Effects Of Communication Technology On Cryptocurrency

Amr Elhefnawy Student ID: 30062701 CPSC 572
Mohamed Elhefnawy Student ID: 30093687 CPSC 572
Abdelhak Khalfallah Student ID: 30081858 CPSC 572

Project summary

In the modern world where social media apps garner and stockpile people's time and attention [1], A person who can influence users on these social apps wields the power to influence trends for particular products from the palm of their hands [2]. When it comes to investments, digital currency, and specifically bitcoin is argued to be the future of finance [3]. And when such an important asset is being discussed on these social media apps, it is crucial to know what the influential people are sharing and how they feel about bitcoin since they have the ability to influence the masses in either direction. In this research, therefore, we examine Twitter as a network of influencers that made bitcoin-related tweets with the power to diffuse their own thoughts and sentiments about bitcoin to their followers[4]. We conducted a social network analysis of our users and their tweets which lets us examine how influential they are and what their sentiment is and therefore by what capacity and direction they might influence the general consensus about bitcoin. The findings of this study show us that communication technology such as twitter and the influential people using this technology contribute to the reputation of digital currencies in a positive way. These findings are important because they show us how the use of global communication platforms might help normalize the idea of a future built on digital currencies.

Research questions

1) Who can affect the decision-making process of people using communication technology the most with regard to investing in bitcoin?
    - We intend to answer this question by finding who the most influential people are in our network of verified users. We accomplish this by finding the top nodes in our network based on 3 different centrality measures, indegree, outdegree, and betweenness centrality. By using these three centrality measures, we can find the most influential users based on how many times they are mentioned, how many times they mention others, and how fast they can propagate information through the network. By finding out this information we identify the people in our network that can have the most influence on people's decision on whether to invest in bitcoin or not.

2) In which direction does communication technology as a whole affect people's

decision-making process toward investing in bitcoin?

- We intend to answer this question by extracting and examining the sentiment of the whole network, the most influential users, and the different communities. Since our network is made up of influential people with bitcoin-related tweets, the sentiment of the influencers may influence their followers and therefore influence their followers' decision on whether to invest in bitcoin or not.

Note: We realized that the research questions we had in the presentation did not represent ‘big-picture goals’. They were more like sub-questions that we asked to help us answer our 2 research questions above. That’s why we re-worded the main questions in such a way that they are still consistent with what we did in our project and our goals.

Introduction

With the rise of social media technologies and the accelerating number of people that use them, along with the fact that social media platforms have been a key piece for the dissemination of information, is a reason why people find themselves attached to social media more than ever before. With this attachment, regular people are more prone to be influenced by social media influencers and famous figures that use these apps. For that reason, it is safe to say that people can get influenced by these social apps more than ever before. Twitter is an online social networking and micro-blogging site and app in which users write short messages that exceed not more than 140 characters called “tweets”. It is a global forum with the presence of eminent personalities from the field of entertainment, industry, and politics.

When it comes to investments, digital currency, and specifically bitcoin, is argued to be the future of finance. And when such an important asset is being discussed on these social media apps, it is crucial to know what the influential people are sharing and how they feel about bitcoin since it has the ability to influence the masses in either direction. In this research, we address two research questions: Who can affect the price of bitcoin the most using communication technology, and in which direction does communication technology affect the price of bitcoin? To answer these questions, we conduct a social network analysis of the dataset we obtained which contains 8827 tweets from verified users (or 1% of all bitcoin-related tweets) made between 6/2/2021 and 10/23/22 (the date we downloaded the dataset). These tweets are marked with #BTC or #bitcoin in their textbase. A hashtag is a word or keyword phrase preceded by a hash symbol (#). It's used within a post on social media to help those who may be interested in your topic to be able to find it when they search for a keyword or particular hashtag. It is believed to help draw attention to a user's posts and encourage interaction.

These 8827 tweets represent non-unique tweets. Since we always just take the first interaction between two users, these tweets would automatically be filtered to give us a unique number of nodes (users) and edges (mentions). The 8827 tweets were made by 3962 unique users and contain 4104 unique mentions where the user mentioning (source node) is always a verified user and the user being mentioned (target node) is not necessarily verified. These tweets represent the interactions (mentions) between verified Twitter users talking about bitcoin and other verified or unverified users who are being mentioned by them. We define our nodes this way since tweets from verified users are usually the influential ones, and anyone being mentioned (verified or not) can be influenced by such a tweet. We conducted this social network analysis of bitcoin-related tweets from verified users to see how these social apps might influence someone's investment decision toward digital currency and more specifically, bitcoin.

Dataset description

Raw data

user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date	text	hashtags	source	is_retweet
DeSota Wilson	Atlanta, GA	Biz Consultant, real (	39929.8369097222	8534	7605	4838	FALSE	44237.9993518519	Blue Ridge Bank sha	['bitcoin']	Twitter Web App	FALSE
CryptoND		ðŸ˜Ž BITCOINLIVE is	43755.8417824074	6769	1532	25483	FALSE	44237.9991666667	ðŸ˜Ž Today, that's th	['Thursday', 'Btc', 'w	Twitter for Android	FALSE
Tdlmatias	London, England	IM Academy : The be	41953.4518171296	128	332	924	FALSE	44237.9963888889	Guys evening, I have		Twitter Web App	FALSE
Crypto is the future		I will post a lot of bu	43736.7001388889	625	129	14	FALSE	44237.9962152778	$BTC A big chance ir	['Bitcoin', 'FX', 'BTC',	dlvr.it	FALSE
Alex Kirchmaier ðŸ‡¦	Europa	Co-founder	42403.5527199074	1249	1472	10482	FALSE	44237.9959027778	This network is secu	['BTC']	Twitter Web App	FALSE
ZerrBenzâ„¢ âš” âœ	Bkk, Thailand	I'm a cat slave ðŸ± Ir	40190.291712963	742	716	2444	FALSE	44237.9954861111	ðŸ’¹ Trade #Crypto	['Crypto', 'Binance', '	Twitter Web App	FALSE
Bitcoin-Bot	Florida, USA	Bot to generate Bitc	43822.7008796296	131	84	5728	FALSE	44237.9953356481	<'fire' &	['Bitcoin', 'Crypto', 'E	BTC_p_bot	FALSE
Cryptocurrencies / E		Stay updated on the	43315.8959259259	4052	1	9	FALSE	44237.9949305556	ðŸ”„ Prices update		Cryptocurrencies pric	FALSE
Mikcoin		Technical Analyst |	44161.9901157407	104	41	238	FALSE	44237.9947337963	#BTC #Bitcoin	['BTC', 'Bitcoin', 'Eth	Twitter Web App	FALSE
DeSota Wilson	Atlanta, GA	Biz Consultant, real (	39929.8369097222	8534	7605	4838	FALSE	44237.994537037	.@Teslaâ€™s #bitcoi	['bitcoin', 'crypto']	Twitter Web App	FALSE
@massumeh18 #Ref	NOVA	Persistent. to the ext	39818.5616550926	1159	2185	30852	FALSE	44237.9944907407	Annnd #btc #Bitcoin	['btc', 'Bitcoin']	Twitter Web App	FALSE
BittrexPrices		Scans Bittrexâ€™s m	43245.8542476852	3131	1	18	FALSE	44237.9940393519	ðŸ”„ Prices update		Bittrex Prices	FALSE
CPUcoin	Cayman Islands	The Sharing Econom	43339.6541666667	5097	791	52	FALSE	44237.9937384259	Join our first virtual		Twitter Web App	FALSE
One Perspective	Market Place	It's all a market of	43109.9095486111	668	1097	6853	FALSE	44237.9934027778	#Bitcoin #BTC $BTC	['Bitcoin', 'BTC', 'AA	Twitter for Android	FALSE
CryptoSquawk	Australia	24x7 Crypto market r	43013.4257986111	1281	25	72	FALSE	44237.9930787037	â¬†ï¸â¬†ï¸ $BTC	['Bitcoin', 'crypto', 'B	CryptoSquawkBot	FALSE
$MOON	Moon	#Bitcoin	40371.3803472222	4	32	139	FALSE	44237.9921064815	Buy #Bitcoin with 5%	['Bitcoin', 'cryptocurr	Twitter Web App	FALSE

