filtering

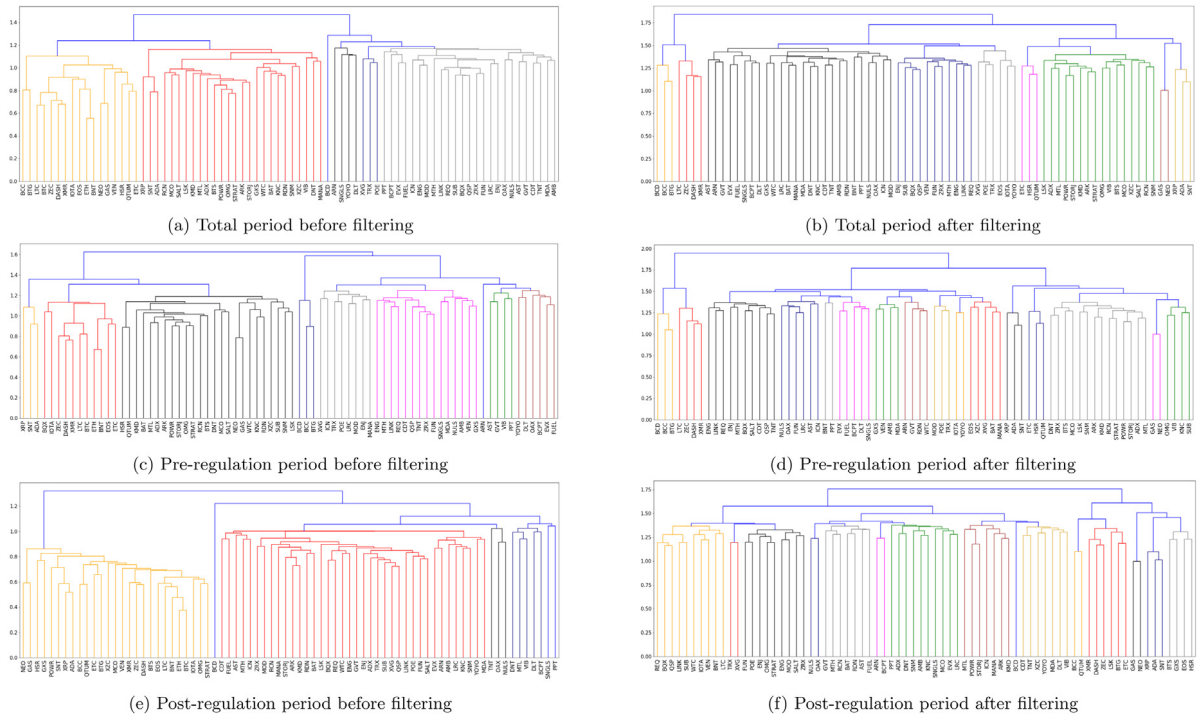
(d) Pre-regulation period after filtering

(e) Post-regulation period before filtering

(f) Post-regulation period after filtering

Fig. 2. Heatmaps of correlation matrices.

the most substantial trading volumes in Korea than in other countries during early 2018 with excessive speculations. Cluster#4 in yellow color indicates the Bitcoin-driven hard folk coins where the BCC, BTG, and BCD are distributed to the owners of #478,558, #491,407, and, #495,866 blocks, respectively. Cluster#5 in red color mostly includes so called dark-cryptocurrencies which allow the anonymous transaction among traders. Cluster#6 in pink color includes the ETC, HSR, and QTUM where the ETC and QTUM are the Ethereum-related cryptocurrencies, and the HSR and QTUM are the

