Utilizando Métricas de Centralidade para Analisar a Distribuição de Riqueza em Transações da Blockchain do Bitcoin

Guilherme A. Scheibe[1], Humberto T. Marques-Neto[1]

[1]Programa de Pós-Graduação em Informática
Departamento de Ciência da Computação
Pontifícia Universidade Católica de Minas Gerais – PUC Minas
Belo Horizonte – MG – Brasil

gascheibe@sga.pucminas.br, humberto@pucminas.br

Abstract. *This study analyzes the application of four centrality metrics in the complex network formed by high-value transactions in the Bitcoin blockchain over a period of 245 days between Jun/2022 and Feb/2023, with the goal of identifying the most important nodes for wealth distribution. The results indicate that addresses that maintain their centrality and importance over time are mainly attributed to large exchanges and custodial entities, which are described in detail in the article. Additionally, the study reveals that between 91% and 94% of the addresses considered central during the analyzed period do not repeat in the results. The study also shows parity among the top results for three centrality metrics used.*

Resumo. *Este estudo investiga a aplicação de quatro métricas de centralidade na rede complexa formada pelas transações de alto valor na blockchain do Bitcoin durante um período de 245 dias entre Junho de 2022 e Fevereiro de 2023, com o objetivo de identificar os nodos mais importantes para a distribuição de riqueza. Os resultados indicam que os endereços que mantêm sua centralidade e importância ao longo do tempo pertencem, principalmente, a grandes corretoras e entidades de custódia, que são descritas em detalhes no artigo. Além disso, o estudo revela que entre 91% e 94% dos endereços considerados centrais durante o período analisado não se repetem nos resultados. O estudo também mostra uma paridade nos principais resultados entre três métricas de centralidade utilizadas.*

1. Introdução

Bitcoin é a primeira criptomoeda conhecida e possui atualmente o maior valor de mercado, sendo aproximadamente 450 bilhões de dólares segundo dados do *CoinMarketCap*[1]. Dentre os 9015 criptoativos existentes e catalogados pelo *CoinMarketCap*, o Bitcoin possui um fator de dominância de cerca de 44% de todo o mercado de criptoativos. Com uma média diária de transações dentro da *blockchain* (*on-chain*) acima de 250 mil, que se relacionam a uma média de 650 mil endereços, o Bitcoin tem grande relevância no mercado de criptomoedas, atuando como um direcionador, visto

[1] https://www.coinmarketcap.com - *Today's Cryptocurrency Prices by Market Cap [5 fev. 2023]*

que suas movimentações de preço invariavelmente afetam os demais criptoativos do mercado [Peshov et al., 2023].

A proposta do Bitcoin foi apresentada por Nakamoto em 2008 e a rede lançada oficialmente em 2009. Todas as transações são armazenadas em um registro público, chamado de livro razão (*ledger*), por meio de uma estrutura que foi posteriormente denominada *blockchain*. Essa estrutura é uma sequência de blocos encadeados, gerados aproximadamente a cada dez minutos, sendo que cada bloco lista as transações processadas com sucesso. Todos os dados obtidos por meio da análise da *blockchain* são chamados de dados *on-chain.*

Com o crescente interesse no Bitcoin e na tecnologia *blockchain*, muitos pesquisadores têm se concentrado em analisar a rede de transações do Bitcoin. A análise da rede de transações pode ajudar a entender o comportamento dos usuários da rede, identificar padrões e anomalias e descobrir possíveis vulnerabilidades. É importante ressaltar que, apesar dos registros das transações do Bitcoin serem públicos e auditáveis, a sua estrutura *blockchain* possui um direcionamento em privacidade. Essa característica gera limitações na análise da rede de transações, as quais são abordadas e esclarecidas adiante.

Neste estudo, foi realizado uma análise da *blockchain* do Bitcoin por um período de 245 dias de 19 de Junho de 2022 a 18 de Fevereiro de 2023, focando em transações de valores iguais ou superiores a 500 BTC e utilizando quatro métricas de centralidade. Foram obtidos acima de 500 mil endereços únicos (nodos) e 3,2 Milhões de relações (arestas) entre eles.

O objetivo do estudo foi identificar endereços relevantes para o funcionamento da rede, bem como explorar o papel das entidades que os detêm. A análise desses nodos nos permitiu observar o fluxo de riquezas, sendo estas as transações de maior valor na rede, e compreender melhor o papel que cada entidade desempenha nesse contexto. Além disso, os resultados confirmaram a baixa reutilização de endereços, o que sugere que os usuários tendem a gerar novas chaves com frequência. Essas descobertas contribuem para uma melhor compreensão do Bitcoin como um sistema distribuído de transferência de valor e podem ter implicações para o desenvolvimento futuro e uso da tecnologia *blockchain*.

