Quantitative Content Analysis of Bitcoin Facebook Groups

September 2019

1 Abstract

Social media sites such as Facebook allow people across the globe to converse and share posts without borders. Users can form groups to interact about a variety of topics, and cryptocurrency is no exception. The focus of this study was to determine the demographics and activity of groups focused on both Bitcoin specifically and cryptocurrencies more broadly, and how they compare to one another. Results indicated that groups found in search results for "bitcoin" were typically larger and more active than those for "cryptocurrency". However, the major finding of this study was that many of the groups served primarily to promote dubious money-making opportunities and in some cases outright scams.

2 Introduction

Bitcoin was described by its anonymous creator(s) Satoshi Nakamoto as a way to send cash digitally without a third party. [3]. The currency started as a response to the Global Financial Crisis (the first-ever transaction on the Bitcoin blockchain was hashed with the message "Chancellor on brink of second bailout for banks" [5]) and had no value at the time it was launched. Ten years later, it trades on dozens of central exchanges globally, as well as on several over-the-counter (OTC) markets. Traditional market venues such as the Chicago Mercantile Exchange and Chicago Board Options Exchange now list Bitcoin futures and options contracts, and Intercontinental Exchange subsidiary Bakkt has come online in the past month to offer physically-settled futures contracts as well. Both exchange trading volumes and on-chain network transaction volumes regularly exceed 1 billion USD globally on a daily basis. These facts and figures serve as evidence that Bitcoin has carved out a niche in the world economy, though its exact role is largely yet to be determined.

This role determination happens largely through digital means, on social media sites such as Facebook as well as messengers and chat apps like Telegram

and Discord. By analyzing the activity in these channels, we hope to gain a better understanding of the cryptocurrency landscape.

3 Literature Review

3.1 Can you see the writing on my wall?

The article examined how, and the extent to which large corporations were using Facebook. The study focused on the number of users that interacted with the pages through their various functions (wall posts, discussion boards, photos and videos). This information was used to assess whether the platform was being used to its fullest extent.

Content analysis on the Facebook member and/or fan pages of Fortune 50 companies was conducted between February 21, 2009 and March 7, 2009. The study excluded pages not representative of a company (i.e. product pages, dissolved corporations). Pages with less than 100 members/fans were also left out. The publicly viewable page discovered in the first five pages of Facebook search results was chosen for coding. The study coded pages, wall posts, and discussion boards into custom variables so they could be used for statistical analysis.

It's essential for modern corporations to use social media (among which Facebook was the largest at the time the study was conducted [2]) to communicate with customers and the community at large. However, the findings of the study indicated that the companies studied (the nation's largest, with therefore the most to gain) largely under-utilized the technology. This implies that there was plenty of potential for growth through expanded use.

3.2 The 2008 Presidential Election, 2.0

The 2008 presidential election was the first to experience a significant impact from social media. The number of active users on the site allowed for analysis of how the voting public communicated about political issues. The key areas of focus for the study were the extent to which political discourse occurred on the platform, and whether candidates were typically viewed in a positive or negative light, and the differences between candidates with respect to the previously listed factors.

To collect data, the study selected the first four groups returned as search results for each candidates; the results were sorted by descending number of group members, and the most popular groups contained a disproportionately high number of members [6]. After the first, every fourth group was selected from the results. The samples were taken from a snapshot of data immediately preceding the election. All elements of the pages (titles, descriptions, posts, photos) were coded as positive or negative. The presence of profanity was also noted. These coded variables were then used to compare the distributions of the two candidates.

Although it's difficult to tease out causation from such a study, several interesting findings correlating Facebook activity to the actual election results were made. Facebook groups for Obama had more activity and viewed him more positively as a candidate. Meanwhile, the groups associated with McCain were determined to be more frequently negative. These results matched the outcome of the actual election, where Obama won handily.

4 Rationale and Research Questions

Most financial assets are governed (albeit loosely) by the concepts of supply and demand: if market participants wish to buy an asset in greater amounts than are currently available, they'll drive the price up and vice versa. For most assets, these characteristics are somewhat inelastic. Corporations around the world need to exchange their local currencies for foreign ones to conduct international business. Food producers need to hedge the prices of ingredients such as sugar and cocoa to accurately forecast their costs. Pension funds need to invest their capital into bonds and equities to generate a return for their investors.

