Are Bitcoin Users Less Sociable? An Analysis of Users' Language and Social Connections on Twitter

Ivan Hernandez, Masooda Bashir, Gahyun Jeon, and Jeremiah Bohr

University of Illinois, Urbana-Champaign, Illinois, USA.
{hernan27,mnb,jeon29,jbohr2}@illinois.edu

Abstract. Bitcoin, a peer-to-peer payment system and digital currency, has seen much growth and controversy in the four years since its introduction. Yet, despite Bitcoin's growing importance, little is known about its users. Our research explores what type of people use this domain and what concepts they tend to emphasize in their language. We analyzed over 50,000 messages from over 6,000 users of the social networking community, Twitter. Our analyses show a consistent pattern that people interested in Bitcoin are far less likely to emphasize social relations than typical users of the site. Specifically, Bitcoin followers (1) are less likely to mention family, friends, religion, sex, and emotion related words in their tweets and (2) have significantly less social connection to other users on the site. These findings offer the first empirical look at what exactly makes Bitcoin users distinct from others and can have implications for the future of the currency.

Keywords: Bitcoin, Twitter, LIWC, Relationships, Text Analysis, Virtual Currency, Social Network Sites.

1 Introduction

Bitcoin, a peer-to-to-peer payment system and decentralized digital currency, has gained an increasing amount of public interest since its inception in 2009. The currency is represented as data within a shared network, and generated by anyone running a Bitcoin mining application over the internet, which can then be transferred directly to people in the network. Over the last four years, its price has fluctuated wildly, going through various cycles of appreciation and depreciation, reaching valuations as high as $1,000 USD per coin to lows of less than a dollar. The trend though has steadily continued upwards, and the currency is receiving more mainstream attention. Popular retailers such as Overstock, Zynga, Wordpress, Baidu, and TigerDirect all accept Bitcoins as a form of payment. Politically, governments across the world have also acknowledged its growing importance. Chinese, Finnish, German, and Canadian governments have all established policies for Bitcoin use, and most other major world governments have issued statements concerning their position on regulation [1]. Yet, despite the currency's growth and potential to impact major markets, we still know very little about the people using Bitcoin. However, it is the users who promote and

C. Stephanidis (Ed.): HCII 2014 Posters, Part II, CCIS 435, pp. 26–31, 2014.

affect the value of the currency. Therefore, understanding their thoughts, feelings, and values may inform us about the future of Bitcoin.

2 Language Analysis

The present research seeks to examine what Bitcoin users think more/less about compared to the typical person. One way to address this question is to study the content of their speech. Psychological research on language finds that the words we use mirror our thoughts and feelings at that moment [2]. Traditionally, analyzing language content involved performing a case-by-case coding of conversations by a trained expert. While this qualitative approach allows for an in-depth understanding of a small sample of conversations, the method was not designed to get a comprehensive picture of an entire culture or group. However, over the past decade, computer-based methods of text-analysis have addressed these issues handling larger amount of text in a faster, broader, and more cost-efficient way.

One of the most widely used text analysis computer programs is the Linguistic Inquiry and Word Count (LIWC) [3]. LIWC analyzes text samples (e.g. a news story, a blog post, an e-mail) on a word-by-word basis and compares each word to an internal dictionary of over 2,000 words divided into different linguistic categories (e.g. positive emotions, personal pronouns, money-related). For every text sample, it outputs the percentage of total words in the text that reflect each linguistic category. For example, if the four-word text sample, "I am happy, today," is given to LIWC, the it would output a value of .25 for positive emotion (i.e. "happy"), and .25 for 1st person pronoun (i.e. "I"), and would give a value of 0 for categories like "money."

While LIWC contains dictionaries that measures traditional linguistic content such as parts-of-speech, one of its greatest benefits is that it also contains dictionaries that measure psychological processes (e.g. anger, cognitive mechanisms, social engagement) in text. These dictionaries were developed using methods similar to those used for traditional psychological scales. First, the developers created a list of emotional and cognitive dimensions often studied in social, health, and personality psychology. Then, using reference books, past psychological scales, and personal opinion, they created an initial list of words that matched each dimension. Independent judges then rated their acceptability, with majority agreement deciding what words remained. These remaining words formed the preliminary dictionaries. To evaluate the dictionaries' psychometric properties, the authors cross-validated them against a corpus of 24,000 text samples totaling over 168 million words. Words were excluded if they were used less than .005 percent of the time or not mentioned in English word frequency reference manuals. The internal consistencies for binary ($\alpha = .83$, $\sigma = .15$) and percentage codings of the words ($\alpha = .40$, $\sigma = .16$) are in ranges common to the field. Further, most dictionaries have a correlations of .40 or greater with judges' ratings of text on the dimension they represent, establishing the dictionaries' predictive validity. Although a non-contextual word count strategy such as LIWC will be more prone to errors than a human coder would, the amount of influence an error has on the results diminishes with larger data. Thus, LIWC has the potential to provide insight into the thoughts and values of people, and its disadvantages are minimized as the data it is given increases.

