

# Influenciadores Digitais da Criptomoeda Bitcoin: um método de coleta e tratamento de dados

Paulo Henrique Ribeiro Costa<sup>1</sup>, Jorge Luiz Bezerra de Araújo<sup>1</sup>,  
Luciano Heitor Gallegos Marin<sup>2</sup>

<sup>1</sup>Universidade de Fortaleza – UNIFOR  
Fortaleza – CE – Brasil

<sup>2</sup>Universidade Federal do Paraná – UFPR  
Curitiba – PR – Brasil

{paulo7676@gmail.com, jorgearaujo@unifor.br, luciano.gallegos@gmail.com}

**Abstract.** *The Bitcoin release has been impacting the global digital market due to its decentralized and secure technology. In the same sense, the digital environment encourages the emergence of cryptocurrency influencers. Is there any way to identify these cryptocurrency influencers, taking into account the large amount of users in online media and social networks? In this article, we describe a method to collect and process texts from an online social network in order to identify these influencers.*

**Resumo.** *A criação do Bitcoin vem impactando o mercado digital global devido à sua tecnologia descentralizada e segura. Neste mesmo sentido, o meio digital incentiva o surgimento de influenciadores digitais de criptomoedas. Haveria alguma forma de identificar estes influenciadores digitais de criptomoedas, levando-se em consideração o grande número de usuários de mídias e redes sociais online? Neste artigo, descrevemos um método para realizar a coleta e o tratamento de textos de uma rede social online com o objetivo de identificar estes influenciadores, visando descobrir influenciadores digitais da criptomoeda Bitcoin.*

## 1. Introdução

Criptomoeda é um tipo de dinheiro digital, que começou a ser conhecido no ano de 2008. Foi neste mesmo período que houve uma grande recessão econômica mundial e, o artigo seminal chamado "Bitcoin: a Peer-to-Peer Electronic Cash System" escrito por Satoshi Nakamoto [Nakamoto 2008] foi publicado. Embora até hoje não se saiba quem é Satoshi Nakamoto, o artigo e suas ideias transformaram as transações financeiras digitais no mundo. No artigo, Satoshi descreve um sistema de pagamentos descentralizado e baseado em criptografia e, no ano de 2009, obtêm-se o primeiro registro de mineração de criptomoeda do mundo: o Bitcoin.

Desde a criação do Bitcoin, o mercado global vem sendo impactado por esta criptomoeda devido à sua tecnologia e método disruptivo, possibilitando a realização de transações monetárias pela internet sem existência de um mediador bancário. A Bitcoin faz uso do Blockchain: uma tecnologia distribuída que visa a descentralização de dados como medida de segurança, armazenando-os em uma lista encadeada de livre acesso

para o público em geral [Bovério and da Silva 2018]. O mesmo meio online digital que permite transações de criptomoedas, também permite que influenciadores digitais atuem.

Os influenciadores digitais são indivíduos com capacidade de influenciar o comportamento de indivíduos e comunidades em um meio online por meio de textos, imagens e áudios [Uzunoglu and Kip 2014]. A Bitcoin, por ter se tornando a criptomoeda mais famosa e negociada do mundo, promoveu o surgimento de outras moedas virtuais e influenciadores digitais que dedicam parte de seu tempo na criação e disseminação de notícias sobre o tema, mostrando o funcionamento do mercado influenciado a negociação de Bitcoins.

Os influenciadores digitais de criptomoedas nem sempre são conhecidos, e muitos deles utilizam pseudônimos para realizar suas comunicações. Assim sendo, haveria alguma forma de identificar estes influenciadores digitais de criptomoedas, levando-se em consideração o grande número de usuários de mídias e redes sociais online? Visando descobrir influenciadores digitais da criptomoeda Bitcoin, descrevemos neste artigo um método para realizar a coleta e o tratamento de textos de uma rede social online com o objetivo de identificar estes influenciadores.

O artigo está organizado da seguinte forma: na seção 2, descreve-se como foi realizado a coleta, o tratamento e o armazenamento de textos extraídos da plataforma Twitter. Já na seção 3, são apresentadas formas para identificar influenciadores pela quantidade de textos compartilhados e retweetados, além da análise da identificação de tópicos discutidos por estes influenciadores digitais sobre Bitcoins. Finalmente, na seção 4, são realizadas considerações acerca do artigo aqui apresentado.

