

# Whose Opinion Matters? Analyzing Relationships Between Bitcoin Prices and User Groups in Online Community

Social Science Computer Review  
2020, Vol. 38(6) 686-702  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/0894439319840716  
journals.sagepub.com/home/ssc



Kyeongpil Kang<sup>1</sup>, Jaegul Choo<sup>1</sup>, and YoungBin Kim<sup>2</sup>

## Abstract

Public interest in cryptocurrencies has consistently risen over the past decade. Owing to this rapid growth, cryptocurrency-related information is being increasingly shared online. As considerable portions of such information in online communities are noise, extracting meaningful information is important. Therefore, judging whose opinion should be considered more important or who the opinion leaders in online communities are is critical. This study analyzed the topics that contain meaningful information, in particular, user groups, by investigating the correlation between topic weights and their price change. The proposed analysis method involves (1) effective classification of the user groups using a hypertext-induced topic selection algorithm, (2) textual information analysis through topic modeling, and (3) the identification of user groups that have a high interest in the Bitcoin price by measuring the correlation between the price and the topics and by measuring the topic similarities between each user group and all users to determine the user group that can effectively represent the entire community. By analyzing the information shared by users, we observed that most users are interested in the price information, whereas users having social influence are not only interested in the price but also in other information.

## Keywords

opinion leader, network analysis, topic modeling, Bitcoin community

In the past decade, new currencies have begun to emerge with the development of the Internet. With the introduction of cryptocurrencies, the economic activities among users on the Internet underwent enormous changes. Mining cryptocurrencies using a unique blockchain technology results in an increase in the total quantity of currencies. Bitcoin was first introduced in 2008 (Nakamoto, 2008), and various types of cryptocurrencies that adopted a similar blockchain technology appeared at the

---

<sup>1</sup> Korea University, Seoul, South Korea

<sup>2</sup> Chung-Ang University, Seoul, South Korea

## Corresponding Author:

YoungBin Kim, Chung-Ang University, Bldg 305 (Faculty Office Building), 84, Heukseok-ro, Dongjak-gu, Seoul 06974, South Korea.

Email: ybkim85@cau.ac.kr

onset of the 2010s (Barber, Boyen, Shi, & Uzun, 2012; Bohme et al., 2015; Grinberg, 2012; Wood, 2014). Consumers are now increasingly using cryptocurrencies such as Bitcoin for all types of payments and are trading them in cryptocurrency markets (Barber et al., 2012; Grinberg, 2012; Y. B. Kim et al., 2016; M. Kim, Kang, Park, Choo, & Elmqvist, 2017; Marella, Lindman, Rossi, & Tuunainen, 2017; Reid & Harrigan, 2013).

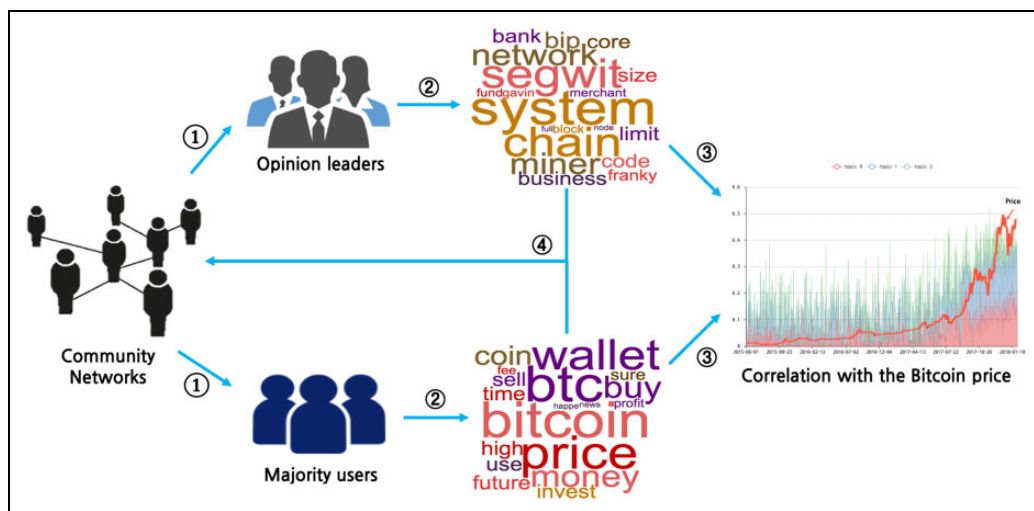
The price history of Bitcoin until now has shown considerable fluctuations. The Bitcoin price rises within a short period while plunging at an alarming rate. Specifically, the Bitcoin currency started 2017 at the levels of 997 USD per Bitcoin and increased to 19,000 USD in December 2017, potentially creating a profit of almost 1,900% in less than a year. In June 2018, the price fell drastically to 6,000 USD, which corresponds to a loss of approximately 70% since December 2017. This trend cannot be sufficiently explained by standard macroeconomic theories (Aalborg, Molnár, & de Vries, 2018; Gerlach, Demos, & Sornette, 2018; Kristoufek, 2013; Li & Wang, 2017).

Originally, the main function of Bitcoin was to act as a substitute for fiat money; however, there exist differences between fiat money and cryptocurrencies. Fiat money is controlled by a central system, but cryptocurrencies are not. Therefore, although a blockchain is used in various fields such as remittance and Darknet, Bitcoin markets are now often recognized as sites for speculators. These markets, unlike the real money market, are currently and considerably influenced by short-term investors, trend chasers, speculators (Kristoufek, 2013), and large numbers of fraud activities (Moeller, Munksgaard, & Demant, 2017). Therefore, the analysis of these traders, activities, and cryptomarkets becomes important to invigorate cryptocurrencies and the blockchain (Demant, Munksgaard, Décary-Héту, & Aldridge, 2018; Moeller et al., 2017). In addition, the majority of Bitcoin investors determine their investment strategy on the basis of the scant information they can obtain because there is no perception of fair prices in this market (Aalborg et al., 2018; Kristoufek, 2013) and there is little information on cryptocurrency markets compared to real markets. Therefore, in this study, we focused on the characteristics of online Bitcoin discussion forums, which are a potentially rich source of information.

Most Bitcoin trading activities take place online, and users share information related to Bitcoin at almost no cost in specific online discussion forums. They effectively decide to buy or sell Bitcoins on the basis of this information (Y. B. Kim et al., 2016; Kim et al., 2017; Phillips & Gorse, 2018). Although responses from many users can be found in real time in these forums, public online discourses have a large amount of noise, misinformation, and spam users; thus, the identification of information that is meaningful and the users who provide high-quality information among the vast user opinion data tend to be difficult (Waldherr, Maier, Miltner, & Günther, 2017). Therefore, the analysis of the relationship between user opinions and price changes is important to verify their meaningfulness and extract verified users and documents from communities.

This study focuses on user opinions and identifies the opinion leaders, also known as influencers, using community networks by analyzing which user group's opinion is highly related to the price change and therefore meaningful. In this article, we propose the analysis method shown in Figure 1. We first construct community networks using the data collected from the Bitcoin forum and effectively classify user groups using the hypertext-induced topic selection (HITS) algorithm based on the information exchange in the Bitcoin online community. After conducting topic modeling for each user group, we analyze which user group demonstrates a strong correlation between the topics from the user group and the Bitcoin price. Finally, we measure the topic similarities between topics from each user group and the topics from all users to determine which user group can effectively represent the entire community.