Figure 1 Raw dataset

The raw data set was taken from a website called Kaggle where data scientists find and publish datasets. The raw data set was available for download on the Kaggle website without any restriction that could limit the use of the data. The dataset contains tweets from 6/2/2021 to 10/23/2022 (the date we obtained the dataset), Figure 1 represents the raw data acquired from Kaggle it held the username of the user who created the tweet, user location, user description, date the user was created, user followers, user friends, if the user is verified, date of the tweet, the text of the tweet, the hashtags in the tweet, the source of the tweet (Utility used to post the Tweet), and if the tweet was retweeted or not.

Data cleaning and data wrangling

user_name	user_created	user_followers	user_friends	user_favourites	user_verified	date	text	cleaned_text	mentioned_users
Alex JimÃ©nez	41024.64367	14469	822	40097	TRUE	44237.98962	#Bitcoinâ€™s Rally	bitcoin ralli signal rise	[]
Royalty	39940.258	31548	2734	26340	TRUE	44237.93615	If Apple Inc ever ma	appl inc ever make mo	[]
Telegraph Technology Intelligence	39842.59073	82628	239	1	TRUE	44237.90271	.@LFDodds has	lfdodd twitter say coul	['@LFDodds']
IGSquawk	41445.38501	49690	655	1667	TRUE	44237.86412	Crypto update:	crypto updat bitcoin et	[]
Alex JimÃ©nez	41024.64367	14469	822	40097	TRUE	44237.78131	#Bitcoinâ€™s Rally	bitcoin ralli signal rise	[]
Blockstream	41663.17059	131970	794	12454	TRUE	44237.752	#BlockstreamAQUA	blockstreamaqua new	['@Liquid_BTC']
Christophe BarraudðŸ›¢	40994.41968	80621	198	687	TRUE	44237.74609	âš #Crypto Brief âš	crypto brief bitcoin eth	[]
Telegraph Technology Intelligence	39842.59073	82628	239	1	TRUE	44237.64781	â€¢ Twitter	twitter consid invest b	[]
CNBC	39853.00256	3977144	843	612	TRUE	44237.64164	"It has to be part of t	part menu say jimcram	['@JimCramer']
Squawk Box	40850.92859	283910	831	1908	TRUE	44237.59196	"It has to be part of t	part menu say jimcram	['@JimCramer']
eToroX	43320.52307	21784	106	2125	TRUE	44237.58334	CEO of @Grayscale (	ceo grayscal sonnensh	['@Grayscale', '@Sonnenshein', '@SquawkCNBC']
Squawk Box	40850.92859	283910	831	1908	TRUE	44237.58029	"We've done a lot of	done lot upfront think	[]
CoinMarketCap	41632.70153	731514	1912	4486	TRUE	44237.53128	This week, @intoth	thi week intotheblock	['@intotheblock']
IGSquawk	41445.38501	49690	655	1667	TRUE	44237.52035	Crypto update:	crypto updat bitcoin et	[]
Maggie Kate Fitzgerald	42011.09839	2829	2044	2249	TRUE	44237.51534	Ark Invest believes T	ark invest believ tesla	[]

Figure 2 Cleaned dataset

The raw data set held over 800000+ tweets with the hashtags BTC and bitcoin. We filtered the dataset using python pandas to show only verified users as they were the study's primary research targets. This filtration concluded in 8827 tweets which were calculated to be 1% of the original raw data set. Our next step was to filter each tweet's text body to find the users being mentioned. Mentions are the act of citing or calling attention to someone within a tweet. Mentions are identified by the @ symbol followed by the name of a user. We search through each tweet's text body and collect all mentions within a tweet. We extract the mentioned users from the tweet's text and add them to a new column that contains the mentioned users for each tweet. We do this to define the edge list for our network where the source of each edge is the username corresponding to the tweet containing a mention and the target of each edge is the username corresponding to the user being mentioned within that tweet. We also create a function to clean the text of the tweets by removing stopwords, stemming the words, removing a word if its length is less than 3, removing a word if it's a space, and removing a word if there are digits in it. This function then returns a tokenized sentence with cleaned words only and the mentioned users within a tweet. This is done to help us apply sentiment analysis on each tweet. We then converted our cleaned dataset to a nodes list and an edge list which was used to create our network.

Clear definitions of nodes and edges, and what type of network you have

Nodes: verified user accounts that made bitcoin-related tweets.
Links: user accounts that have been mentioned within those tweets.

The nodes in our nodes list are defined as verified users who made bitcoin-related tweets and users being mentioned in those bitcoin-related tweets. Our edge list is defined as a list of source nodes and target nodes where the source of each edge will contain the data in the ["user_name"] column of our dataset and the target of each edge will contain the data in the ["user_mentions"] column. That way if a user mentioned another user in their tweet then there's an edge between them and the edge contains the cleaned tweet's text. We have a directed network since we need to differentiate between the users mentioning other users and the users being mentioned.

Basic statistics

Number of nodes and edges

The number of nodes in our network is 3,962 (users) and the number of edges is 4104 (mentions). This is because some tweets don't have mentions, we didn't exclude those from our dataset because these people can be mentioned by other people in the dataset.

Number of connected components

In the constructed network we have 3962 Strongly Connected Components and 483 Weakly Connected Components. The number of strongly connected components is telling us that we don't have nodes that are 2-way connected. Meaning, we don't have a case where user1 mentions user2, and user2 also mentions user1.