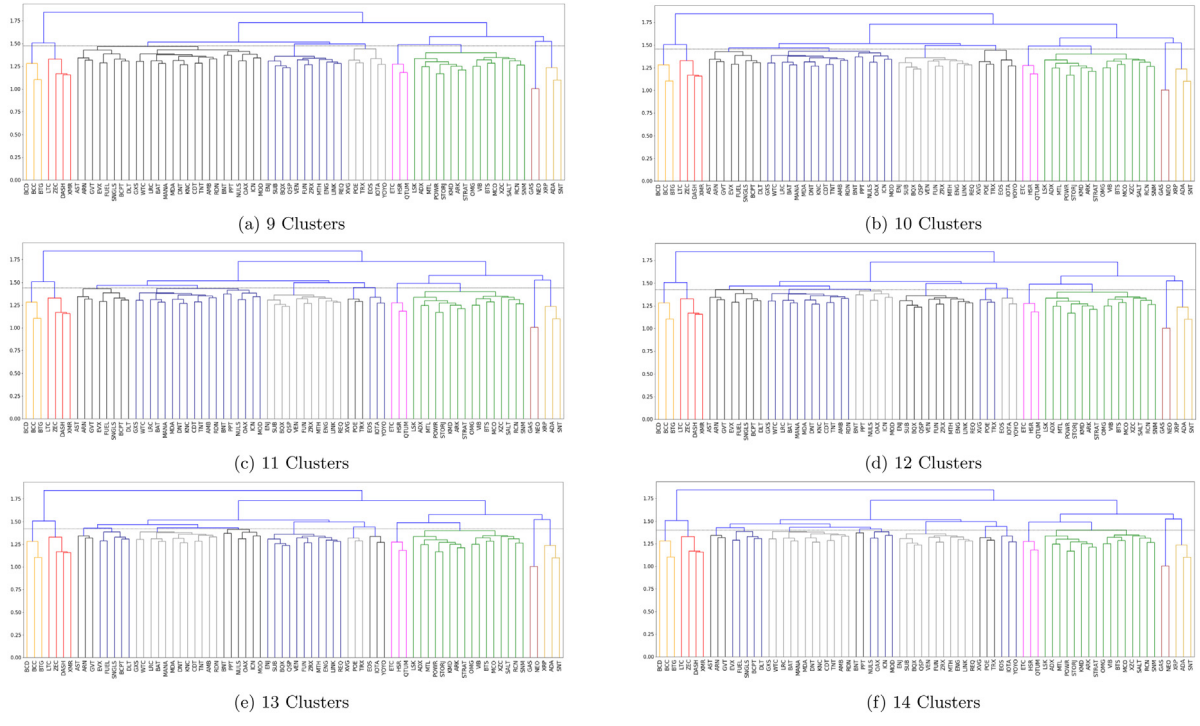


(a) Total period before filtering (b) Total period after filtering

(c) Pre-regulation period before filtering (d) Pre-regulation period after filtering

(e) Post-regulation period before filtering (f) Post-regulation period after filtering

Fig. 3. Agglomerative hierarchical clustering for six matrices.



(a) 9 Clusters (b) 10 Clusters

(c) 11 Clusters (d) 12 Clusters

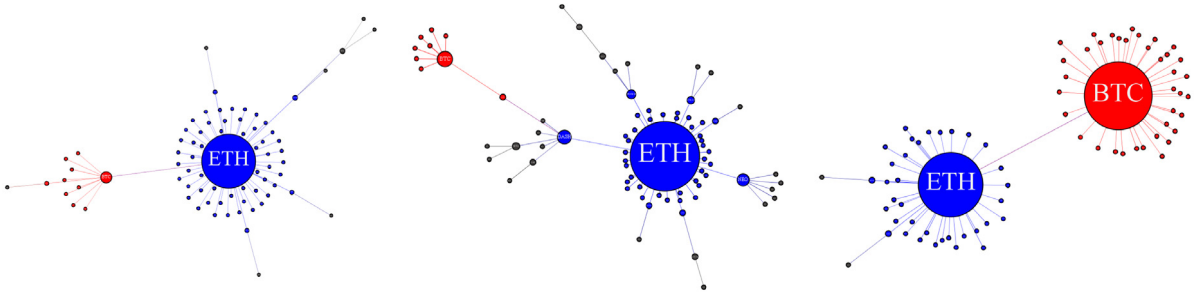
(e) 13 Clusters (f) 14 Clusters

Fig. 4. Agglomerative hierarchical clustering for total period.

platform-based cryptocurrencies. Furthermore, the six major clusters are also presented in a similar pattern for the Pre-regulation period in Fig. 3d. However, Cluster#2, #3, and #5 are only observed for the Post-regulation period in Fig. 3f.

Table 10
Correlations of major clusters.