The findings of this study based on an analysis of documents written by users in the Bitcoin forum and responses over the past 3 years demonstrate that most users are only interested in the price information, whereas the top users, defined as the opinion leaders in this article, who shared



**Figure 1.** Overview of the proposed analysis approach. After forming community networks, we (1) divide users into an opinion leader group and a majority user group, (2) conduct topic modeling using documents from each user group, (3) calculate the Pearson correlation between topic weights and the Bitcoin price, and (4) measure the similarities between topics from each group and topics for all users.

information frequently, show significant interest in diverse information and actively share it with each other.

## Related Work

Various events in the Bitcoin market cannot be sufficiently explained by standard macroeconomic theories and this market is characteristically online. Since 2013, there has been considerable research on connecting Bitcoin with various sources of online information to analyze and utilize Bitcoin and its market. Numerous studies have covered a wide range of topics from the trends and various social events related with the appearance of Bitcoin to the characteristics of the Bitcoin price. These studies analyzed user sentiments related to Bitcoin on social media such as Twitter or quantified web search queries on search engines such as Google Trends and analyzed the fluctuations in the price and transaction volume to determine any relation (Cheah & Fry, 2015; Ifrim, Shi, & Brigadir, 2014; Karalevicius, Degrande, & De Weerd, 2018; Kaminski, 2014; Kristoufek, 2013; Mai, Shan, Bai, Wang, & Chiang, 2018; Phillips & Gorse, 2017; Yelowitz & Wilson, 2015; Zhang, Wang, Li, & Shen, 2018). These studies are more likely to be used by various online sources as material for understanding and analyzing Bitcoin. Despite this, there have been almost no studies seeking to identify users who share information that could be useful. An average analysis of entire groups should comparatively involve noise and is likely to degrade the quality of the analysis. Therefore, finding users and groups that share meaningful information is necessary in order to address this problem.

People with similar interests such as making money and investment gather together, write articles on these topics, and exchange ideas in social communities (Bickart & Schindler, 2001; Bernstein et al., 2011; Kim et al., 2017; Panzarasa, Opsahl, & Carley, 2009). Previous studies on user opinions in online discussion forums and price fluctuation performed sentimental and quantitative analyses or focused on certain phases where the Bitcoin price soared. Several studies attempted to investigate how the Bitcoin price is related to users' feelings or opinions by analyzing Bitcoin forums (Y. B.

Kim et al., 2016; Kim et al., 2017; Y. B. Kim et al., 2015; Phillips & Gorse, 2018). Although Kim's works form the motive for our research, we mainly focused on finding opinion leaders using network analysis and analyzing different behaviors and interest between opinion leaders and other majority users and not on predicting prices or volumes using causality analysis between the price and the word frequencies. Furthermore, studies on the prediction of the Bitcoin value or trading volume were performed (Alessandretti, ElBahrawy, Aiello, & Baronchelli, 2018; Guo & Antulov-Fantulin, 2018; Indera, Yassin, Zabidi, & Rizman, 2017; Y. B. Kim et al., 2016; Kim et al., 2017; Li, Chamrajnagar, Fong, Rizik, & Fu, 2018; Matta, Lunesu, & Marchesi, 2015; McNally, Roche, & Caton, 2018; Sin & Wang, 2017). These studies focused more on the prediction of the Bitcoin price fluctuation itself rather than on finding information that could be useful, such as opinions or opinion leaders within the forum. Many studies also sought to analyze community networks and users' opinions to identify opinion leaders or to find the main theme and maximize network marketing effects (Alvarez-Galvez, 2016; Choi, 2015; Ho et al., 2016; Junbo, Min, Fan, & Xufa, 2005; Jiang et al., 2014; Maier, Waldherr, Miltner, Jähnichen, & Pfetsch, 2018; Wu et al., 2015; Wang, Du, & Tang, 2016; Zhao, Kou, Peng, & Chen, 2018). Users who actively participate in the community and have high reputation are called opinion leaders and they play an important role in leading the consensus in the social community (Chan & Misra, 1990; Johnson, Safadi, & Faraj, 2015; Ming Yu, 2002). Rogers (2010) showed that opinions shared by opinion leaders propagate much faster in social communities than those of public users. Several surveys have created a simulation environment to understand the flow of opinions and identify opinion leaders (Zhao et al., 2018) and have attempted to identify the opinion leaders in a group (Ho et al., 2016; Jiang et al., 2014; Wu et al., 2015; Wang et al., 2016). Several authors focused on developing improved algorithms based on PageRank to identify opinion leaders in social communities (Jiang et al., 2014; Wang et al., 2016).

## Method

People actively communicate with each other by writing posts and posting comments on other users' posts in online communities such as the Bitcoin forum. In this section, we introduce the process of data collection and preprocessing and analyze the difference between user groups by constructing community networks and performing topic modeling.

### *Data Collection and Preprocessing*

First, we crawled posts and comments including metadata such as the post time and username from a Bitcoin forum (<https://bitcointalk.org/index.php?board=1.0>) using Python, Requests, and the BeautifulSoup library. Users in this forum discuss Bitcoin, the blockchain system of Bitcoin, prices, and investments, among other topics. As Bitcoin developers and core members have discussed the development of Bitcoin since its initial stages in this forum, this community can be considered as a representative community among various Bitcoin communities. We collected a total of 36,898 posts and 353,679 comments written by 32,304 users from November 22, 2009, when the forum was created to February 2, 2018, when we finished crawling. We crawled data in a legitimate manner in compliance with terms at <https://bitcointalk.org/robots.txt>. In addition, the crawled data do not include any personal information.

After collecting the data, we applied data preprocessing to the raw text for (1) parsing each document into word tokens, (2) stemming each word using the WordNet lemmatizer in the Python NLTK (Natural Language Toolkit) library (Loper & Bird, 2002), (3) converting each word into lowercase, and (4) removing non-English characters, special characters, and stop words, which are words that have fewer than two characters and words that appeared fewer than 10 times. We also

discarded posts and comments that are five words or less. After data preprocessing, we collected 390,572 posts and comments with a vocabulary of 12,021 terms.

Subsequently, we constructed community networks and analyzed the network topology using the Python NetworkX library (Hagberg, Swart, & Chult, 2008). There exist network or link analysis algorithms such as PageRank and HITS (Kleinberg, 1999; Page, Brin, Motwani, & Winograd, 1999). In this study, the HITS algorithm was employed to measure the degree of significance of each user. We then extracted both the highest and lowest ranked users and performed topic modeling.