Degree distribution

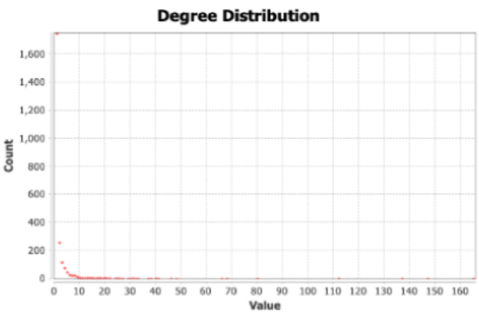


Figure 3

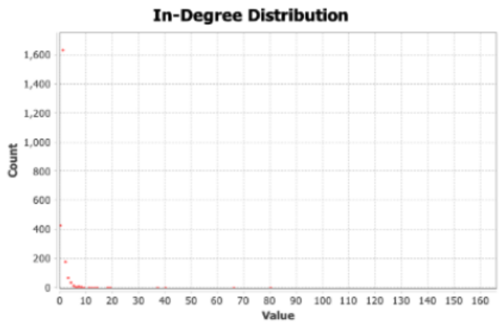


Figure 4

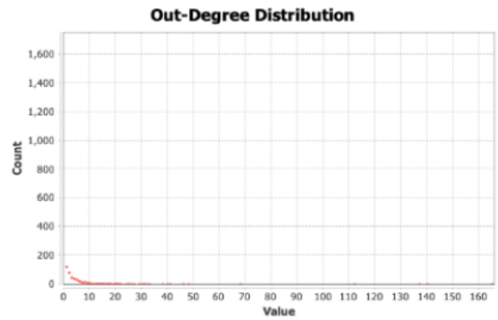


Figure 5

Figure 3 shows us that our degree distribution is left-skewed which means that it follows the real scale-free network structure since it has a power-law distribution. This also tells us that we observe preferential attachment in our network. Figure 4 shows our network's indegree distribution. This is telling us that most of our users (1,600+) were mentioned once and only a few users were mentioned multiple times. Figure 5 shows our network's outdegree distribution. This is telling us that most of our users mention 0-3 people. And it starts decreasing as the number of mentions increases. We only have one node that mentioned 160+ users.

Clustering coefficient

Our average clustering coefficient is 0.0020439531270307724 which tells us that most of the node's neighbors in our network are not connected.

Path length

The average path length in our network is 1.289. This tells us that every 1.289 steps, we reach another node in our network which shows us that our network follows small-world properties.

Network visualization

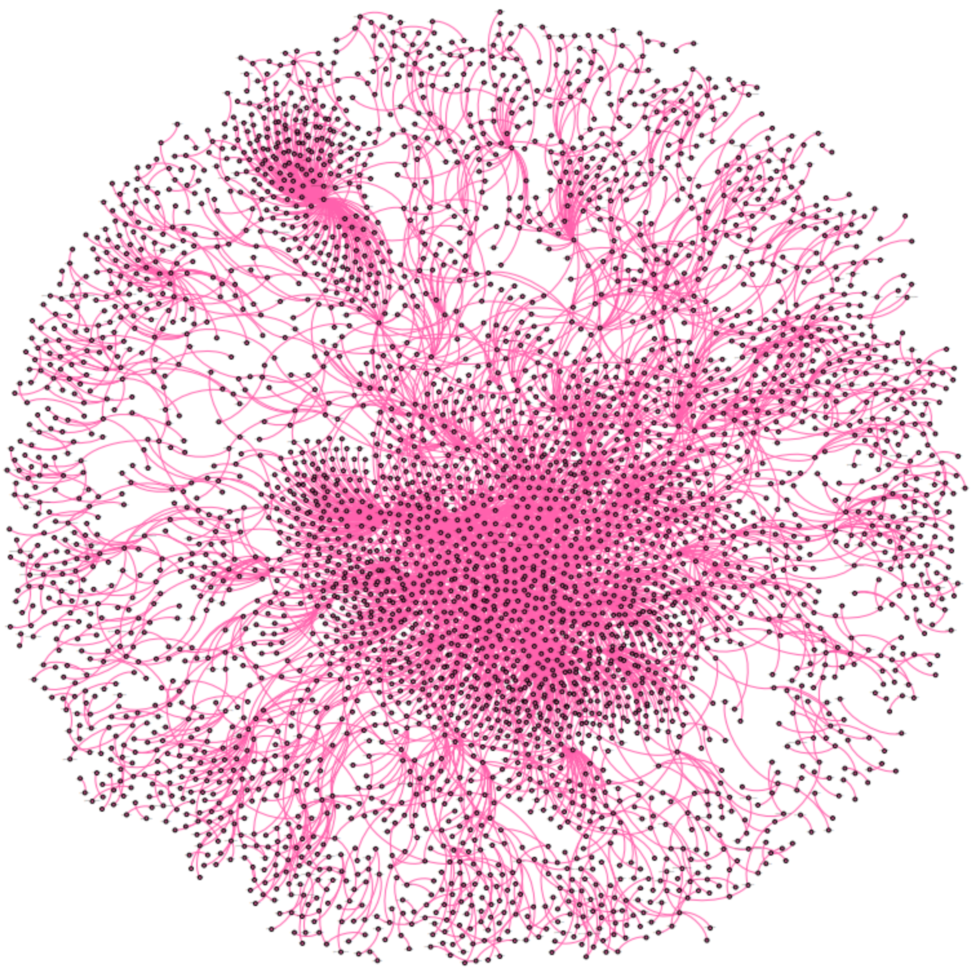


Figure 6

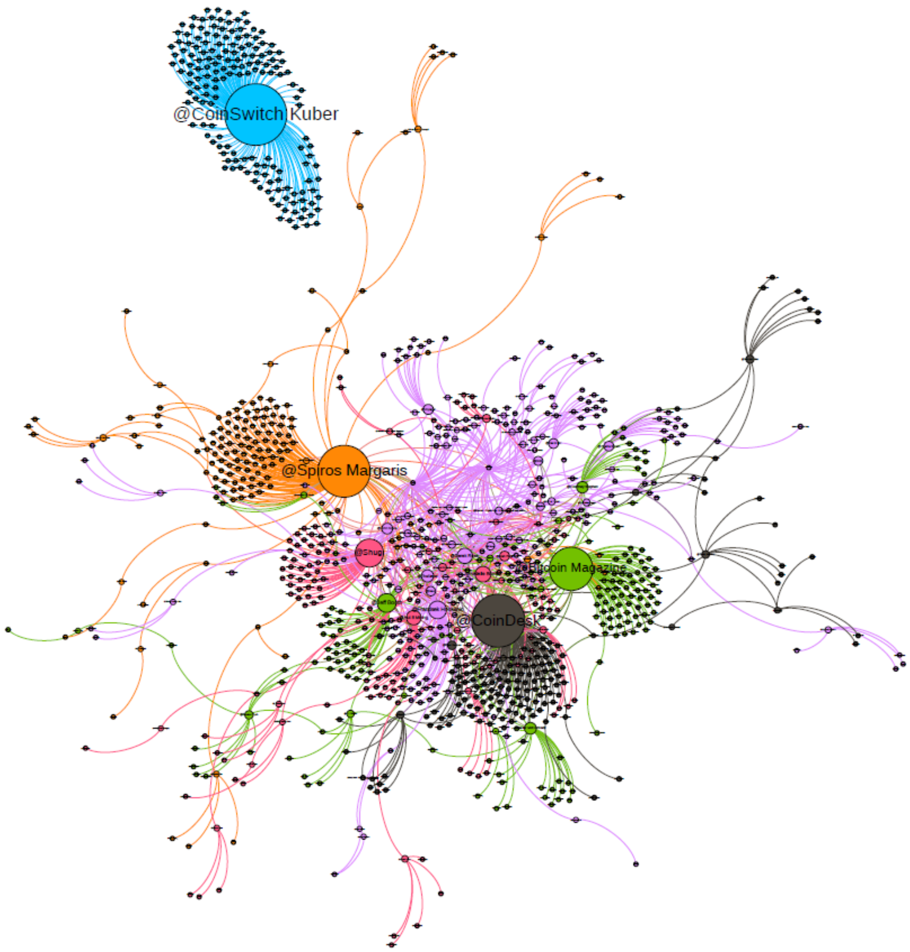


Figure 7

Figure 7 shows us the outdegree-based visualization. Meaning, these are the nodes that mention the most users in our network. The bigger a node is, the more users it mentions.