	Cluster#1	Cluster#2	Cluster#3	Cluster#4	Cluster#5	Cluster#6	Others
(a) Total period before Bitcoin–Ethereum filtering							
mean correlation	0.5591	0.7649	0.6315	0.5227	0.7373	0.6592	0.4564
max correlation	0.7014	0.7649	0.6883	0.6751	0.7693	0.6857	0.8458
min correlation	0.4231	0.7649	0.5903	0.4132	0.7012	0.6281	0.2897
(b) Pre-regulation period before Bitcoin–Ethereum filtering							
mean correlation	0.4739	0.6908	0.4946	0.4663	0.6603	0.5800	0.3470
max correlation	0.5947	0.6908	0.5787	0.5969	0.7082	0.6052	0.7745
min correlation	0.3101	0.6908	0.4462	0.3411	0.6107	0.5320	0.1495
(c) Post-regulation period before Bitcoin–Ethereum filtering							
mean correlation	0.6793	0.8248	0.8473	0.5915	0.8207	0.7531	0.6123
max correlation	0.8283	0.8248	0.8658	0.7714	0.8320	0.7818	0.9293
min correlation	0.5332	0.8248	0.8199	0.5012	0.8013	0.7381	0.4345
(d) Total period after Bitcoin–Ethereum filtering							
mean correlation	0.1717	0.4967	0.3173	0.2849	0.2569	0.2457	0.0987
max correlation	0.3204	0.4967	0.3964	0.3914	0.3318	0.3035	0.2350
min correlation	0.0896	0.4967	0.2579	0.1854	0.1754	0.1901	−0.0155
(e) Pre-regulation period after Bitcoin–Ethereum filtering							
mean correlation	0.1956	0.5007	0.3057	0.3408	0.2869	0.2822	0.1003
max correlation	0.3413	0.5007	0.3902	0.4461	0.3743	0.3660	0.2554
min correlation	0.0949	0.5007	0.2477	0.2200	0.1890	0.2060	−0.0346
(f) Post-regulation period after Bitcoin–Ethereum filtering							
mean correlation	0.1484	0.5036	0.4420	0.1524	0.2192	0.2186	0.1391
max correlation	0.3065	0.5036	0.4869	0.2125	0.3182	0.2560	0.3233
min correlation	0.0424	0.5036	0.3880	0.1089	0.1288	0.1934	0.0205



(a) Total period (b) Pre-regulation period (c) Post-regulation period

Fig. 5. Minimum spanning tree before Bitcoin–Ethereum filtering.