## Network Analysis

We assume that there exists an edge if user  $u_i$  posted a comment on an article by another user  $u_j$  given the user set  $\mathbb{U} = \{u_1, \dots, u_n\}$ . In other words, a set of directed edges can be defined as:

$$\mathbb{E} = \{(v_{u_i}, v_{u_j}) | \text{If user } u_i \text{ write comments on } u_j\text{'s posts}\}. \quad (1)$$

Therefore, a directed graph  $\mathbb{G} = \{\mathbb{V}, \mathbb{E}\}$  can be defined as a community network regarding a user node set  $\mathbb{V} = \{v_{u_1}, \dots, v_{u_n}\}$  and an association set  $\mathbb{E} = \{(v_{u_i}, v_{u_j})\}$ . After community networks are built, we calculate the degree of social influence or importance of each user node using the HITS algorithm. Unlike the PageRank algorithm, the HITS algorithm calculates both the authority and hub scores per node to examine the degree of significance of each node. The authority scores herein indicate the extent of information or influence that incoming neighbor nodes receive from and the attention they pay to a particular node. In contrast, hub scores indicate the extent of information that each node provides or its influence on outgoing neighbor nodes and the attention it pays to these outgoing neighbor nodes. The authority and hub scores can be calculated per user node  $v_{u_i}$  using the following equations:

$$\text{Authority}(v_{u_i}) = \sum_{v_{u_j} \in \mathbf{In}(v_{u_i})} \text{hub}(v_{u_j}), \quad (2)$$

$$\text{Hub}(v_{u_i}) = \sum_{v_{u_j} \in \mathbf{Out}(v_{u_i})} \text{authority}(v_{u_j}), \quad (3)$$

where  $\mathbf{In}(v_{u_i})$  represents the neighbor nodes that direct attention to  $v_{u_i}$  and  $\mathbf{Out}(v_{u_i})$  represents the neighbor nodes that  $v_{u_i}$  directs attention to. The HITS algorithm calculates Equations 2 and 3 repeatedly until the authority and hub scores converge as shown in Algorithm 1.

Using the Python NetworkX library, we calculated the authority score for each user node using the HITS algorithm to assess how much attention a user gets from other users. The degree of attention gained from others indicates more influence than activity because there may be spam users or malicious users. Hence, we used the authority score and not the hub score to measure influential power of each user.

As shown in the Network Analysis subsection in the Experimental Results section, the distribution of authority scores follows Pareto's law (Koch, 2011). In many studies (Lipovetsky, 2009; Koch, 2018; Parmenter, 2007), the 80/20 rule is often adopted to split two polarity groups. Therefore, we defined users who account for the top 20% authority scores as opinion leaders  $\mathbb{O}$ ; the other users were referred to as majority users  $\mathbb{M}$ , and users including those in the opinion leader group and majority user group were defined as the user group  $\mathbb{T}$ . In this manner, we analyzed the differences between topics from each user group and the information or interest they mainly share by performing topic modeling. Furthermore, we determined that the opinion leader group shares more abundant information than the majority user group.

**Algorithm 1.** HITS algorithm.**Input:** Community networks  $G$ **Output:** Hub and authority scores

---

```

1: Initialize all authority and hub scores as 1
2: repeat
3:   for  $v_{u_i} \in U$  do
4:     for  $v_j \in \text{In}(v_i)$  do
5:        $\text{authority}(v_i) = \text{authority}(v_i) + \text{hub}(v_j)$ 
6:     end for
7:     normalize  $\text{authority}(v_i)$  to 1
8:     for  $v_j \in \text{Out}(v_i)$  do
9:        $\text{hub}(v_i) = \text{hub}(v_i) + \text{authority}(v_j)$ 
10:    end for
11:    normalize the sum of  $\text{hub}(v_i)$  to 1
12:  end for
13: until authorities and hubs converge

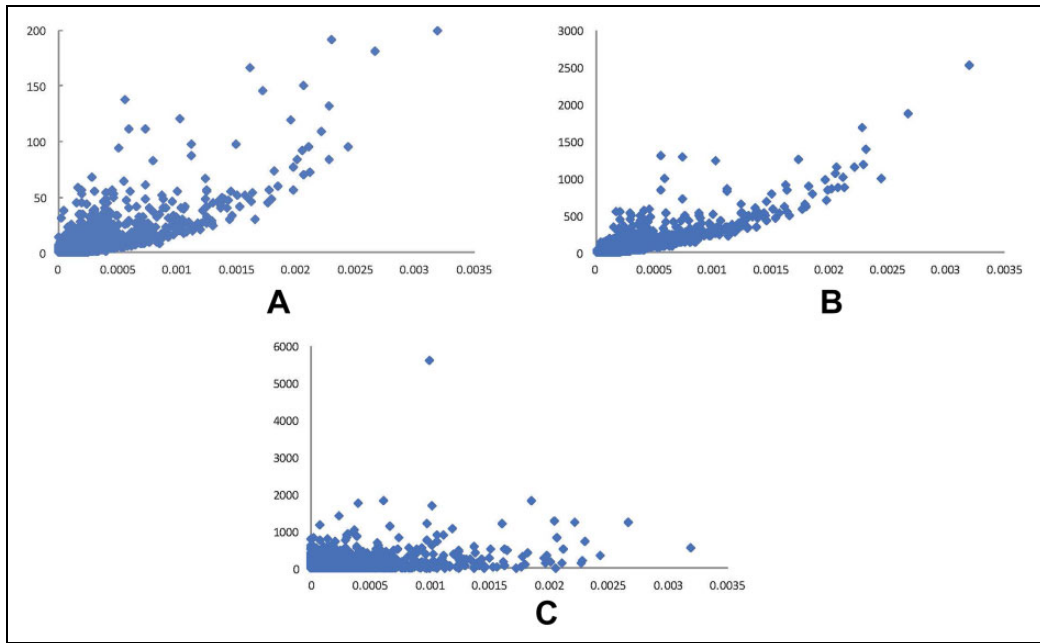
```

---

**Topic Modeling**

Topic modeling is a machine learning technique that extracts coarse-grained and abstract topics that prominently appear in a large-sized document corpus. Generally, topics can be calculated using unsupervised learning algorithms. For example, topic modeling includes probabilistic techniques such as probabilistic latent semantic indexing (p-LSI; Hofmann, 1999), latent Dirichlet allocation (Blei, Ng, & Jordan, 2003), and matrix factorization techniques such as latent semantic analysis (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990) and nonnegative matrix factorization (NMF; Lee & Seung, 1999, 2001). By conducting topic modeling, topics and key words that prominently appear per topic can be extracted from the document corpus. Using these topics and key words, we can efficiently analyze large documents by examining only the summarized information or by interactive visualization using topic information. For example, several studies (Choo, Lee, Reddy, & Park, 2013; Chaney & Blei, 2012; Kim et al., 2017; Kuang, Choo, & Park, 2015) provide visualization platforms that help experts or analysts to carry out efficient analyses using interactive tools such as hovering and data linking.