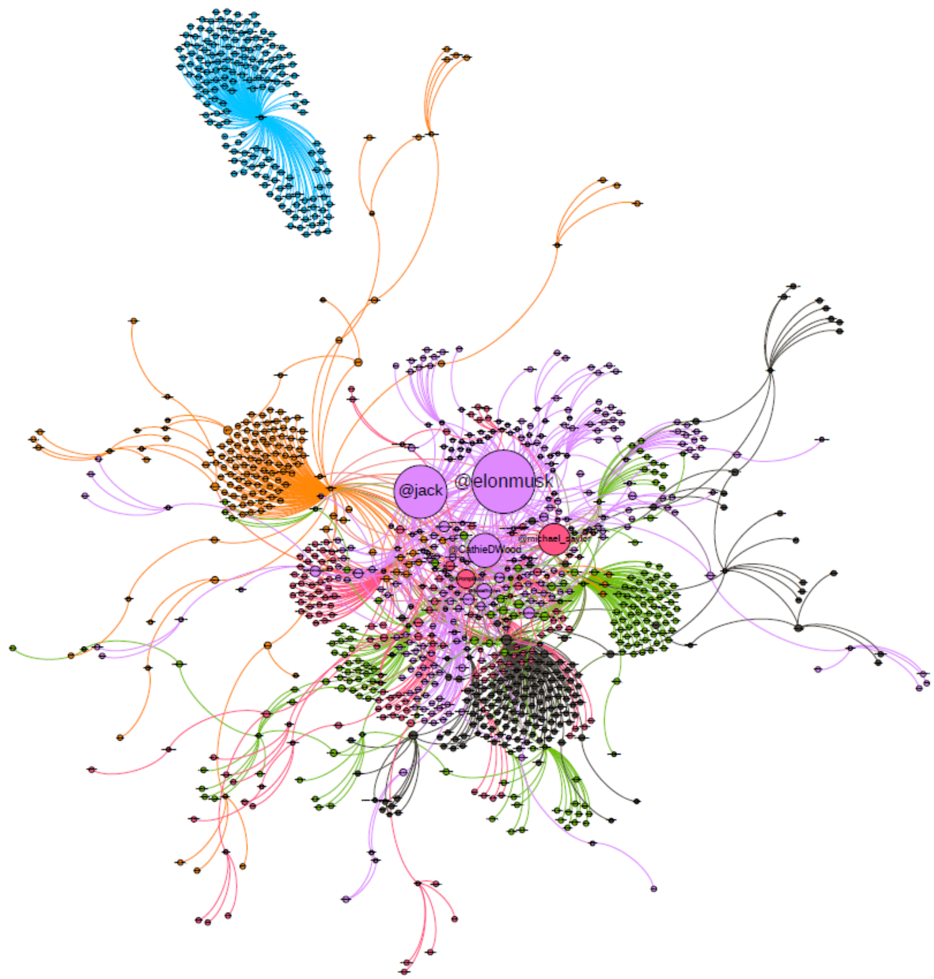


Figure 8

Figure 8 shows us the indegree-based visualization. Meaning, these are the nodes that are mentioned the most in our network. The bigger a node is, the more times it's mentioned.

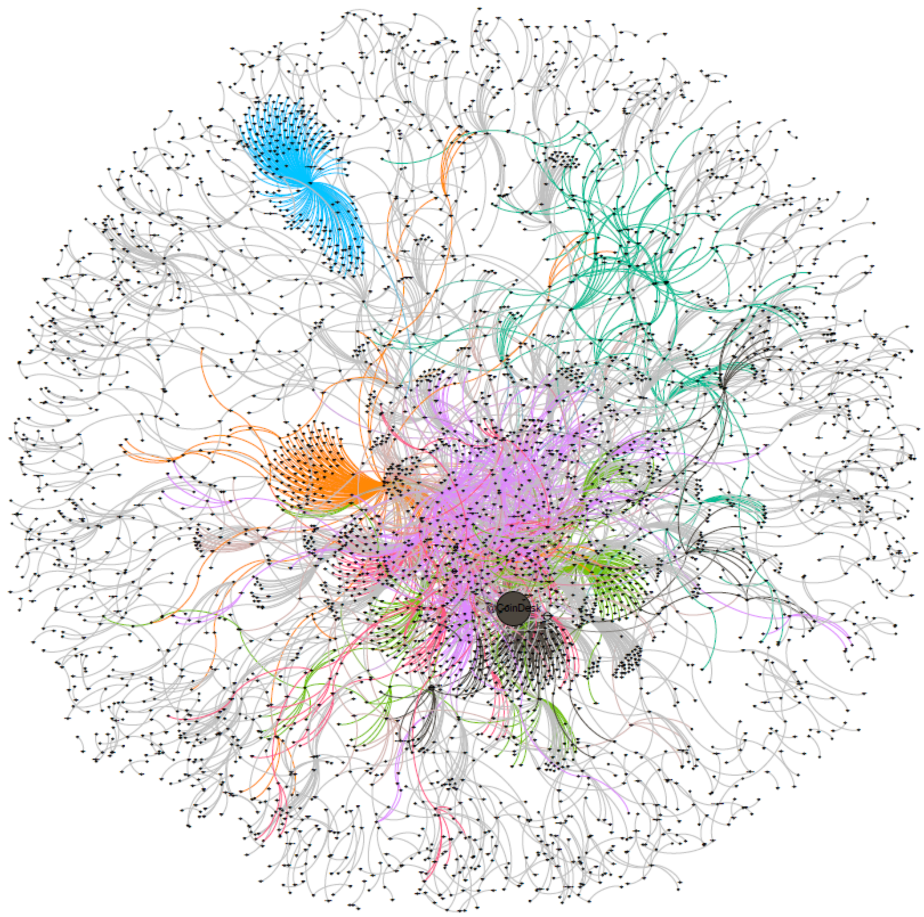


Figure 9

Figure 9 shows us the betweenness centrality-based visualization. Meaning, these are the nodes that propagate information the fastest in our network. The bigger a node is, the faster it propagates information.