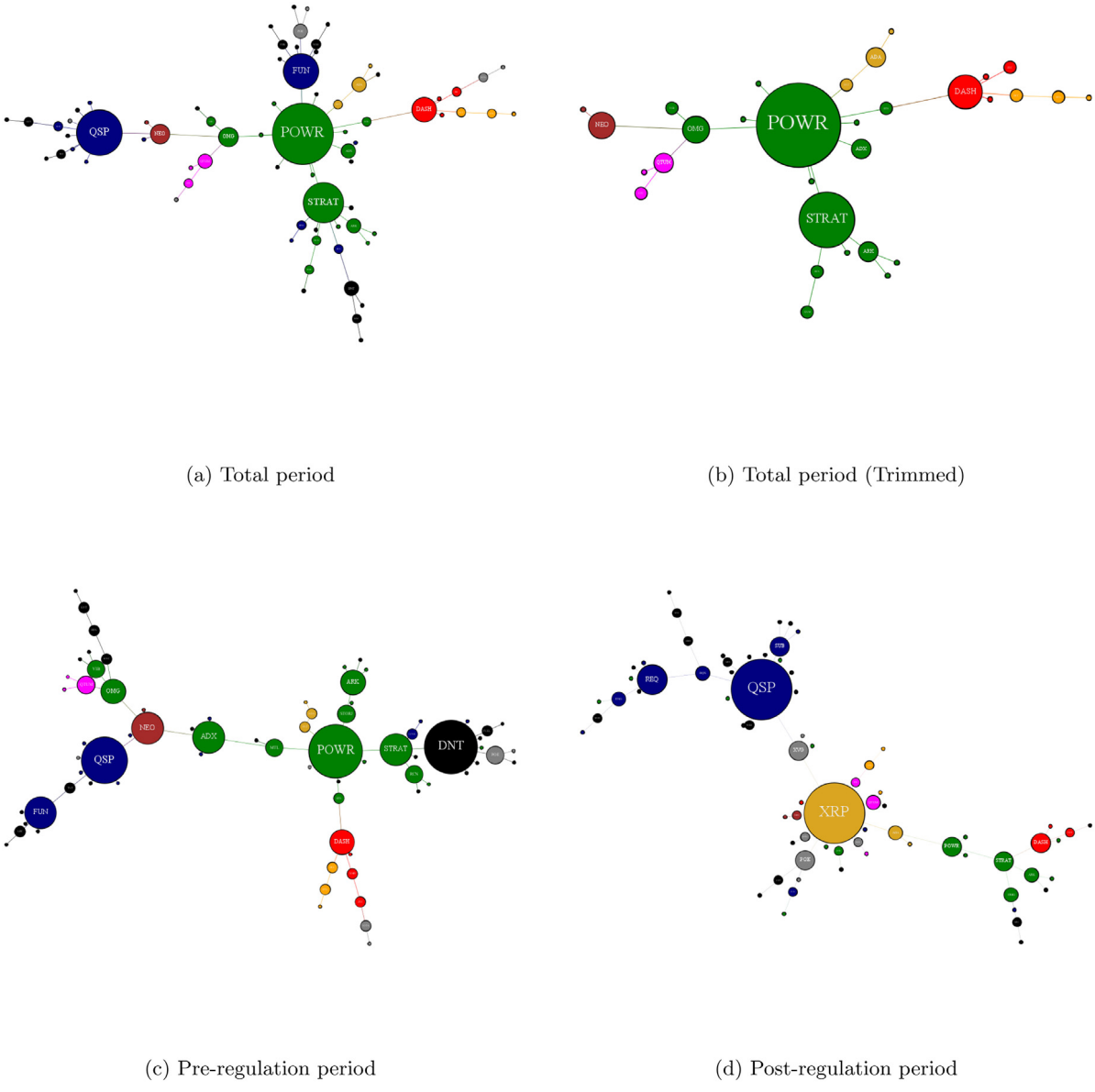
Interestingly, Cluster#5 includes the DASH, ZEC, and XMR and excludes the LTC. Given that the LTC is not a dark-coin, it seems that Cluster#5 develops into stronger homogeneity. Lastly, the correlations of major clusters and other coins are summarized in Table 10.

The mean, maximum, and minimum correlations decrease after filtering in all clusters where a higher correlation is detected in the Post-regulation period than in the Pre-regulation period. Both the before and after filtering, the mean correlations are higher in the major clusters than others for all periods. Also, the maximum and minimum correlations are higher in the Total and Pre-regulation period than that in the Post-regulation. Based on the results, we conclude that the structure of the market has been changed after the announcements of various regulations.

3.3. Minimum spanning tree analysis

The results of clustering analysis can be further investigated based on the correlation-based MSTs of cryptocurrency markets. The results of the Total, Pre-, and Post-regulation periods before filtering are plotted in Fig. 5.

Note that the red and blue colors indicate the connected nodes to the Bitcoin and Ethereum, respectively. According to MSTs, most of the cryptocurrencies are closely connected to the Bitcoin and Ethereum for all periods. Specifically, the Ethereum forms a stronger cluster than Bitcoin in the Total and Pre-regulation period, whereas the Bitcoin-centered cluster



(a) Total period

(b) Total period (Trimmed)

(c) Pre-regulation period

(d) Post-regulation period

Fig. 6. Minimum spanning tree after Bitcoin–Ethereum filtering.

is dramatically enlarged in the Post-regulation period. Although the cryptocurrencies are divided into small clusters in the Pre-regulation period, the center of them is directly connected from the Ethereum. In order to identify the homogeneity among minor cryptocurrencies, we plot the MSTs of correlation matrices after filtering in Fig. 6 based on the return series in Eq. (8) and distance in Eq. (10)

Note that the color scheme of nodes is analogous to the clustering analysis in Fig. 4. At first, the MST of Total period in Fig. 6a shows more diversified clusters with specified color patterns than that in before filtering. Fig. 6b is trimmed MST using the major clusters in Table 9. Based on the original and trimmed MSTs for the Total period, Cluster#1 in green color can be considered as the major cluster due to its location in the center of MST. The MST of the Pre-regulation period in Fig. 6c shows a similar connectedness to the Total period with some disconnections, whereas that of the Post-regulation period in Fig. 6d exhibits different connectedness. The Post-regulation period presents the detachment of Cluster#1 and the rise of XRP at the center of the network. Note that the XRP quickly became a cryptocurrency with the third highest market capitalization as soon as it appeared in the market. However, the MST of the Post-regulation still shows the strong connectedness in Cluster#2, #3, and #5 as appeared in the clustering analysis. Besides, we conduct a side experiment to investigate if the structural change in the Post-regulation period is solely incurred from the stepwise filtering mechanism used in Table 8. Thus, we construct the MST of the Post-regulation period based on the combined filtering and compare

Table 11
Statistics of minimum spanning trees.