Esse trabalho está dividido da seguinte forma: a seção 2 apresenta os trabalhos relacionados. A seção 3 caracteriza o processo transacional na *blockchain* do Bitcoin. A seção 4 apresenta a metodologia utilizada para construção da rede e análise dos dados. A seção 5 apresenta a análise dos resultados e finalmente a seção 6 apresenta as conclusões do trabalho.

2. Trabalhos Relacionados

Silva et al. [2018] apresentam um estudo de caracterização da rede Bitcoin no qual identifica uma grande concentração de riqueza, sendo que no período da análise 90% dos Bitcoins convergiram para apenas 1,7% dos endereços. Além disso, foi observado que 85% dos endereços apresentam apenas duas transações, o que pode indicar um comportamento de migração dos Bitcoins entre endereços próprios. Um dado analisado e de relevância para o estudo está relacionado à caracterização das transações na rede, onde é feita a distribuição de transações por faixa de valor.

Emery e Latapy [2021] conduziram uma análise ampla da *blockchain* do Bitcoin, capturando e processando todos os dados disponíveis. Embora as técnicas utilizadas para a obtenção de dados não tenham sido replicadas neste artigo, algumas estatísticas foram revisadas juntamente com as de Silva et al. [2018], por possuírem dados mais atualizados.

O estudo de Ho et al. [2020] apresenta uma análise das redes de 120 criptomoedas, incluindo o Bitcoin, com foco na utilização de medidas de centralidade para previsão de preço de curto prazo. Os resultados indicam que as métricas de centralidade são úteis na identificação de nodos importantes ou influentes na rede. Além disso, o estudo também investiga a correlação entre os movimentos de diferentes criptomoedas com base nos resultados obtidos na análise da rede.

Tao et al. [2022] realizam uma análise de transações do Bitcoin utilizando métricas de redes complexas, incluindo a distribuição por grau (*degree distribution*), coeficiente de agrupamento (*clustering*), comprimento do caminho mais curto, componentes conectados, centralidade, associatividade e coeficiente do clube dos ricos (*rich-club coefficient*). Para gerar a rede, os autores utilizaram uma técnica para distribuir os pesos entre os endereços de entrada e saída de cada transação.

Kondor et al. [2014] analisam a estrutura das transações do Bitcoin através de métricas e características da rede, incluindo a distribuição por grau, correlações entre os graus e agrupamento.

Di Francesco Maesa et al. [2017] propõem um algoritmo de agrupamento (*clustering*) e utiliza diversas métricas de centralidade, entre outras análises, para identificar propriedades da rede de transações do Bitcoin. Os resultados obtidos confirmam a presença de nodos centrais, que estão relacionados a nodos populares.

Esse trabalho busca expandir o uso das métricas de centralidade com o foco no estudo dos nodos importantes em transações de alto valor, não se limitando a caraterização global da *blockchain* ou utilização as métricas como base para aplicação de outros conceitos. Além disso, o trabalho possui uma visão mais atualizada em relação aos conceitos e resultados obtidos na literatura, considerando que vários dos trabalhos que relacionam as métricas de centralidade com a *blockchain* do Bitcoin são prévios a 2018.

3. Rede de Transações Bitcoin

Essa seção apresenta o processo transacional e de relação de endereços em transações da *blockchain* do Bitcoin, considerando os aspectos relevantes ao entendimento do trabalho e as decisões tomadas.

A *blockchain* do Bitcoin é uma rede descentralizada que registra e valida transações. Conforme representado na Figura 1, os blocos são as unidades básicas de armazenamento de dados no qual cada bloco é composto por um cabeçalho e uma lista de transações. O cabeçalho contém metadados, incluindo o *hash* do bloco anterior, um *timestamp* e um número aleatório chamado *nonce*, gerado através do processo de mineração. A lista contém as transações confirmadas e adicionadas ao bloco [NAKAMOTO, 2008].