In this study, we conducted topic modeling using NMF. Before conducting topic modeling, a term-document matrix  $A \in \mathfrak{R}^{|\mathbb{V}| \times |\mathbb{D}|}$ , which stores the frequency of each word per document, should be constructed, where  $\mathbb{V} = \{w_1, w_2, \dots, w_{|\mathbb{V}|}\}$  is the vocabulary set appearing in all documents,  $\mathbb{D} = \{d_1, d_2, \dots, d_{|\mathbb{D}|}\}$  is the set of documents, and  $a_{ij}$  in  $A$  is the frequency of each word  $w_i$  per document  $d_j$ . The  $L_2$  length of each column vector in the term-document matrix  $A$  was normalized to 1. NMF factorizes  $A$  into the nonnegative matrices  $W \in \mathfrak{R}^{|\mathbb{V}| \times K}$  and  $H \in \mathfrak{R}^{K \times |\mathbb{D}|}$  as follows:



**Figure 2.** Scatterplots of the numbers of posts and comments according to the authority scores. (A) Number of posts. (B) Number of received comments. (C) Number of written comments.

$$\min_{W,H} A - WH^T{}_F^2 + \alpha W_F^2 + \beta H_F^2, \text{ where } W \geq 0, H \geq 0, \quad (4)$$

where  $\|\cdot\|_F^2$  represents the Frobenius norm and  $\|A\|_F = (\sum_{i=1}^m \sum_{j=1}^m |a_{ij}|^2)^{1/2}$ , that is, an  $L_2$  regularization term was added to reduce the overfitting problem. For updating parameters  $W$  and  $H$ , we used the coordinate descent algorithm.

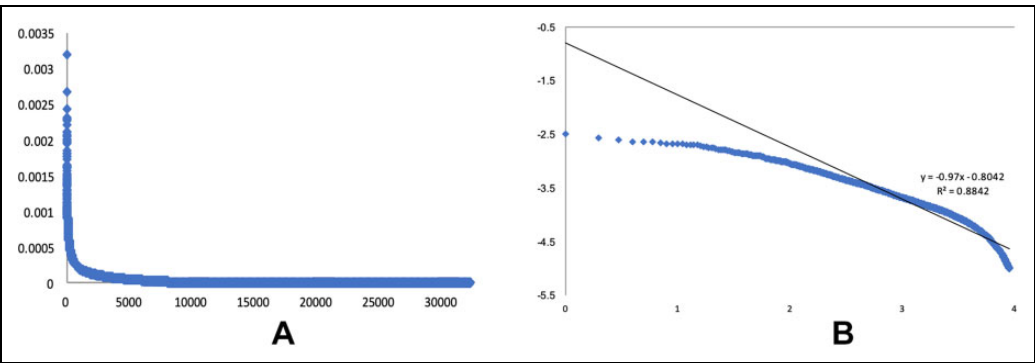
In addition to conducting topic modeling using the term-document matrix, we performed topic modeling by establishing the **term-date matrix**  $X \in \mathbb{R}^{|\mathbb{V}| \times |\mathbb{T}|}$ , where  $X_{ij}$  is the frequency of word  $w_i$  in all documents generated on each date  $t_j$  during the entire period  $\mathbb{T} = \{t_1, \dots, t_{|\mathbb{T}|}\}$  to calculate the correlations between a topic from each user group and the fluctuation in the Bitcoin price. We performed topic modeling using the Python NumPy (Walt, Colbert, & Varoquaux, 2011) and scikit-learn (Pedregosa et al., 2011) libraries.

## Experimental Results

After constructing the community networks described in the Network Analysis subsection given in the Method section, we analyzed the community networks to differentiate opinion leaders and majority users in the community, as described in the Network Analysis subsection in the Experimental Results section. Subsequently, we performed topic modeling to compare the opinion leaders to other majority users, as explained in the Topic Modeling subsection given in the Experimental Results section.

### Network Analysis

As described in the Network Analysis subsection given in the Method section, the HITS algorithm was applied to the community networks, and the authority score was calculated for each user. Figure 2



**Figure 3.** Distribution of authority scores. (A) Authority score versus the user rank. (B) Log–Log scaled graph of Authority scores.

illustrates the total number of posts, the number of comments written by each user, and the number of comments tagged to each user’s posts according to the authority scores.

Figure 2A and B shows that both the numbers of posts and comments received are highly correlated with the authority score, but Figure 2C shows a low correlation between the number of comments written and the authority score. In particular, the correlation between the number of comments received and the authority score is positive and much higher than the other correlations. This shows that authority scores tend to be high if a user receives many comments or attention from other users in the community.

In addition, the distribution of authority scores according to the user rank is shown in Figure 3A. This figure shows that a small number of top-ranked users account for most authority scores. We plot Figure 3A on a log–log scale (as shown in Figure 3B) and observe that the shape of the distribution almost follows a linear form and the coefficient of determination is approximately  $R^2 = .884$ . Thus, the distribution of the authority score follows Pareto’s law.

According to Pareto’s principle (Koch, 2011), we selected users who account for the top 80% of authority scores (3,637 users, approximately 11.26%) and defined them as the opinion leader (O), and 192,292 posts are written by them. We also defined the other users (28,667 users; 197,344 posts) as the majority set (M) to examine the difference. Moreover, we conducted experiments to determine whether topics from O can effectively represent the entire user group (total set T, 32,304 users; 390,572 posts).

First, we extracted words that are relatively more frequently used by opinion leaders rather than majority users and also words that are used more frequently by majority users. Table 1 indicates that opinion leaders use more terminology such as technical words than majority users. They are likely to be Bitcoin developers, Bitcoin miners, or businessmen producing new cryptocurrencies or initial coin offerings. In contrast, words that are observed to be more frequently used by majority users than opinion leaders are related to the price, trading, and investment. This indicates that majority users are more interested in the price, transaction, and investment of Bitcoin.

We further analyzed whether opinion users are interested in the price, transaction, or investment and if so, whether they can represent the entire community by covering a variety of topics through topic modeling.

**Topic Modeling**

We performed topic modeling considering the period from June 1, 2015, to February 1, 2018, for the entire set of users, opinion leaders, and majority users. We conducted topic modeling with 5, 10, 15,



**Table 1.** Top 25 Key Words That Are Relatively More Frequently Used by Each User Group Compared to the Other Group.

More Frequently Used by Opinion Leaders			More Frequently Used by Majority Users		
Word	Ratio (O)	Ratio (M)	Word	Ratio (O)	Ratio (M)
Block	0.448	0.311	Bitcoin	3.650	4.166
Node	0.168	0.092	Price	0.410	0.608
Core	0.136	0.067	btc	0.682	0.831
Miner	0.258	0.202	Wallet	0.435	0.548
Segwit	0.132	0.078	Good	0.384	0.476
Chain	0.149	0.105	Crypto	0.116	0.202
Network	0.261	0.219	Buy	0.267	0.350
Consensus	0.060	0.024	Country	0.167	0.249
Code	0.129	0.094	Now	0.513	0.594
Size	0.101	0.067	Money	0.533	0.610
Bank	0.325	0.296	Currency	0.412	0.484
Solution	0.099	0.074	Coin	0.490	0.557
Full	0.098	0.074	Market	0.243	0.304
Change	0.181	0.158	Time	0.571	0.628
Foundation	0.053	0.030	Invest	0.051	0.106
Blockstream	0.030	0.007	High	0.122	0.171
Limit	0.074	0.051	Future	0.165	0.212
Fund	0.133	0.111	Sell	0.141	0.186
Blockchain	0.270	0.248	use	0.504	0.549
System	0.246	0.225	Cryptocurrency	0.073	0.115
Merchant	0.087	0.067	Sure	0.200	0.242
Developer	0.065	0.045	Investment	0.087	0.128
Blocksize	0.029	0.011	Profit	0.094	0.134
Protocol	0.061	0.043	Fee	0.360	0.399
Business	0.179	0.162	News	0.117	0.155