## 2. Materiais e Métodos

Inicialmente, cogitou-se coletar textos diretamente de perfis de influenciadores digitais de criptomoedas do Twitter em períodos específicos, nos quais temas sobre a criptomoeda Bitcoin estava entre as mais comentadas no Twitter. Embora a seleção de tópicos no Twitter (Trending Topics) para coleta e tratamento de dados seja um método já disseminado [Costa and Marin 2021], existe uma limitação em saber quais tópicos estão, de fato, relacionados ao tema "criptomoeda Bitcoin". Além disto, existe a possibilidade de ignorar-se influenciadores digitais de criptomoedas que disseminaram textos em outros períodos não considerados em nossa coleta de textos, incluindo-se a utilização de palavras-chaves distintas das por nós considerada. Por estes motivos, optou-se por selecionar palavras-chave conhecidas sobre a criptomoeda Bitcoin baseada em palavras-chave do Trending Topics do Twitter e, a partir delas, descobrir os influenciadores digitais desta criptomoeda.

A identificação de tweets sobre a criptomoeda Bitcoin foi feita utilizando a função de busca avançada do Twitter, que é um método de busca que possibilita a adição de parâmetros para um refinamento de resultados<sup>1</sup>. Foi utilizado as palavras-chave "bitcoin" e "btc" para a coleta de textos do Twitter, levando-se em consideração tweets com mais de 100 retweets/compartilhamentos. Este critério sobre retweets foi escolhido como parâmetro devido ao maior potencial de encontrar influenciadores em grande disseminação de mensagens em comparação aos comentários e curtidas, já que estes atos são exibidos para todos os seus seguidores que, também podem realizar a mesma ação.

<sup>1</sup>Twitter Advanced Search: <https://twitter.com/search-advanced?lang=en>

A coleta de textos do Twitter foi realizado utilizando a linguagem de programação Python<sup>2</sup>, juntamente com o webdriver Selenium<sup>3</sup>, capaz de manejar um navegador web de forma iterativa, tal como um usuário humano o faz, possibilitando a realização automática de login no Twitter e da busca avançada, adicionando-se os parâmetros e palavras-chave previamente determinados. Por meio da linguagem de marcação XML(Extensible Markup Language), o Selenium permitiu coletar do Twitter os seguintes valores de cada tweet (cada texto postado): nome de usuário, url, data da postagem, texto e quantidade de likes, retweets e comentários. As informações de identificação dos usuários foram anonimizados, evitando-se assim problemas de privacidade de dados. Ressalta-se a importância deste procedimento, pois o Twitter limita a coleta de textos com mais de 30 dias de publicação, e no máximo de 5 mil tweets por mês. Como não sabíamos a quantidade de textos existentes a partir dos parâmetros por nós selecionados, a utilização do webdriver Selenium foi preferível.

Coletou-se tweets em um período de 2 anos, de 1º de Janeiro de 2020 até 31 de dezembro de 2021, e todos os valores obtidos foram transformados em um arquivo no formato Json e armazenados em um banco de dados não estruturado conhecido como MongoDB<sup>4</sup>. O MongoDB é um software de banco de dados gratuito e não-estruturado, orientado a documentos e de código aberto, que permite a modelagem de informações de modo mais natural, onde os dados são aninhados em hierarquias e continuar a ser indexáveis e de fácil busca. Todos os textos coletados foram tratados por meio da técnica Latent Dirichlet Allocation (LDA) [Blei et al. 2003] de processamento de linguagem natural, permitindo a remoção de acentos e "stopwords", normatização de palavras e letras em minúsculo, remoção de preposições, números e artigos, e a realização da tokenização dos textos, processos feito de forma semelhante a [Castro and Ferreira 2019].O LDA pode ser utilizado por meio da biblioteca gensim<sup>5</sup> do Python.