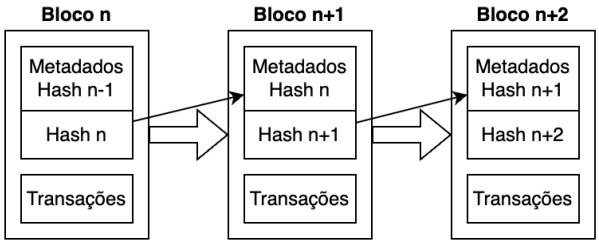

Figura 1. Representa o encadeamento dos blocos na *blockchain* do Bitcoin. Exceto o primeiro (Bloco Genesis), cada bloco contém o *hash* do bloco anterior e são gerados, aproximadamente, a cada 10 minutos, produzindo em média 144 blocos por dia.

3.1. Transações

As transações são compostas de entradas e saídas, onde cada entrada se refere a um conjunto de Bitcoins previamente recebidos pelo endereço do usuário. A *blockchain* não mantém um registro do saldo total de um endereço. Cada transação recebida é registrada e representa uma saída de uma transação anterior que não foi gasta, chamada de UTXO (*Unspent Transaction Output*). Cada UTXO é uma unidade indivisível e que precisa ser gasta em sua totalidade. Vários UTXOs podem ser usados como entrada de uma mesma transação para compor o valor total de saída. Caso a soma dos UTXO ultrapasse o valor transferido, somado à taxa de mineração, então a diferença é transferida para o emissor da transação criando um novo UTXO chamado de troco (*change*).

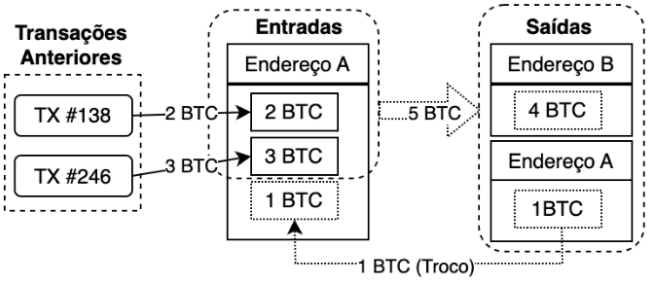

Figura 2. Exemplo de uma transação na *blockchain* do Bitcoin. O endereço A possui dois UTXO de 2 e 3 BTC. Para realizar uma transação de 4 BTC para o endereço B, ambos UTXO são incluídos na transferência, criando um novo UTXO de 4 BTC em B e um UTXO de 1 BTC em A, que é recebido como troco.

Outras duas características da *blockchain* do Bitcoin que foram exploradas e que impactam na metodologia desse estudo foram as propostas de melhorias do Bitcoin BIP0032 [Wuille, 2012] e BIP0044 [Palatinus e Rusnak, 2014]. O BIP0032 define o padrão de geração de chaves privadas e endereços públicos derivados de uma única semente, com base em uma hierarquia determinística. Isso permite que um usuário gere infinitos endereços públicos a partir de uma semente, o que provê segurança e principalmente privacidade nas transações. No exemplo da Figura 2, ao invés de enviar o troco de volta para o Endereço A, o portador da carteira do endereço A poderia enviar esse valor para um endereço C de sua propriedade, derivado da sua carteira, não sendo possível identificar se algum dos endereços de destino pertence ao emissor da transação.

Já o BIP0044 expande os conceitos implementados no BIP0032 através de regras para os caminhos de derivação das chaves, incluindo diferentes tipos de contas. Nessa perspectiva, quando implementado pelas carteiras Bitcoin, as melhorias permitem, por exemplo, que cada transação de um mesmo usuário gere um novo endereço de retorno para o UTXO de "troco" ou que uma carteira realize transferências ou consolidações entre endereços que são de uma mesma carteira. Esse comportamento, quando considerado como uma rede complexa, gera nodos (endereços) e arestas (transações) que afetam a estrutura da rede, porém não afetam a posse ou a propriedade dos Bitcoins transferidos.

De acordo com os estudos de Silva et al. [2018], no conjunto de transações analisadas, 85% dos endereços possuíam apenas duas transações, sendo uma entrada e uma de saída, o que pode indicar o uso de carteiras determinísticas, que automatizam a implementação da hierarquia determinística, de forma que os endereços utilizados não são reaproveitados.

Outra característica que foi considerada na análise da rede de transações é a possibilidade de cada transação possuir diferentes endereços de entrada e saída, sendo que a soma de todas as entradas deve ser igual a soma de todas as saídas, incluindo a recompensa do minerador, conforme a Figura 3(a).