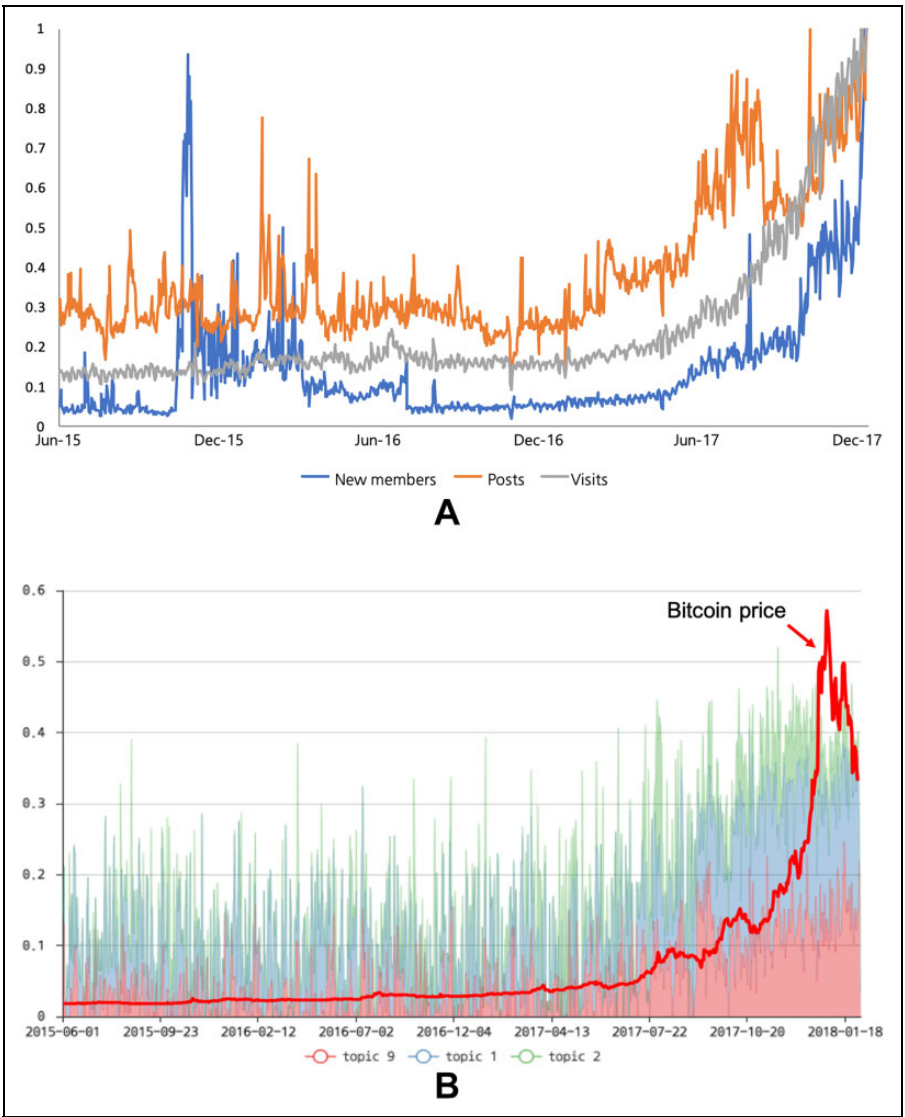
Note. The List of Words is Sorted by the Relative Ratio.

**Table 2.** Pearson Correlation Between Topic Weights and the Bitcoin Prices.

Total Users		Opinion Leaders		Majority Users	
Topic	Correlation	Topic	Correlation	Topic	Correlation
5	.688	3	.584	9	.650
9	-.301	10	.191	1	.492
6	-.275	2	-.163	2	.243
10	-.166	8	-.119	4	.148
7	-.143	7	.114	3	.120

and 20 topics. When the number of topics is more than 10, very similar topics occurred; therefore, 10 topics were observed to have the best fit. The hyperparameters  $\alpha$  and  $\beta$  for the  $L_2$  regularization term in Equation 4 were set as 0.1. We extracted the topic weights on each date and calculated the Pearson correlation with the actual Bitcoin price fluctuation. Furthermore, we drew a time-series graph of the highly correlated topics and the Bitcoin prices.

Table 2 implies that the topics generated by majority users are highly correlated with the price and are positively correlated with the quantities. In other words, the majority users are highly



**Figure 4.** Time-series graphs for community statistics and the weights of majority users' topics highly correlated to Bitcoin prices. (A) Number of new members, posts, and visits on each date. Each value is normalized to the range from 0 to 1. (B) Weights for each topic extracted from majority users and real prices.

interested in topics related to the price. In contrast, entire users and opinion leaders show a relatively low correlation than majority users and there exist a few negative correlations with the corresponding topics. These trends are also shown in Figure 4B, which depicts the weights for the top three topics from majority users that show the highest correlation to the fluctuation in the Bitcoin price. These top three topics are mainly related to the price, transaction, and investment of Bitcoin. In Figure 4B, we can clearly observe that the fluctuation of each topic that is highly correlated with the Bitcoin price is similar to the fluctuation in the Bitcoin price. These trends can also be seen in Figure 4A. After June 2017, when the Bitcoin price had drastically increased, the numbers of new members, posts, and visits also increased. These new members were not originally interested in Bitcoin, but it is likely that they became interested since the price soared. These results indicate that majority users

**Table 3.** Topic Similarities and the Average Scores for the Opinion Leader Group and Majority User Group.

Group	1	2	3	4	5	6	7	8	9	10	Avg.
O	1.000	0.969	1.000	0.992	0.994	0.995	0.990	0.999	0.998	0.998	0.994
M	1.000	0.723	1.000	0.995	0.998	0.997	0.995	0.999	0.999	0.657	0.936

are more interested in topics related to the price, whereas opinion leaders are interested in topics related to the price as well as other diverse topics such as the blockchain and mining, among others.

Therefore, we conducted further experiments to determine whether opinion leaders can effectively represent the entire community. The degree to which a user group represents the entire community can be calculated using the topic similarity for each topic with its one-to-one matching topic from the entire community. However, it is a nontrivial problem to find the one-to-one matching pairs. To solve this problem, surveys such as Greene, O'Callaghan, and Cunningham (2014) used the Hungarian algorithm, also called the Kuhn–Munkres algorithm or Munkres assignment algorithm. The Hungarian algorithm is an optimization algorithm used to solve an assignment problem. The Hungarian algorithm is calculated as follows:

$$\min_{i,j} \sum_i \sum_j C_{i,j} X_{i,j}, \quad (5)$$

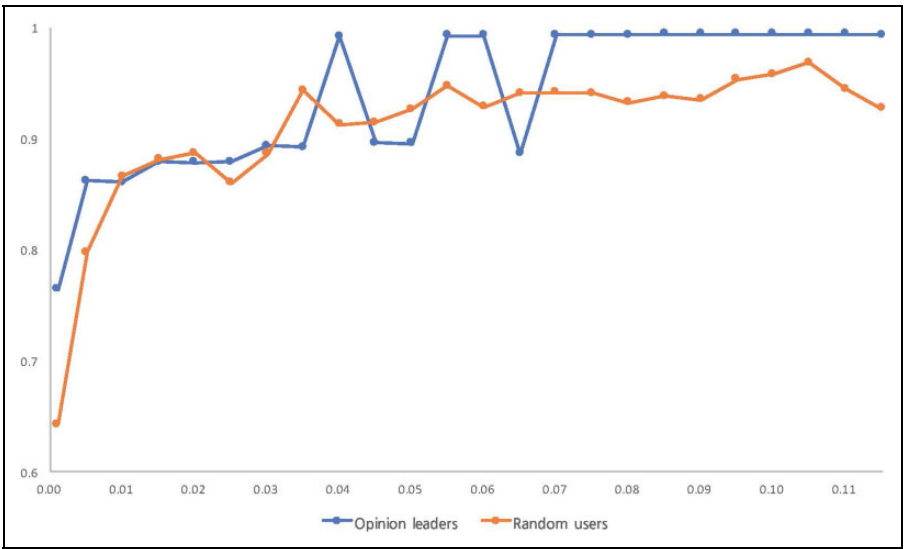
where  $C_{i,j}$  is the cost to assign the  $i$ th item to the  $j$ th place or the distance between the  $i$ th and  $j$ th items;  $X_{i,j}$  is 1 if  $i$  is assigned to  $j$  or 0 otherwise. For the cost, we can calculate the distance between two topics by calculating the cosine or Euclidean distance. In this study, the results for the cosine and Euclidean distances are the same because all topic weights are nonnegative.

Greene et al. used the Jaccard distance with the top-ranked key words for several surveys by applying the Hungarian algorithm. Following this, we found one-to-one matching topic pairs for the opinion user group, majority user group, and total user group. Thus, we obtained  $2 \times 10$  one-to-one matching topic pairs, that is,  $\{(t_{1,O}, t_{1,T}), \dots, (t_{10,O}, t_{10,T})\}$  and  $\{(t_{1,M}, t_{1,T}), \dots, (t_{10,M}, t_{10,T})\}$ , where  $t_{i,O}$ ,  $t_{i,M}$ , and  $t_{i,T}$  represent a topic from the opinion user group, majority user group, and total user group, respectively. After finding the one-to-one matching pairs, we calculated the topic similarity for each topic from articles written by opinion leaders and each topic from articles written by majority users by calculating the cosine similarity between the two topic vectors.

Table 3 summarizes the topic similarities and average scores for the opinion leader group and majority user group. Most topics selected by opinion leaders and majority users are similar to topics by entire users. However, two topic similarities from majority users, as highlighted in Table 3, were far lower. These topics are mainly related to mining network updates and questions about the identity of Satoshi Nakamoto (2008) who invented the original version of Bitcoin and the blockchain. In spite of the use of a smaller number of articles (192,292 articles from opinion leaders and 197,344 articles from majority users), the average topic similarity from opinion leaders was higher than that from majority users. Thus, this result suggests that opinion leaders have a greater ability to effectively represent the entire community than majority users.

## Discussion

As presented, we revealed several interesting findings regarding opinion leaders and majority users in the Bitcoin forum. In the Network Analysis subsection in the Experimental Results section, we determined that the authority scores calculated from community networks follow the Pareto distribution. In addition, we investigated users who have the highest authority scores. As described



**Figure 5.** Average topic similarity according to the proportion of users.

earlier, a high authority score indicates that they are likely to get more attention or have a high reputation and communicate with other users actively. In practice, most of these users are Staff, Legendary, or Hero members in the Bitcoin forum (in the Bitcoin forum, there are positions such as staff, legendary, hero member, Sr. member, full member, member, Jr. member, newbie, and brand new for each user based on the number of activities and posts they have written) and active users who simultaneously work within the Bitcoin forum and other cryptocurrency communities. We can presume that they have abundant information about the trends of Bitcoin and other cryptocurrencies and they can transmit this information to this community. According to the results in Table 1, opinion leaders use professional and technical words relatively more frequently than others, whereas majority users use words related to the Bitcoin prices or transactions. It implies that, in reality, opinion leaders are very likely to be experienced and have a deeper understanding of Bitcoin and cryptocurrencies.

In the Topic Modeling subsection given in the Experimental Results, we compared the topics extracted from opinion leaders and those from majority users in the Bitcoin forum. The results of the Pearson correlation demonstrate that majority users tend to focus on the price, transaction, and investment of Bitcoin. On the basis of the topic similarity, we realized that opinion leaders can represent the entire community much more effectively than majority users. Then, can we represent the entire community using a small number of opinion leaders? In other words, would it be more reasonable than selecting users randomly? In addition, we hypothesized that users who account for the top 80% of authority scores are opinion leaders, whereas Zhao et al. (2018) assumed that there are 5% of the members of a social e-commerce community are opinion leaders when conducting simulation experiments. Then, how many users need to be defined as opinion leaders to represent the entire community?

To answer these questions, in a similar manner to the Topic Modeling subsection given in the Experimental Results, we calculated the average topic similarities from opinion leaders and presented them using a line graph according to the percentage of users. We also calculated the average topic similarities from randomly selected users as a baseline model. However, when we randomly select users, the variation in the topic similarity should be high. Therefore, a bootstrapping technique was applied. Bootstrapping is a resampling technique for reducing the statistical variance. Here, we conducted experiments 10 times and used the average results.

Figure 5 shows that the topic similarity of opinion leaders converges to almost 1.0 after using only 7% of users. However, the topic similarity of randomly selected users increases much more slowly than that of opinion leaders and converges to a lower value than that of opinion leaders. Therefore, we can represent the entire community with only a small percentage of opinion leaders. However, at some points, such as ratios of 0.45 and 0.65, the topic similarity slightly decreases. We note that this is because there are several users having different interests even in the same opinion leader group. We leave this problem of various groups existing among opinion leaders for future work.

This study focused on online users in a Bitcoin forum. Most Bitcoin transactions are conducted online and many people seek and share information on a forum. We confirmed that the environments characterized by a large number of users and a large amount of shared information may have opinion leaders, the existence of which facilitates the analysis. Such analyses utilizing opinion leaders are expected to be helpful in various ways in processing information related to Bitcoin. For example, we can analyze the huge amount of information more efficiently. Because we analyze only a small percentage of opinion leaders to analyze the entire community, we only use the posts written by these opinion leaders, and the number of posts is less than half of all posts in the community. In addition, the number of documents targeted for analysis is very important because topic modeling consumes considerable memory and time resources.

The findings confirmed in various Bitcoin forums with unique characteristics are expected to contribute to finding the opinion leaders in communities with various types of environments, identifying their roles, and extracting meaningful information. For example, the identification of opinion leaders in a marketing-oriented forum may maximize the marketing effect by targeting these opinion leaders, as they have more socially influential power than majority users.

## Conclusion

This study identified opinion leaders in a Bitcoin forum and compared them to other majority users by analyzing the topics extracted from posts written over the past 3 years. The results of this study showed that the information delivered by a small number of opinion leaders may represent the opinion of the entire community in large social communities such as the Bitcoin forum where information is exchanged vigorously.