Results

Our first research question was: Who can affect people's decision-making process the most with regard to investing in bitcoin? We answered this question by finding the most influential verified users in our network. We find the most influential verified users by finding the top nodes in our network based on 3 different centrality measures: indegree, outdegree, and betweenness centrality. By using these three centrality measures, we can find the most influential users based on how many times they are mentioned, how many times they mention others, and how fast they can propagate information through the network. These three are considered as influence measures because they either: 1) Show how many verified users are trying to mention you and grab your attention, 2) Show how many people you are mentioning and thus potentially changing their sentiment, 3) Show us how fast you can propagate information and thus reach the most amount people the fastest. By finding out this information we identify the people in our network that might have the most influence on people's decision on whether to invest in bitcoin or not.

node	Indegree
@elonmusk	80
@jack	66
@CathieDWood	40
@michael_saylor	37
@APompliano	19
@Olympics	19
@Tesla	18
@Tokyo2020	18
@WeAreTeamIndia	14
@BustaRhymes	14
@MicroStrategy	13
@CNBC	12
@Grayscale	11
@naomiosaka	11
@novogratz	9
@nayibbukele	9
@MikeTyson	9
@ianuragthakur	8
@SpaceX	8
@Bitcoin	8

Figure 10

node	Outdegree
@CoinSwitch Kuber	165
@CoinDesk	140
@Spiros Margaris	137
@Bitcoin Magazine	112
@Shug	68
@Herbert R. Sim	48
@Peter Schiff	46
@Jeff Booth	41
@Jay Gould	40
@Frantisek Hrinkanic	38
@Tiffany Hayden	33
@Natalie Brunell	32
@Squawk Box	31
@Derek Ross	31
@David Gokhshtein	30
@Hut 8 Mining	29
@Mark Jeffrey ⚡🚀	27
@sportsfile	26
@The Exchange	25
@Real Vision	25

Figure 11

node	Betweeness centrality
@CoinDesk	1052.5
@Grayscale	180.0
@BlockFi	103.0
@TheStreet	76.5
@CNBC	70.0
@Blockstream	43.0
@Benzinga	33.0
@CoinMarketCap	23.0
@Binance	9.0
@reason	6.0
@Bitstamp	6.0
@Hedgeye	5.0
@Gemini	4.0
@Ledger	3.0
@DailyFX	1.0
@CryptoCompare	1.0
@Telegraph Technology Intelligence	0.0
@LFDodds	0.0
@Liquid_BTC	0.0
@JimCramer	0.0

Figure 12

Figure 9 is the top 20 users based on indegree, which means these are the top 20 users being mentioned the most in our network. Figure 10 is the top 20 users based on outdegree which means these are the users mentioning the most people in our network. Figure 11 is the top 20 users based on the highest betweenness centrality. This means these are the users that can propagate information the fastest in our network. Using these 3 centrality measures we were able to identify the most influential people in our network and so these are the people who can affect people's decision-making process the most with regard to investing in bitcoin.

Our second research question was: In which direction does communication technology as a whole affect people's decision-making process toward investing in bitcoin? We answered this question by extracting and examining the sentiment of all available tweets, the most influential users, and the different communities. Since our network is made up of influential people with bitcoin-related tweets, the sentiment of the influencers may influence their followers and therefore influence their followers' decision on whether to invest in bitcoin or not.

Sentiment based on all available tweets

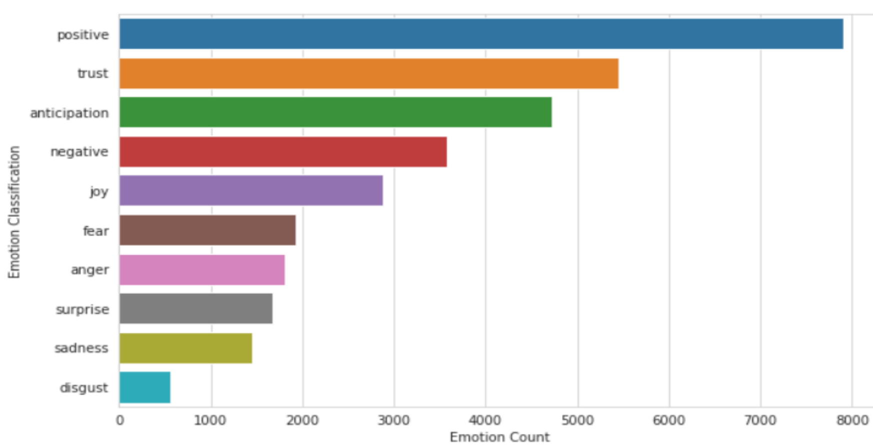


Figure 13

Firstly, by applying sentiment analysis to all available tweets in our network, we found that the overall sentiment toward bitcoin was mostly positive. So, If a person reads tweets from verified users there's a higher chance that that tweet is of a positive sentiment since most of the words used when talking about bitcoin in our network are positive. This shows us that people who read the tweet are more likely to be influenced in a positive way which may increase the chances of them investing in bitcoin.

Secondly, we got the sentiment of the most influential users in our network. We defined the most influential users based on 3 centrality measures: 1) indegree-based top users, which are the verified users being mentioned the most. 2) Outdegree-based top users, which are the verified users mentioning the most people. 3) Betweenness centrality-based top users, which are the verified users who are propagating information the fastest in our network. We get the sentiment of the tweets relating to the top 20 nodes for all 3 measures and combine them to find out the overall sentiment of all influential users.

Top 20 indegree-based user sentiment

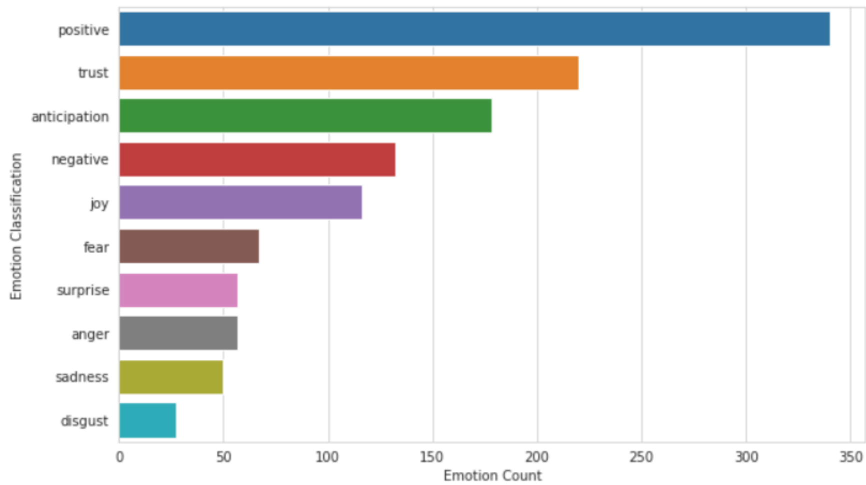


Figure 14

By getting the sentiment of the top 20 users based on indegree, we know that the influential users being mentioned the most are being mentioned in tweets that have a mostly positive sentiment towards bitcoin. There is a possibility that those influential users may be affected by the sentiments of the people mentioning them since the people mentioning them are verified users as well. Now, when that influential person makes a tweet about bitcoin with their new sentiment, they may influence their followers to have that same sentiment.