The situation for Bitcoin is different, and perhaps somewhat analogous to gold. Very few people actually *need* Bitcoin for their day-to-day activities. Most financial needs are adequately met by legacy systems (banks and payment processors like PayPal, Visa, etc.). The asset therefore derives its value predominantly from traders' and investors' opinion on its adoption and usage in the future. The field of behavioral finance (functionally more a branch of psychology) has shown that these opinions are often just that, and not always rational [1].

For this reason, market participants' sentiment is critical in determining what the value of Bitcoin actually is. With no solid way of determining a fundamental "fair value" and a smallish market capitalization relative to more traditional assets, the price can fluctuate rapidly at the whim of trigger-happy traders.

With that in mind, this investigation set out to determine what information publicly available on Facebook might indicate about those actively interacting about Bitcoin. First and foremost, the groups under study were split into two groups: those focused around "Bitcoin" specifically, and "Cryptocurrency" more broadly. Within the space, there are those (nicknamed "maximalists") who believe that as the original cryptocurrency, Bitcoin deserves their sole attention. An opposing school of thought is that Bitcoin has undesirable characteristics that can be addressed with other, newer currencies commonly referred to as "altcoins". The following research questions were thus considered:

- What is typical size and activity level for these Facebook groups?
- What is the primary focus of the group (trading, mining, etc.)?
- What are the demographics of recently active users (age, gender)?
- What do users typically post about?

- How much engagement is generated on user posts?
- What is the general sentiment in user posts?

5 Methodology

5.1 Data Collection

Data collection from Facebook is impeded by the fact the company prohibits all automatic scraping of its sites without express written permission, as per its robot.txt file [4]. In addition to this, they drastically limited access to their APIs in the wake of the Cambridge Analytica scandal, making it difficult even for those willing to take their chances with Facebook's formidable legal team. The situation worsens further for European researchers, where General Data Protection Regulation increases the chances of litigation for illegal scraping.

To complete this investigation in full compliance with all laws and regulations, the quantitative content analysis was conducted "by hand" a.k.a. manually recording all findings into a spreadsheet. The first step was to find eligible Facebook groups. This was done by entering either "bitcoin" or "cryptocurrency" into the Facebook search feature. Among the result options are Pages and Groups, which serve different purposes on the site. While Pages are official spaces for organizations to communicate with their followers, Groups are places for users to interact around a common theme/topic. Groups were determined to be better suited candidates for this analysis. Only public Groups were considered because user posts are not visible in private Groups.

It was unclear what criteria was used to sort the results, although they appeared to roughly follow number of group members in descending order. For each search term, every fourth Group result was selected. This method of sampling was chosen to generate a random (and hopefully representative) sample from which to draw insights, since the algorithm used to sort the results wasn't explicitly known. 15 Groups were chosen for each search keyword in this fashion. For each Group considered, information was gathered from the "About" and "Discussion" pages. The former contained metadata about the Groups themselves, while the latter displayed posts from users. On the discussion page, the top ten posts displayed were used to gather statistics. 150 posts were analyzed for each search term. Information from both pages was coded for analysis.

5.2 Content Analysis

Page information that was either numerical or explicitly defined was left unaltered. From the "About" page, this included:

- number of Group members
- number of posts made on the day when accessed

- number of members that had joined in the past 30 days

From the discussion page, the following data was used as-is:

- gender of poster (found in the "About" section of each user profile)
- number of comments on each post
- length of posts (number of lines displayed on desktop browser, capped at 5)

The remainder needed to be assigned categories (see Appendix for possible values of each variable). Each group's focus was determined by reading the description on the "About" page and choosing the best match if possible. If the page contained rules for posts either in an explicit "Rules" section, or outlined in the description, this was recorded.

Information about individual posts needed to be coded into categories, as well. First, each poster's age was placed into three broad buckets: under 25, 25-40, and over 40. This was done subjectively based on user's photos and information such as education, job status, and family (children) found on their personal pages. This obviously induces some loss in accuracy for this category, but for broad demographic analysis it was still thought to be useful. The type of post was recorded based on what types of media it contained (text, images, videos, links). The post content was also assigned a category, based on popular topics of discussion in the cryptocurrency space (trading, investing, mining).