In a future study, as mentioned in the Discussion section, we intend to find the opinion leaders using a more efficient method and divide them into several topic groups to analyze the opinion leaders for each topic group. The analysis of how online communities influence the value of cryptocurrencies and the blockchain is an interesting research problem (Munksgaard & Demant, 2016). Furthermore, we will investigate methods for finding opinion leaders more efficiently not only in the Bitcoin forum but also across other online social communities with different characteristics.

## Authors' Note

The data set generated and analyzed during the current study is available in the repository at <https://tinyurl.com/y6vgycst>. The source code generated and analyzed during the current study is available in the author's Github repository at <https://github.com/rudvlf0413/opinion-leader-analysis>.

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2016R1C1B2015924) and the Chung-Ang University Research Grants in 2018.

## References

- Aalborg, H. A., Molnár, P., & de Vries, J. E. (2018). What can explain the price, volatility and trading volume of Bitcoin? *Finance Research Letters*. doi:10.1016/j.frl.2018.08.010
- Alessandretti, L., ElBahrawy, A., Aiello, L. M., & Baronchelli, A. (2018). Machine learning the cryptocurrency market. *arXiv preprint arXiv:1805.08550*.
- Alvarez-Galvez, J. (2016). Network models of minority opinion spreading: Using agent-based modeling to study possible scenarios of social contagion. *Social Science Computer Review*, 34, 567–581.
- Barber, S., Boyen, X., Shi, E., & Uzun, E. (2012). Bitter to better: How to make Bitcoin a better currency. *Lecture Notes in Computer Science*, 7397, 399–414.
- Bernstein, M. S., Monroy-Hernández, A., Harry, D., André, P., Panovich, K., & Vargas, G. (2011). 4chan and/b: An analysis of anonymity and ephemerality in a large online community. *Paper presented at the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 50–57). Barcelona, Spain: AAAI Press.
- Bickart, B., & Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*, 15, 31–40.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Bohme, R., Christin, N., Edelman, B., & Moore, T. (2015). Bitcoin: Economics, technology, and governance. *Journal of Economic Perspectives*, 29, 213–238.
- Chan, K. K., & Misra, S. (1990). Characteristics of the opinion leader: A new dimension. *Journal of Advertising*, 19, 53–60.
- Chaney, A. J.-B., & Blei, D. M. (2012). Visualizing topic models. *Paper presented at the Sixth International AAAI Conference on Weblogs and Social Media*. Dublin, Ireland: AAAI Press.
- Cheah, E.-T., & Fry, J. (2015). Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics Letters*, 130, 32–36.
- Choi, S. (2015). The two-step flow of communication in Twitter-based public forums. *Social Science Computer Review*, 33, 696–711.
- Choo, J., Lee, C., Reddy, C. K., & Park, H. (2013). Utopian: User-driven topic modeling based on interactive nonnegative matrix factorization. *IEEE Transactions on Visualization and Computer Graphics*, 19, 1992–2001.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41, 391–407.
- Demant, J., Munksgaard, R., Décary-Héty, D., & Aldridge, J. (2018). Going local on a global platform: A critical analysis of the transformative potential of cryptomarkets for organized illicit drug crime. *International Criminal Justice Review*, 28, 255–274.
- Gerlach, J. C., Demos, G., & Sornette, D. (2018). Dissection of Bitcoin's Multiscale Bubble History from January 2012 to February 2018. *arXiv preprint arXiv:1804.06261*.
- Greene, D., O'Callaghan, D., & Cunningham, P. (2014). How many topics? Stability analysis for topic models. In T. Calders, F. Esposito, E. Hüllermeier, & R. Meo (Eds.), *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 498–513). Berlin, Germany: Springer.
- Grinberg, R. (2012). Bitcoin: An innovative alternative digital currency. *Hastings Science and Technology Law Journal*, 4, 159.

- Guo, T., & Antulov-Fantulin, N. (2018). Predicting short-term Bitcoin price fluctuations from buy and sell orders. *arXiv preprint arXiv:1802.04065*.
- Hagberg, A., Swart, P., & Chult, S. D. (2008). Exploring network structure, dynamics, and function using NetworkX (No. LA-UR-08-05495; LA-UR-08-5495). In G. Varoquaux, T. Vaught, & J. Millman (Eds.), Los Alamos, NM: Los Alamos National Lab. Retrieved from: <https://www.osti.gov/servlets/purl/960616>.
- Ho, Y.-C., Liu, H.-M., Hsu, H.-H., Lin, C.-H., Ho, Y.-H., & Chen, L.-J. (2016). Automatic opinion leader recognition in group discussions. In *Technologies and Applications of Artificial Intelligence (TAAI 2016) Conference on* (pp. 138–145). Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Hofmann, T. (1999). Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* (pp. 289–296). Burlington, MA: Morgan Kaufmann.
- Ifrim, G., Shi, B., & Brigadir, I. (2014, April 8). Event detection in Twitter using aggressive filtering and hierarchical tweet clustering. In S. Papadopoulos, D. Corney, & L. Aiello (Eds.), *Second Workshop on Social News on the Web (SNOW)*. Seoul, Korea: ACM.
- Indera, N., Yassin, I., Zabidi, A., & Rizman, Z. (2017). Non-linear autoregressive with exogeneous input (narx) Bitcoin price prediction model using pso-optimized parameters and moving average technical indicators. *Journal of Fundamental and Applied Sciences*, 9, 791–808.
- Jiang, L., Li, F., Ge, B., Xiao, W., Tang, J., & Hu, Y. (2014). Detecting opinion leaders in online communities based on an improved PageRank algorithm. *Applied Mechanics and Materials*, 543, 3524–3527.
- Johnson, S. L., Safadi, H., & Faraj, S. (2015). The emergence of online community leadership. *Information Systems Research*, 26, 165–187.
- Junbo, G., Min, Z., Fan, J., & Xufa, W. (2005, September). Analysis of opinion leader in on-line communities. In R. Khosla, R. J. Howlett, & L. C. Jain (Eds.), *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 1153–1159). Berlin, Heidelberg: Springer.
- Kaminski, J. (2014). Nowcasting the Bitcoin market with Twitter signals. *arXiv preprint arXiv:1406.7577*.
- Karalevicius, V., Degrande, N., & De Weerd, J. (2018). Using sentiment analysis to predict interday Bitcoin price movements *The Journal of Risk Finance* 19 56–75
- Kim, M., Kang, K., Park, D., Choo, J., & Elmqvist, N. (2017). Topiclens: Efficient multi-level visual topic exploration of large-scale document collections. *IEEE Transactions on Visualization and Computer Graphics*, 23, 151–160.
- Kim, Y. B., Kang, K., Choo, J., Kang, S. J., Kim, T., Im, J., ... Kim, C. H. (2017). Predicting the currency market in online gaming via lexicon-based analysis on its online forum. *Complexity*, 2017, 4152705.
- Kim, Y. B., Kim, J. G., Kim, W., Im, J. H., Kim, T. H., Kang, S. J., & Kim, C. H. (2016). Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PLoS ONE*, 11, e0161197.
- Kim, Y. B., Lee, J., Park, N., Choo, J., Kim, J.-H., & Kim, C. H. (2017). When Bitcoin encounters information in an online forum: Using text mining to analyse user opinions and predict value fluctuation. *PLoS ONE*, 12, e0177630.
- Kim, Y. B., Lee, S. H., Kang, S. J., Choi, M. J., Lee, J., & Kim, C. H. (2015). Virtual world currency value fluctuation prediction system based on user sentiment analysis. *PLoS ONE*, 10, e0132944.
- Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46, 604–632.
- Koch, R. (2011). *The 80/20 Principle: The secret of achieving more with less: Updated 20th anniversary edition of the productivity and business classic*. Hachette England: Nicholas Brealey Publishing.
- Koch, R. (2018). *Living the 80/20 way*. Central Jakarta, Indonesia: Gramedia Pustaka Utama.
- Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports*, 3, 3415.
- Kuang, D., Choo, J., & Park, H. (2015). Nonnegative matrix factorization for interactive topic modeling and document clustering. In M. Celebi (Ed.), *Partitional Clustering Algorithms* (pp. 215–243). Berlin, Germany: Springer.
- Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401, 788.