	Before filtering			After filtering		
	Total	Pre-regulation	Post-regulation	Total	Pre-regulation	Post-regulation
mean correlation	0.4935	0.3827	0.6458	0.1074	0.1128	0.1388
max correlation	0.8458	0.7745	0.9293	0.4967	0.5007	0.5036
min correlation	0.2384	0.1305	0.3547	−0.0239	−0.0616	−0.0323
Normalized tree length	0.8473	0.9636	0.6640	1.2508	1.2304	1.2212
mean occupation layer	1.2500	1.7105	1.500	2.7297	3.7162	3.1757
max connection	58	42	38	12	8	12
Center node	ETH	ETH	BTC	POWR	POWR	XRP

it with the current MST. The result shows that the proportion of matched edges between both MSTs is roughly 78.08%(57 out of 73 edges). Also, the structure of the MST based on the combined filtering is distinctly different from those of the Total and Pre-regulation periods. Therefore, we conclude that the structural change in the Post-regulation period is not the result of different filtering mechanism.

Lastly, the statistics of MSTs are provided in Table 11. The correlations and normalized tree length of MSTs are reduced and lengthened after filtering, which implies the strong influences of Bitcoin and Ethereum in the cryptocurrency market. The mean occupation layer of MSTs is small for all periods before filtering. Smaller the mean occupation layer, higher the density of MST is. According to Fig. 5, the strong connectedness toward the Bitcoin and Ethereum incurs high densities. In contrast, the mean occupation layer is ample for all periods after filtering. Lastly, the center node of MSTs before filtering is limited to the Bitcoin and Ethereum, whereas that of after filtering is changed to the POWR in Cluster#1 for the Total and Pre-regulation periods and the XRP in Cluster#3 for the Post-regulation period. Besides, the low maximum connections in the MSTs suggest the presence of several central nodes for homogeneous clusters.

4. Conclusion

Throughout the paper, we analyze the structure of the cryptocurrency market using the agglomerative hierarchical clustering and MST. The duration of the experiment includes the strong bull and bearish markets dramatically switched by the announcements of various regulations. Hence, we detect the structural breakpoint based on the Chow test statistics and classify the duration into the Total, Pre-, and Post-regulation periods. Furthermore, since it is difficult to analyze the market structure with extremely high correlations among cryptocurrencies, we propose Bitcoin–Ethereum filtering to eliminate the linear influences from the Bitcoin and Ethereum to other cryptocurrencies. Then, the market is investigated using the correlation matrices of non-filtered and filtered return series for different periods.

Our paper is novel in the following reasons. At first, in our best knowledge, this is the first attempt to examine the structure of the cryptocurrency market by excluding the influences of the Bitcoin and Ethereum via the OLS filtering. Based on the results of clustering and MST analysis, we discover that the Bitcoin and Ethereum lead the market as major assets, which produce a relatively simple structure of the cryptocurrency market with two clusters. However, due to the strong influence of two major assets, it is hard to observe the homogeneous clusters among minor cryptocurrencies. Therefore, we apply the clustering and MST based on the filtered return series and observe the six homogeneous clusters in the Total period. Also, we discover the explicable homogeneity in every six clusters based on its characteristics in creation and trade. Lastly, we confirm that most of the homogeneous clusters are retained in the Total and Pre-regulation periods, whereas only Cluster #2, #3, and #5 are selectively maintained in the Post-regulation period. Based on these results, we conclude that the structure of the cryptocurrency market is transformed after the announcement of regulations from various countries. Since the breakpoint is concurrent with the burst of the Bitcoin bubble, the structural changes in the Post-regulation period can be interpreted as the result of the downfall of the cryptocurrency market.