3 Twitter Data

In the present research we use the microblogging website, Twitter, as our source of daily language. Twitter is currently the 11th most popular website in the world [4] and allows users to post short form messages ("tweets") that describe their thoughts, sentiments, and concerns at a given moment. Studying Twitter data offers several methodological advantages for this project. By using Twitter, instead of survey questions, we lower the risk of demand characteristics and observer effects in people's responses. Rather, this data is more representative of a user's mindset as it occurs over a longer period of time and in an unprompted, casual setting. Therefore, Twitter data is more suited for studying a person's state of mind on a day-to-day basis, and recent research finds convergent results between data collected on Twitter and psychological studies [5,6]. Additionally, Twitter's popularity allows researchers to study populations that would be difficult to access through university participant pools or field surveys. Users interested in certain people/topics (e.g. Barack Obama, Pope Francis, Apple) subscribe ("follow") to updates from those figures/companies/topics. Therefore, the followers from those account provide a sample of the desired population to study. Lastly, the magnitude of the data is significantly greater than in traditional contexts, allowing for a more comprehensive analysis of a population. Twitter thus provides a unique opportunity to study psychological constructs on a large scale that is not possible through traditional survey and laboratory methods.

4 Method

4.1 Data Collection

Data collection began on February 1st, 2014 and ended on February 4th, 2014. Using the Twitter Advanced Programming Interface (API) via the Twython package for Python [7], we obtained a list of the of the followers to the most popular Bitcoin exchange at the time, MTGOX (@MTGOX), which had approximately 25,000 followers. To increase the likelihood the users in our sample were human and not automated accounts, we only included users who posted to Twitter from a web browser or a mobile application. From this list, we randomly sampled 14,956 users. Of those users, 34%, (5,101) made their information publically available. We then accessed their entire message history on Twitter since the creation of their account. Additionally, we recorded the number of accounts the user followed, and that followed the user.

We also collected the tweets of 4,988 randomly sampled users who posted from a browser or mobile device. This sample came from the Twitter Streaming API, which provides a real-time sample of all tweets posted publicly at a given moment. Upon connecting to the API, we compiled a list of the users in stream. Therefore, this sample of users serves as a representation of a typical person who posts on Twitter and as a control group to the users interested in Bitcoin.

4.2 Data Preparation and Cleaning

Due the noise that non-English and non-active accounts introduce into text analysis, we followed the exact data cleaning procedures that similar research has used [6]. Because the LIWC dictionaries are based on the English language, we removed all timelines that were not in English. We filtered users by coding the percentage of words in a user's timeline that matched the stop words (e.g. prepositions, articles, pronouns) of 14 European languages. The collection of stop words came from the natural language toolkit (NLTK) package for Python. We removed any user that had a higher percentage of stop words in any non-English language than the percentage of English stop words. Because shorter text may not be as reliable and some accounts may not be active, we only included timelines that contained at least 20 tweets in their history. Lastly, we removed all numbers, special characters, hyperlinks, and punctuation from the text and converted each letter to lowercase. After data cleaning, our dataset contained 2,673 MTGOX followers, and 4,180 control users.

4.3 Measures

The current research took an exploratory approach to examine what psychological terms/categories/identifiers Bitcoin Twitter users differ on compared to the average Twitter user. We used the Linguistic Inquiry and Word Count (LIWC) program to measure how much their language emphasizes a variety of psychological processes. The internal dictionary contains six categories of psychological processes: "social," "affective," "cognitive," "perceptual," and "biological." Those conceptual dictionaries contain various subdictionaries that capture the different facets of the general domain. For example, the "social" dictionary is a collection of subdictionaries relating to "friends," "family," and "humans" in general. LIWC also contains dictionaries relating to personal concerns including: "work," "achievement," "leisure," "home," "money," "sex," and "religion." (see Pennebaker et al for a full list of the LIWC dictionaries and sample words[3]).

5 Results and Discussion

Due to the large sample size (N = 6,853 users), all tests of mean differences were significant at the conventional .05 level. Therefore, for the results, we discuss the pattern of effect size differences (*Cohen's d* and confidence intervals of the effect) between Bitcoin followers on Twitter and the typical user. Our results show a clear and consistent pattern, where Bitcoin users are less likely to emphasize concepts pertaining to relationships with others.