- Lee, D. D., & Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In: T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in Neural Information Processing Systems* (pp. 556–562). Denver, USA: MIT Press
- Li, T. R., Chamrajnagar, A. S., Fong, X. R., Rizik, N. R., & Fu, F. (2018). Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. *arXiv preprint arXiv:1805.00558*.
- Li, X., & Wang, C. A. (2017). The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin. *Decision Support Systems*, 95, 49–60.
- Lipovetsky, S. (2009). Pareto 80/20 law: Derivation via random partitioning. *International Journal of Mathematical Education in Science and Technology*, 40, 271–277.
- Loper, E., & Bird, S. (2002). NLTK: The natural language toolkit. *arXiv preprint arXiv:cs/0205028*.
- Mai, F., Shan, Z., Bai, Q., Wang, X., & Chiang, R. H. (2018). How does social media impact Bitcoin value? A test of the silent majority hypothesis. *Journal of Management Information Systems*, 35, 19–52.
- Maier, D., Waldherr, A., Miltner, P., Jähnichen, P., & Pfetsch, B. (2018). Exploring issues in a networked public sphere: Combining hyperlink network analysis and topic modeling. *Social Science Computer Review*, 36, 3–20.
- Marella, V., Lindman, J., Rossi, M., & Tuunainen, V. (2017). Bitcoin: A social movement under attack. *Selected Papers of the IRIS, Issue Nr 8, 1*. <https://aisel.aisnet.org/iris2017/1>
- Matta, M., Lunesu, I., & Marchesi, M. (2015). Bitcoin spread prediction using social and web search media. In A. Cristea, J. Masthoff, A. Said, & N. Tintarev (Eds.), *Posters, Demos, Late-breaking Results and Workshop Proceedings of the 23rd Conference on User Modeling, Adaptation, and Personalization* (pp. 1–10). Dublin, Ireland: Sun SITE Central Europe.
- McNally, S., Roche, J., & Caton, S. (2018). Predicting the price of Bitcoin using machine learning. In *Parallel, Distributed and Network-based Processing (PDP), 2018 26th Euromicro International Conference on* (pp. 339–343). Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Ming Yu, C. (2002). Socialising knowledge management: The influence of the opinion leader. *Journal of Knowledge Management Practice*, 3, 76–83.
- Moeller, K., Munksgaard, R., & Demant, J. (2017). Flow my FE the vendor said: Exploring violent and fraudulent resource exchanges on cryptomarkets for illicit drugs. *American Behavioral Scientist*, 61, 1427–1450.
- Munksgaard, R., & Demant, J. (2016). Mixing politics and crime—The prevalence and decline of political discourse on the cryptomarket. *International Journal of Drug Policy*, 35, 77–83.
- Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. Retrived from Bitcoin Project Website: <https://bitcoin.org/bitcoin.pdf>
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*. Technical report SIDL-WP-1999-0120, Stanford Digital Library Technologies Project.
- Panzarasa, P., Opsahl, T., & Carley, K. M. (2009). Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. *Journal of the American Society for Information Science and Technology*, 60, 911–932.
- Parmenter, D. (2007). *Pareto's 80/20 rule for corporate accountants*. Hoboken, NJ: John Wiley.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12, 2825–2830.
- Phillips, R. C., & Gorse, D. (2017, November). Predicting cryptocurrency price bubbles using social media data and epidemic modelling. In *Computational Intelligence (SSCI 2017) IEEE Symposium Series on* (pp. 1–7). Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Phillips, R. C., & Gorse, D. (2018). Cryptocurrency price drivers: Wavelet coherence analysis revisited. *PLoS One*, 13, e0195200.



- Reid, F., & Harrigan, M. (2013). An analysis of anonymity in the Bitcoin system. In Y. Altshuler, Y. Elovici, A. Cremers, N. Aharony, & A. Pentland (Eds.), *Security and Privacy in Social Networks* (pp. 197–223). Berlin, Germany: Springer.
- Rogers, E. M. (2010). *Diffusion of innovations*. New York, NY: Simon & Schuster.
- Sin, E., & Wang, L. (2017). Bitcoin price prediction using ensembles of neural networks. In *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (pp. 666–671). Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Waldherr, A., Maier, D., Miltner, P., & Günther, E. (2017). Big data, big noise: The challenge of finding issue networks on the web. *Social Science Computer Review*, 35, 427–443.
- Walt, S. V. D., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, 13, 22–30.
- Wang, C., Du, Y. J., & Tang, M. W. (2016). Opinion leader mining algorithm in microblog platform based on topic similarity. In *Computer and Communications (ICCC 2016) 2nd IEEE International Conference on* (pp. 160–165). Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum Project Yellow Paper*, 151, 1–32.
- Wu, C., Li, C., Yan, W., Luo, Y., Mao, X., Du, S., & Li, M. (2015). Identifying opinion leader in the Internet forum. *International Journal of Hybrid Information Technology*, 8, 423–434.
- Yelowitz, A., & Wilson, M. (2015). Characteristics of Bitcoin users: An analysis of Google search data. *Applied Economics Letters*, 22, 1030–1036.
- Zhang, W., Wang, P., Li, X., & Shen, D. (2018). Quantifying the cross-correlations between online searches and Bitcoin market. *Physica A: Statistical Mechanics and its Applications*, 509, 657–672.
- Zhao, Y., Kou, G., Peng, Y., & Chen, Y. (2018). Understanding influence power of opinion leaders in e-commerce networks: An opinion dynamics theory perspective. *Information Sciences*, 426, 131–147.

## Author Biographies

**Kyeongpil Kang** is a doctoral student in the Department of Computer Science and Engineering at Korea University. His research focuses on data mining, machine learning, and social mining. Email: rudvlf0413@korea.ac.kr

**Jaegul Choo** is an assistant professor in the Department of Computer Science and Engineering at Korea University. He was a research scientist at Georgia Tech from 2011 to 2015, where he also received an MS in 2009 and a PhD in 2013. His research focuses on visual analytics for machine learning and deep learning. Email: jchoo@korea.ac.kr

**YoungBin Kim** is an assistant professor in the Graduate School of Advanced Imaging Science, Multimedia & Film at Chung-Ang University. He earned his MS in 2012 and PhD in 2017 from Korea University. His research focuses on social media analysis, text mining, and deep-learning-based information processing. Email: ybkim85@cau.ac.kr