### 3. Resultados

A coleta dados do Twitter resultou em um total de 18581 tweets (textos) extraídos, dos quais 3.356 são do ano de 2020, e 15.225 tweets do ano 2021. O tweets coletados referem-se a 4.124 usuários distintos, onde a identificação dos influenciadores digitais da criptomoeda Bitcoin foi feita a partir dos seguintes critérios: quantidades de tweets publicados por usuários, levando-se em consideração se a conta é verificada no Twitter(conta autêntica de um indivíduo com interesse público), e a quantidade de seguidores destes perfis.

O grupo de usuários coletados permitiu a observação de que apenas 100 usuários do Twitter foram responsáveis por 10.122 tweets, ou 54% dos textos coletados, como representado no início do eixo horizontal da Figura 1. Nesta Figura, podemos observar uma tendência de função logarítmica, onde poucos usuários (influenciadores digitais da criptomoeda Bitcoin) possuem alto poder de publicação, disseminação e compartilhamento de twittes, enquanto muitos usuário possuem baixo poder de publicação, disseminação e compartilhamento de twittes. Este tipo de comportamento é semelhante ao notado por Alfred Lotka em relação a alta disseminação de artigos científicos: proporcionalmente, pou-

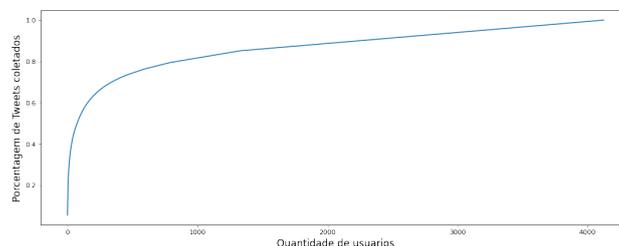
<sup>2</sup>Python versão 3.10.0: <https://www.python.org/downloads/release/python-3100/>

<sup>3</sup>Webdriver Selenium: <https://www.selenium.dev>

<sup>4</sup>MongoDB: <https://www.mongodb.com>

<sup>5</sup>Biblioteca gensim: <https://radimrehurek.com/gensim/>

cos cientistas contribuem fortemente com a disseminação científica (e sua produção), enquanto muitos cientistas pouco contribuem com a disseminação científica. [Lotka 1926].



**Figura 1. Quantidade de Usuários x Porcentagem Tweets**

Os textos coletados do Twitter dos usuários identificados como influenciadores digitais da criptomoeda Bitcoin permitiu a análise dos termos mais utilizados por eles e, agrupá-los. Observou-se de 2 grupos de influenciadores por meio dos termos utilizados: o primeiro grupo está relacionado à discursão sobre o tema envolvendo Bitcoins e criptomoedas (ex: price, Bitcoin, value, money), e o segundo grupo relacionado a termos de incentivo na compra, uso e venda de Bitcoins (ex: good luck, winner, give away, follow).

#### 4. Considerações Finais

Nesse artigo, foi descrito um método de coleta e tratamento de textos do Twitter para a identificação de influenciadores digitais da criptomoeda Bitcoin. Este método permitiu a identificação de 100 influenciadores por meio da coleta de textos durante o período de 2 anos. Observou-se a existência de 2 grupos de influenciadores: aqueles que incentivam a compra, venda e negociação de Bitcoins, e aqueles que querem discutir o tema. Pretende-se, como trabalho futuro, aprofundar a análise das características destes influenciadores, por meio de novas coletas de dados.

#### Referências

- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.
- Bovério, M. A. and da Silva, V. A. F. (2018). Blockchain: uma tecnologia além da criptomoeda virtual. *Revista Interface Tecnológica*, 15(1):109–121.
- Castro, P. C. d. and Ferreira, B. R. (2019). Análise de sentimentos de tweets para identificar tendências de bitcoins no mercado financeiro.
- Costa, P. H. R. and Marin, L. H. G. (2021). Sistema para coleta e tratamento textos brasileiros sobre polarização política. In *Anais Estendidos do XVII Simpósio Brasileiro de Sistemas de Informação*, pages 01–04. SBC.
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington academy of sciences*, 16(12):317–323.
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review*, page 21260.
- Uzunoglu, E. and Kip, S. M. (2014). Brand communication through digital influencers: Leveraging blogger engagement. *International journal of information management*, 34(5):592–602.