Despite its findings, our research has limitations in its consideration of the OLS method, which only eliminates the linear influences of Bitcoin and Ethereum. We believe that the non-linear relationship should be considered in further study. Also, the duration of the experiment is relatively short since the cryptocurrency trade has recently been organized and activated. Therefore, we are planning to perform a similar analysis based on a more extended period of data in mid-frequency. Notwithstanding all the limitations, we believe that our research contributes to the research in the cryptocurrency market in the early stage.

Acknowledgements

This research was supported by the internal research project funded by the SNU Institute for Research in Finance and Economics, Republic of Korea (No. 0666A-20180003) and the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (No. 2018R1C1B5043835).

References

[1] S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system.
[2] P. Ciaian, M. Rajcaniova, et al., Virtual relationships: Short-and long-run evidence from bitcoin and altcoin markets, J. Int. Financ. Markets Institutions Money 52 (2018) 173–195.
[3] A. Urquhart, The inefficiency of bitcoin, Econom. Lett. 148 (2016) 80–82.
[4] A.F. Bariviera, M.J. Basgall, W. Hasperué, M. Naiouf, Some stylized facts of the bitcoin market, Physica A 484 (2017) 82–90.
[5] T. Takaishi, Statistical properties and multifractality of bitcoin, Physica A 506 (2018) 507–519.
[6] A.C. da Silva Filho, N.D. Maganini, E.F. de Almeida, Multifractal analysis of bitcoin market, Physica A (2018).
[7] W. Zhang, P. Wang, X. Li, D. Shen, The inefficiency of cryptocurrency and its cross-correlation with dow jones industrial average, Physica A 510 (2018) 658–670.
[8] L. Kristoufek, On bitcoin markets (in) efficiency and its evolution, Physica A 503 (2018) 257–262.
[9] A.K. Tiwari, R. Jana, D. Das, D. Roubaud, Informational efficiency of bitcoin–an extension, Econom. Lett. 163 (2018) 106–109.
[10] R.N. Mantegna, Hierarchical structure in financial markets, Eur. Phys. J. B 11 (1) (1999) 193–197.
[11] V. Plerou, P. Gopikrishnan, B. Rosenow, L. Amaral, H. Stanley, Collective behavior of stock price movements–a random matrix theory approach, Physica A 299 (1–2) (2001) 175–180.
[12] P. Gopikrishnan, B. Rosenow, V. Plerou, H.E. Stanley, Quantifying and interpreting collective behavior in financial markets, Phys. Rev. E 64 (3) (2001) 035106.
[13] J. Kwapień, S. Drożdż, Physical approach to complex systems, Phys. Rep. 515 (3–4) (2012) 115–226.
[14] T. Peron, F.A. Rodrigues, Collective behavior in financial markets, Europhys. Lett. 96 (4) (2011) 48004.
[15] R.K. Pan, S. Sinha, Collective behavior of stock price movements in an emerging market, Phys. Rev. E 76 (4) (2007) 046116.
[16] N. Basalto, R. Bellotti, F. De Carlo, P. Facchi, S. Pascazio, Clustering stock market companies via chaotic map synchronization, Physica A 345 (1–2) (2005) 196–206.
[17] T.W. Liao, Clustering of time series data–a survey, Pattern Recognit. 38 (11) (2005) 1857–1874.
[18] M. Tumminello, F. Lillo, R.N. Mantegna, Correlation, hierarchies, and networks in financial markets, J. Econ. Behav. Organ. 75 (1) (2010) 40–58.
[19] A. Kocheturov, M. Batsyn, P.M. Pardalos, Dynamics of cluster structures in a financial market network, Physica A 413 (2014) 523–533.
[20] S.S. Jung, W. Chang, Clustering stocks using partial correlation coefficients, Physica A 462 (2016) 410–420.
[21] M.R. Berthold, F. Höppner, On clustering time series using euclidean distance and pearson correlation, arXiv preprint arXiv:1601.02213.
[22] T. Heimo, J. Saramäki, J.-P. Onnela, K. Kaski, Spectral and network methods in the analysis of correlation matrices of stock returns, Physica A 383 (1) (2007) 147–151.