Top 20 outdegree-based user sentiment

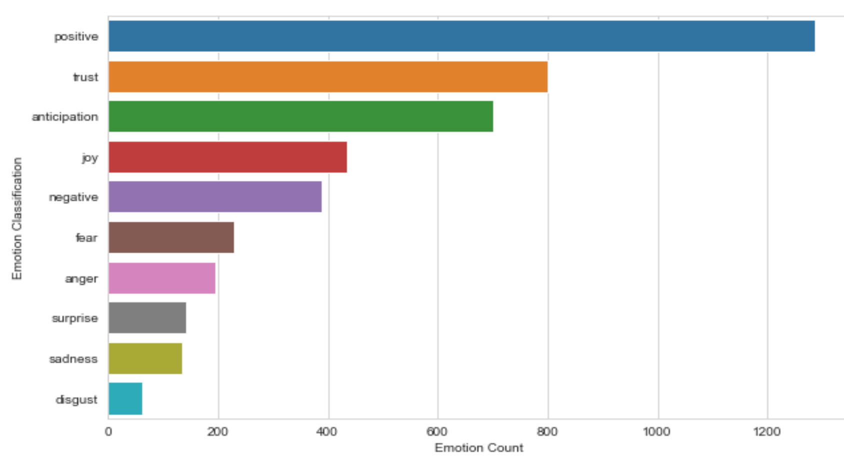


Figure 15

By getting the sentiment of the top 20 users based on outdegree, we know that the verified users that are mentioning the most people, are mentioning them in a mostly positive sentiment. So the people being mentioned are more likely to have a positive sentiment as well and in turn, may be more inclined to invest in bitcoin.

Top 20 betweenness centrality-based user sentiment

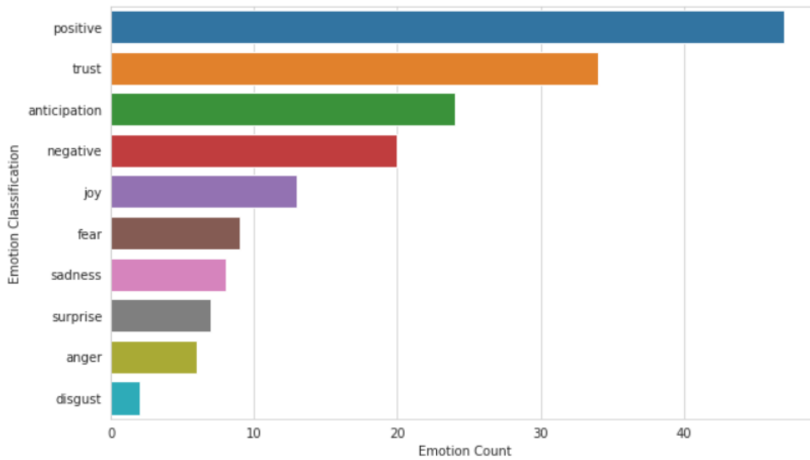


Figure 16

By applying sentiment analysis on the top 20 users that have the highest betweenness centrality, we find that the verified users that can propagate information the fastest in our network are doing so in a mostly positive sentiment as well. This means that the people receiving the information propagated by those verified users may be influenced in a positive way as well.

By examining the sentiments of the tweets made by the top 20 outdegree nodes and the top 20 betweenness nodes, we can see that they are always mostly positive. This means that all of the most influential verified users are sharing information about bitcoin in a mostly positive way.

By examining the sentiments of the tweets mentioning the top 20 indegree nodes, we can see that they are also mostly positive. This means that all of the most influential verified users being mentioned by tweets from other verified users may be reading these tweets about bitcoin which have a mostly positive sentiment. If these tweets do affect their decision-making process toward bitcoin and make a tweet with that new sentiment, then it will most likely also affect their followers' decision-making process toward purchasing bitcoin in a positive way**.**

This shows us that the most influential people using this communication technology platform and sharing bitcoin-related information may affect people's decision-making process toward investing in bitcoin **in a positive way**.

Thirdly, by using community detection (Cdlib rb_pots algorithm) we can find the groups of people interacting with each other the most. Then we get the sentiment of each community and based on that, we assume that if people are part of a community that has a certain sentiment then those people may be influenced to share the same sentiment as the community they're in. This may influence their decision on whether to invest in bitcoin or not. Ex: if I'm a part of a community that has a mostly positive sentiment towards bitcoin then I am more likely to be influenced to think about bitcoin in a positive way.

Figure 17

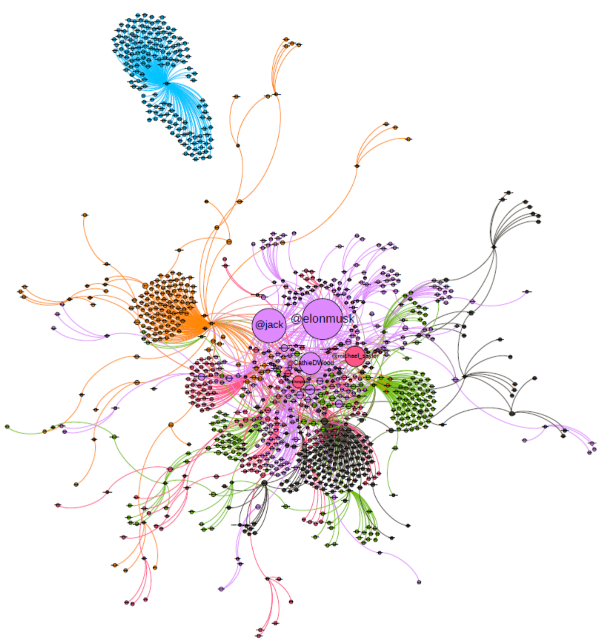


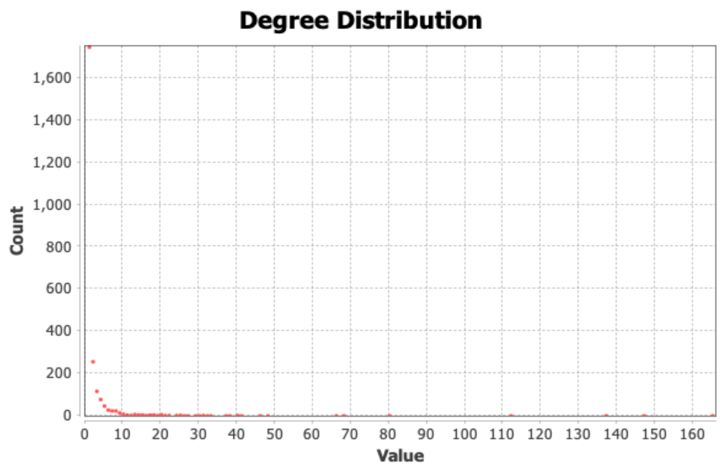
Figure 18

Figure 19

Figure 17 tells us the community ID, the corresponding color for each community, and the percentage of users in our network that are a part of each community. We only take these communities into account since they're the most important and we disregard the rest of the communities as they don't give us useful information. In community ID 0 (purple cluster) we can see the top users being mentioned and the positive and negative words being used. The bigger a word is the more times it's being used in this community. Influencers can use the information from the word clouds to target communities that have a more positive sentiment towards bitcoin and use the positive words being used to influence their followers to think about bitcoin in a positive way (pump and dump crypto scams). Similarly, if someone wants to influence their followers to think about bitcoin in a negative way, they can look at the negative words being used and construct a tweet to influence their followers to think about bitcoin in a negative way. Thus, this information can be used to help influencers play a role in people's decision-making process when investing in bitcoin.