Compared to the typical site user, Bitcoin users were far less likely to talk about content in LIWC's "social" dictionary (d = -1.34, 95% CI = [-1.40, -1.29]). This dictionary contains all of the words found in the subdictionaries for family (e.g. brother, mother), friends (e.g. buddy friend), and humans (e.g. adult, baby) as well as all non-first-person-singular personal pronouns and verbs that suggest human interaction (e.g. talking, sharing). Additionally, the subdictionaries for family (*d* = -.92, 95% CI = [-.97, -.87]), friends (*d* = -1.0, 95% CI = [-1.05, -.95]), and humans (*d* = -.69, 95% CI = [-.74, -.64]) reflected the same pattern, indicating that the differences are not restricted to one specific social aspect. Related dictionaries show a similar trend. Bitcoin users are less likely to talk about sexual topics (*d* = -1.28, 95% CI = [-1.33, -1.23]), religion (*d* = -.48, 95% CI = [-.53, -.43]), and the home (*d* = -0.35, 95% CI = [-.4, -.31]. As a further examination, we calculated the 50 words that Bitcoin users say more often relative to typical users, and did the same for the typical users. These differences also suggest a lower social emphasis in what Bitcoin users say (figure 1).

Fig. 1. Words that showed the biggest positive difference in frequency between typical site users compared to Bitcoin users (left) and Bitcoin users compared to typical site users (right). Font size indicates the relative magnitude of the difference between groups for that word.

We also examined differences in language for concepts related to social processes, such as emotions (both positive and negative). Past research finds that emotions serve social functions and arise in response to social issues [8]. Their presence in speech may therefore serve as another indicator of social connection. Bitcoin followers were much less likely to have emotional content (*d* = -1.13, 95% CI = [-1.18, -1.08]) in their tweets. In addition to examining overall emotion differences, we found the same trend for the emotional subdictionaries. Bitcoin followers were less likely to mention both positive (*d* = -.93, 95% CI=[-.98, -.88]) and negative emotion words (*d* = -1.33, 95% CI=[-1.39 , -1.28]). The negative emotion dictionary is composed of subdictionaries relating anxiety, anger, and sadness. Bitcoin users were less likely to mention sadness (*d* = -.66, 95% CI=[-.71, -.61]) and anger (*d* = -1.16, 95% CI=[-1.21, -1.11]), but were equal on anxiety (*d* = .003, 95% CI=[-.08, .09]). The swearing dictionary shows a similar trend, where Bitcoin users are less likely to use expletives (*d* = -1.10, 95% CI=[-1.15, -1.05]) Therefore, Bitcoin users' tweets contained generally less emotional content, with the exception of anxiety.

This lack of social focus was also reflected in the users' connection with other users on the site. We compiled a list containing the number of people the Bitcoin followers/control users followed as well as how many people followed them. Due to the strong left skew in the distribution of each, the log-transformed the number of people each user followed and that followed them. We again found a strong difference between Bitcoin enthusiasts and typical users. People interested in Bitcoin on Twitter followed less people (d = -1.04, 95% CI = [-1.06, -1.02]) and had less people following them as well (d = -.58, 95% CI = [-.60, -.57]).

6 Conclusion

Our results show that people interested in Bitcoin on Twitter are much less likely to emphasize socially related dimensions in their daily language. Concepts pertaining to family, friends, humans, home, religion, sex, swearing, and emotions were mentioned less frequently, compared to a typical site user. We also found that they were less engaged with others on the site compared our control sample. This research is the first to suggest what psychological differences exist between Bitcoin users and the average person. Based on our findings, one possibility is that people interested in Bitcoin are distinctively less socially involved and emotionally expressive. Future research may seek to examine these differences further in different contexts as well as the implications less social connection may have. While Bitcoin has become increasingly more popular, the public may still be largely unaware of its details. This lack of information may be due in part to its users' lower social connectedness. Further, the adoption of the currency may be slowed if its users' social networks are less dense compared to others'. Collectively, our findings suggest many new possible avenues for research on Bitcoin users, and offer insight into how this currency may develop in the future.

References

1. BitLegal Index, http://bitlegal.io/list.php
2. Pennebaker, J.W.: The Secret Life of Pronouns: What Our Words Say About Us. Bloomsbury Press, New York (2013)
3. Pennebaker, J.W., et al.: The Development and Psychometric Properties of LIWC 2007. LIWC. Net, Austin (2007)
4. Twitter.com – Site Info, http://alexa.com/siteinfo/twitter.com
5. Golder, S.A., Macy, M.W.: Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures. Science 333(6051), 1878–1881 (2011)
6. Ritter, R.S., et al.: Happy Tweets: Christians Are Happier, More Socially Connected, and Less Analytical Than Atheists on Twitter. Social Psychological and Personality Science 5(2), 243–249 (2014)
7. Twython 3.1.1, http://pypi.python.org/pypi/twython
8. Keltner, D., Haidt, J.: Social Functions of Emotions at Four Levels of Analysis. Cognition & Emotion 13(5), 505–521 (1999)