[23] W.-Q. Huang, X.-T. Zhuang, S. Yao, A network analysis of the chinese stock market, Physica A 388 (14) (2009) 2956–2964.
[24] J.G. Brida, W.A. Risso, Hierarchical structure of the german stock market, Expert Syst. Appl. 37 (5) (2010) 3846–3852.
[25] S. Radhakrishnan, A. Duvvuru, S. Sultornsanee, S. Kamarthi, Phase synchronization based minimum spanning trees for analysis of financial time series with nonlinear correlations, Physica A 444 (2016) 259–270.
[26] P. Coletti, Comparing minimum spanning trees of the Italian stock market using returns and volumes, Physica A 463 (2016) 246–261.
[27] V. Tola, F. Lillo, M. Gallegati, R.N. Mantegna, Cluster analysis for portfolio optimization, J. Econom. Dynam. Control 32 (1) (2008) 235–258.
[28] S. Nanda, B. Mahanty, M. Tiwari, Clustering Indian stock market data for portfolio management, Expert Syst. Appl. 37 (12) (2010) 8793–8798.
[29] D. Cheong, Y.M. Kim, H.W. Byun, K.J. Oh, T.Y. Kim, Using genetic algorithm to support clustering-based portfolio optimization by investor information, Appl. Soft Comput. 61 (2017) 593–602.
[30] C. Iorio, G. Frasso, A. D'Ambrosio, R. Siciliano, A p-spline based clustering approach for portfolio selection, Expert Syst. Appl. 95 (2018) 88–103.
[31] L.S. Junior, I.D.P. Franca, Correlation of financial markets in times of crisis, Physica A 391 (1–2) (2012) 187–208.
[32] T. Kauê Dal'Maso Peron, L. da Fontoura Costa, F.A. Rodrigues, The structure and resilience of financial market networks, Chaos 22 (1) (2012) 013117.
[33] R.H. Heiberger, Stock network stability in times of crisis, Physica A 393 (2014) 376–381.
[34] J.W. Song, B. Ko, P. Cho, W. Chang, Time-varying causal network of the Korean financial system based on firm-specific risk premiums, Physica A 458 (2016) 287–302.
[35] J.W. Song, B. Ko, W. Chang, Analyzing systemic risk using non-linear marginal expected shortfall and its minimum spanning tree, Physica A 491 (2018) 289–304.
[36] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, Dynamic asset trees and black monday, Physica A 324 (1–2) (2003) 247–252.
[37] W. Jang, J. Lee, W. Chang, Currency crises and the evolution of foreign exchange market: Evidence from minimum spanning tree, Physica A 390 (4) (2011) 707–718.
[38] J. Dias, Spanning trees and the eurozone crisis, Physica A 392 (23) (2013) 5974–5984.
[39] A. Nobi, S.E. Maeng, G.G. Ha, J.W. Lee, Effects of global financial crisis on network structure in a local stock market, Physica A 407 (2014) 135–143.
[40] T.U. Kuzubaş, I. Ömercikoğlu, B. Saltoğlu, Network centrality measures and systemic risk: An application to the turkish financial crisis, Physica A 405 (2014) 203–215.
[41] M. Majapa, S.J. Gossel, Topology of the South African stock market network across the 2008 financial crisis, Physica A 445 (2016) 35–47.
[42] D. Stosic, D. Stosic, T.B. Ludermir, T. Stosic, Collective behavior of cryptocurrency price changes, Physica A 507 (2018) 499–509.
[43] W.F. Sharpe, The sharpe ratio, J. Portfolio Manage. 21 (1) (1994) 49–58.
[44] W.A. Fuller, Introduction to statistical time series, 428, John Wiley & Sons, 2009.
[45] P.C. Phillips, P. Perron, Testing for a unit root in time series regression, Biometrika 75 (2) (1988) 335–346.
[46] G.C. Chow, Tests of equality between sets of coefficients in two linear regressions, Econometrica (1960) 591–605.
[47] J.H. Ward Jr, Hierarchical grouping to optimize an objective function, J. Amer. Statist. Assoc. 58 (301) (1963) 236–244.
[48] J.B. Kruskal, On the shortest spanning subtree of a graph and the traveling salesman problem, Proc. Amer. Math. Soc. 7 (1) (1956) 48–50.
[49] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, A. Kanto, Dynamics of market correlations: Taxonomy and portfolio analysis, Phys. Rev. E 68 (5) (2003) 056110.