Comparison to a suitable null model

Our network

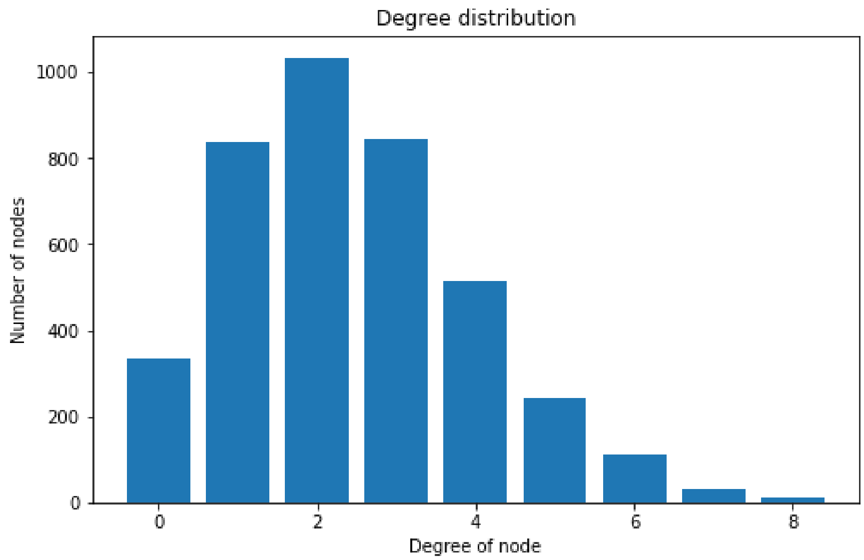


Number of node	3962
Number of edges	4101
Average clusetring	0.0020439531270307724
Network desnity	0.000261318679946235

Figure 20

Our degree distribution is left-skewed which tells us that it's following a real scale-free network structure. The average clustering is close to 0 which tells us that most of the node's neighbors in our network aren't connected. And the network density is telling us that we have a sparse network.

Erdos Renyi random network

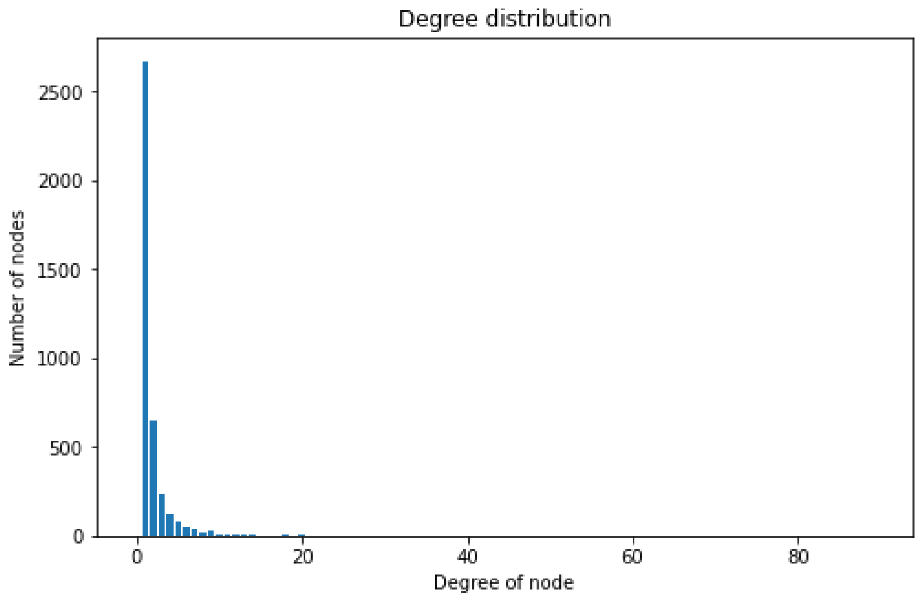


Number of node	3962
Number of edges	4843
Average clusetring	0.00023136463065791686
Network desnity	0.0003085994554936884

Figure 21

In Figure 21, the density and average clustering are low for the random network as well because the number of connected edges is the same. For the degree distribution plots, we can see there are no nodes that have a very high degree. The highest degree node in the ER model is 8. This is because the nodes aren't connected on the basis of any preference or they do not get connected to other nodes on the basis of any particular reason, instead, they're connected on a random basis which means almost all nodes have the same number of edges compared to some specific nodes which have higher degrees. Thus, we can't take this model as a benchmark.

Barabasi Albert random network

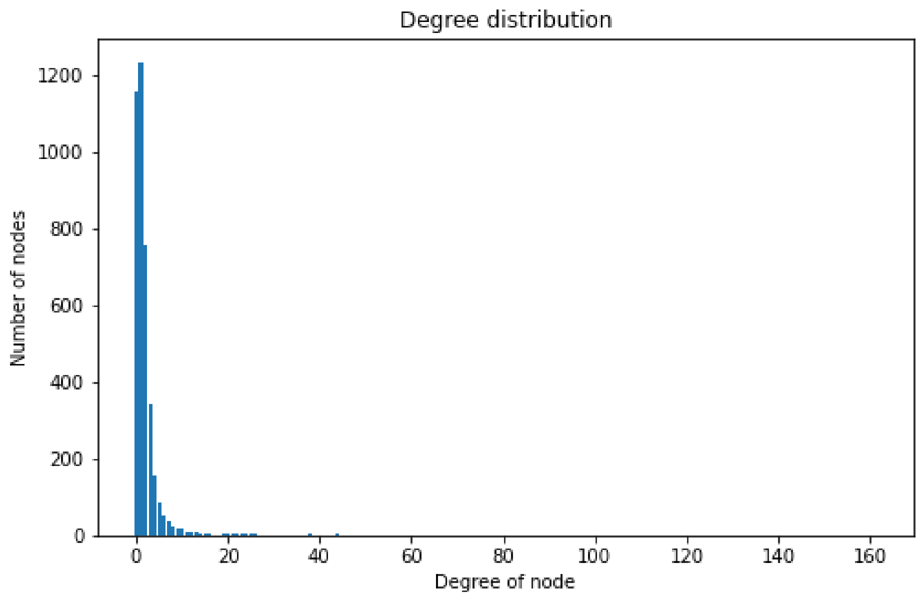


Number of node	3962
Number of edges	3961
Average clusetring	0.0
Network desnity	0.0005047955577990914

Figure 22

In Figure 22, the degree distribution of the Barabasi Albert random graph with preferential attachment looks like our original network but from the clustering coefficient and density values, we can see that the nodes still don’t preserve their degrees to be a benchmark random network for our model.
Here the average clustering value = 0 which tells us that the nodes aren't well connected to their neighbors compared to the nodes we have in our original network.