An important note on this category relates to the "opportunity" category. This designation means that a post's primary purpose was to offer the reader a chance for some type of gain, usually financial. Posts in this category typically carried messages such as "I'm not offering you millions of dollars but I will teach you how to earn minimum bitcoin of $1000 in 3days [sic] with your smart phone. Inbox me or WhatsApp [user's phone number]". While such propositions may be regarded skeptically by most, likely drawing accusations of being scams, there was a possibility that the poster was offering some legitimate service. The separate category "outright scam" was reserved for those posts that blatantly acknowledged participation in less-than-legal activity. These included gems like the following: "If you are a scammer inbox me I want to spend 50$ [sic] in so called bitcoin mining scam".

Finally, the sentiment of each post was subjectively determined to be either "bullish" (indicating the user thought the price of an asset was going to increase), "bearish" (opposite of bullish), or neutral.

6 Results

As one can observe in 1 and 2, Groups returned as search results for "bitcoin" generally had more members and more activity than those for "cryptocurrency". This is demonstrated by every numerical statistic recorded for each; bitcoin

groups were both larger in absolute terms and had more members join recently. In addition, these groups had more recent posts with more comments on each.

Post activity by gender was biased, with males accounting for over 65% of the total posts, while females accounted for less than 30% (see appendix). The ages, though possibly inaccurate due to subjective coding, were fairly evenly distributed; under 25 and 25-40 both account for roughly 32% of posts, with the over 40 crowd making up 19%.

The most interesting finding, however, was the content of the posts. Across both search terms, "opportunity"-related posts dominated the feeds. Over 70% of all activity fell into this category. The next closest, news, made up just over 10%. Trading and Investing combined for less than 6%; disappointing, given what the study set out to accomplish. As a result, most of the posts were neutral sentiment. Among those that did contain sentiment, bullish posts out-numbered bearish ones 32-4.

		bitcoin	cryptocurrency
# Members	mean	44591.6	13483.9
	std	34699.9	17949.6
	min	13653	289
	max	126067	69859
Posts Today	mean	500.3	42
	std	383.5	66.3
	min	0	0
	max	1133	251
Members in Last 30 Days	mean	14488.6	697.1
	std	31024.4	887.7
	min	788	19
	max	126067	3533

Table 1: Group Stats Comparison

		bitcoin	cryptocurency
Post_Length	mean	3.8	2.9
	std	1.6	1.8
	min	1	1
	max	5	5
#_Comments	mean	26.4	7.6
	std	24.8	9.7
	min	1	1
	max	87	33

Table 2: Post Stats Comparison

7 Conclusion / Future Work

The findings of this study did not meet the goal it was set out to achieve. What began as an investigation into the demographics and user sentiment towards Bitcoin and other cryptocurrencies revealed large groups of people whose main goal was to peddle their services to others. Based on available information about how Bitcoin functions, it seems implausible that the authors of such posts would be able to deliver on the claims they made. It seems more likely that the majority of these posts are ill-intentioned, and seek to take advantage of those with litte expertise on the subject. It's difficult to estimate how successful these users are in their attempts.

Follow-up studies on the topic may want to use different keywords in their searches, such as "bitcoin trading" or "bitcoin development" to find more serious, focused information. If Bitcoin is truly going to be the global reserve currency of the future, it still has a long way to go.

References

[1] D. Kahneman. *Thinking, Fast and Slow.* Farrar, Straus and Giroux, 2011.

[2] Tina McCorkindale. Can you see the writing on my wall? A content analysis of the Fortune 50's Facebook social networking sites. *Public Relations Journal*, 2010.

[3] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2009.

[4] Octoparse. 5 things you need to know before scraping data from facebook. Accessed: 2019-09-28.

[5] Bitcoin Wiki. Genesis block. Accessed: 2019-09-28.

[6] Julia K. Woolley, Anthony M. Limperos, and Mary Beth Oliver. The 2008 presidential election, 2.0: A content analysis of user-generated political facebook groups. *Mass Communication and Society*, pages 631–652, 2010.