Degree-preserved random network



Number of node	3962
Number of edges	4101
Average clusetring	0.012858068210956848
Network desnity	0.00052263735989247

Figure 23

In Figure 23, the degree sequence from our original graph has been used here to generate this random network model. From the results, we can see that the density and clustering values of this network are more consistent with the real scale-free network properties with the preserved degrees from our original graph. We can also see that the average clustering here is better compared to our original model so we can use this model as a benchmark.

Discussion

Discuss your results and where appropriate frame them in existing literature.

The analysis we did for our network was more qualitative than quantitative since we make a lot of assumptions based on the sentiments of the most influential users. We found that the overall sentiment toward bitcoin in our network was mostly positive. All the tweets in our network are coming from verified users, which means all the users making those tweets have some level of influence. And so, if the majority of those users have a positive sentiment towards bitcoin, then the people that view those tweets (their followers and potentially others) may also be influenced to think about bitcoin in a positive way. We can't prove that since we don't know whether the people who saw the tweets purchased bitcoin after reading it or not. But we assume that because a tweet is coming from a verified user it has some level of credibility, so people are more likely to be influenced by it. A recent example of an algorithm using Twitter sentiment to make trades based on the tweets of an influential person is called TRUMP2CASH [5]. This algorithm, which trades based on Trump's tweet, has, according to its historical benchmark, an annualized return of 59% since its inception. These results indicate that the effect of the tweets of influential people is worth looking into as a predictive element. Moreover, The study of Oh and Sheng [6] shows that micro-blog sentiments contain information that can be valuable to investment decision-making and that people do in fact use social media to influence their investment decisions. We also found that the sentiment of the top 20 verified users mentioning the most people is mostly positive as well (outdegree-based sentiment). This means that those top users are mentioning people in tweets that have a more positive sentiment towards bitcoin. As a result, the people being mentioned may be influenced to think about bitcoin in a more positive way as well. Similarly, with the sentiments of the top 20 verified users who can propagate information the fastest in our network (betweenness-based sentiment), these users also have a mostly positive sentiment towards bitcoin which tells us that they may influence people's thoughts about bitcoin in a more positive way, faster than any other users in our network. We also found that the top 20 verified users being mentioned the most are mentioned within tweets that have a mostly positive sentiment (indegree-based sentiment). This tells us that those users may themselves get influenced by the sentiment of the verified users mentioning them. This may influence their decision on purchasing bitcoin as well.

What did you learn about your network?

We learned, from our degree distribution, that our network follows the real scale-free network properties. We also observe preferential attachment. Since Elon Musk was the most mentioned user, if we were to scale this network, Elon Musk would be mentioned by even more people. From our network density, we learned that we have a sparse network which makes sense because most verified users mention 1-3 people. Moreover, since our average clustering value is close to 0, we know that most of the node's neighbors in our network aren't connected.

Did you successfully answer your research questions? If not, why not?

We successfully answered our research questions by conducting a social network analysis of our network.

We answered our first question by finding the top nodes in our network based on 3 different centrality measures, indegree, outdegree, and betweenness centrality. These measures helped us identify the people in our network that can have the most influence on people's decision on whether to invest in bitcoin or not.

We answered our second question by extracting and examining the sentiment of the whole network, the most influential users, and the different communities, thus allowing us to learn the different ways communication technology and more specifically Twitter can influence people who read bitcoin-related tweets on their platform and in what direction (positive or negative). We found out that this influence is mostly positive for all three different ways.

Are there any limitations inherent to your data or your approach?

There are no limitations that come with our dataset. However, an inherent limitation in the approach is that there is no way to quantitatively find out if someone was influenced by a tweet or not. Thus, we have to assume that influential people talking about bitcoin will in fact influence a subset of people in regard to their decision-making process when investing in bitcoin.

What would future work want to consider?

In terms of future work, we would need a dataset of bitcoin prices to pair with the tweets in our dataset. This will tell us if there was any change in price after a tweet was made. Using this information we could perform a quantitative analysis of bitcoin price fluctuations based on tweets shared at a certain time and their sentiments. Secondly, we can get the retweet information for each tweet which would help us get more accurate details as to what number of users are reading a certain tweet. This would allow us to rank a user's influence based on their outreach as well which would give us a more accurate representation of who the most influential people are in our network.

Methods

1) We use networkx to get the basic measures of our network and the null models used to compare to our network. We define a function called basic_details() which takes a graph object as an input and outputs the basic details such as the number of nodes, number of edges, average clustering, and network density. We also define a function called degree_dist() which takes a graph object as an input and outputs a histogram with the degree distribution which takes into account the number of nodes and degree of a node.

1) We get the overall sentiment of all the tweets in our network to get a general idea of what people think about bitcoin. We use NLTK libraries to clean the text of the tweets in our dataset and tokenize the sentences which helps us apply sentiment analysis on each tweet.

2) We use NRCLex to get a sentiment analysis of all tweets in the network (get_emotions function in our code). Once we analyze the sentiment of the whole network, we will know how Twitter as a whole affects the decision-making process of someone using the platform.

3) We then use networkx to get the top 20 people being mentioned (indegree), the top 20 people mentioning others (outdegree) to find the most influential users, and the top 20 nodes that can propagate information the fastest within the network (highest betweenness centrality).

4) We then define a function called get_selected_tweets() which selects the tweet data from the edge information for a particular centrality measure. We also define a function called

print_sentiments() which helps us get the word cloud visualizations and visualize bar plots of the sentiments for each of the top-ranked central nodes.

5) Once we analyze the sentiment of the influential people, we will know how the decision-making process of the people following those influential users will be affected.

6) We then use the cdlib rb_pots algorithm to generate communities based on shared mentions. We then find the overall sentiment of the different communities since different communities could have different sentiments towards bitcoin. This will allow us to discover the sentiment of each community which allows us to predict the sentiment of the users that are a part of each community.

7) We then define a function called positive_negative_wc() which helps us display the positive and negative words being used in each community which we can use to analyze the sentiments each community has towards bitcoin.

8) Once we analyze the sentiment of the communities, we will know how the decision-making process of the people interacting in each community may be affected

Note: the community detection algorithm returns different results every time it's ran. The results we presented above are one of the possible outputs that we get, however, this doesn't affect the results since they remain consistent every time we run the algorithm.

Code

https://github.com/mohamedhefnawy1/Effects-Of-Communication-Technology-On-Cryptocurrency

References

Our Dataset: https://www.kaggle.com/datasets/kaushiksuresh147/bitcoin-tweets

[1] https://www.oberlo.ca/statistics/how-many-people-use-social-media

[2] https://www.shs-conferences.org/articles/shsconf/abs/2020/02/shsconf_glob2020_03014/shsconf_glob2020_03014.html

[3] https://www.weforum.org/agenda/2022/11/cryptocurrency-us-midterms/

[4] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7719982/

[5] Max, B.(2017). https://github.com/maxbbraun/trump2cash

[6] Oh, C. and Sheng, O. (2011)
https://www.scirp.org/(S(351jmbntvnsjt1aadkozje))/reference/referencespapers.aspx?referenceid=2661934