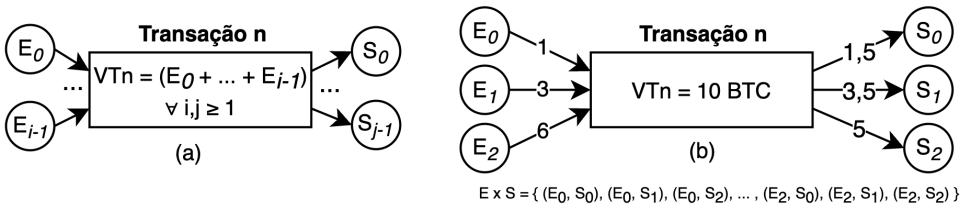

Figura 3. Transações na *blockchain* do Bitcoin. (a) Uma transação pode possuir *i* ≥ 1 entradas e *j* ≥ 1 saídas. VTn = Valor Total da Transação n, representando a soma de todas entradas. (b) Relação E x S, onde cada elemento do conjunto é atribuído a uma aresta direcionada no grafo da rede de transações

A relação entre as entradas e saídas é determinada pelo produto cartesiano Entradas x Saídas. Quando i = 1, é possível identificar através do valor de cada saída o quanto em BTC foi transferido para cada endereço, determinando assim o peso de cada aresta. Quando j =1, é possível identificar o peso de cada aresta através do valor em BTC de cada entrada. Conforme representado na Figura 3(b), quando i,j > 1, não é possível determinar uma relação entre de i → j de peso, uma vez que os valores de todas as entradas são somados e consolidados pela transação e depois distribuídos entre os valores de saída.

Há uma exceção para a regra de i ≥ 1, na qual cada bloco possui uma transação chamada de *coinbase*, utilizada para remunerar o minerador responsável por gravar o bloco na *blockchain*. Como a recompensa é gerada como parte da mineração, essa transação não possui um endereço de entrada.

4. Metodologia

Essa sessão descreve a metodologia utilizada para construção da rede de transações e os critérios adotados. A Metodologia também contempla as ferramentas utilizadas e os métodos de análise empregados na rede de transações.

4.1. Construção da Rede

Entre as referências revisadas, os dados para análise foram obtidos através da coleta dos blocos da *blockchain* do Bitcoin através de APIs em sites como o *Blockchain.com* [Silva et al., 2018]. Outros autores utilizaram técnicas de obtenção de dados através de um *Full Node* da *blockchain* (bitcoin-core), realizando consultas RPC e processamento dos arquivos de dados [Emery e Latapy, 2021] [Ebrahimi e Babveyh, 2018] ou através de dados pré-processados [Pereira e Couto, 2022] [Di Francesco Maesa et al., 2017].

Para este trabalho, foi utilizado o portal Luabase (luabase.com) que processa os dados da *blockchain* do Bitcoin, armazenando em tabelas em uma base de dados relacional. Esse processo simplifica tanto a obtenção quanto a correlação entre os dados, uma vez que pode ser feita através de consultas SQL customizadas, acessadas através de API. A obtenção dos dados foi realizada através de ferramentas desenvolvidas para acesso e consulta a sete APIs criadas especificamente para esse estudo, relacionando as informações de blocos, transações de entrada, transações de saída e transações consolidadas. Durante a análise, outros sites como o *Blockchain.com* e um *full* node (Bitcoin-core) foram consultados para confirmar a exatidão dos dados do Luabase. Durante a execução da pesquisa, após a metade do mês de fevereiro de 2023, o projeto Luabase foi descontinuado e posteriormente desativado, não podendo mais ser utilizado como ferramenta de obtenção de dados.

Para este estudo, foram obtidas todas as transações realizadas entre o período de 19 de junho de 2022 a 18 de fevereiro de 2023, totalizando 245 dias ou aproximadamente 8 meses. Foi constatado um volume transacional diário médio de 250 mil transações e 650 mil endereços. Estudos foram realizados considerando diferentes valores mínimos transacionais, a partir de 0 a 1000 BTC, onde foi observado uma estabilidade no volume transacional a partir de 500 BTC, refletindo na quantidade de endereços e relações.

Com o objetivo deste estudo de buscar uma análise dos nodos de maior influência e limitar o volume de dados para execução dos algoritmos de centralidade, foram consideradas todas as transações onde o $VT_n \geq 500$ BTC (Valor total da transação) e o produto cartesiano E x S de todos os endereços de entrada e saída de cada transação. Foi obtida, também, uma quantidade média de 2225 endereços únicos (nodos) por dia, sendo a quantidade média de relações (arestas) entre os endereços de 13231. Os maiores valores obtidos no período para um dia foram de 12390 endereços, em 13/11/2022, durante o período de instabilidade do setor de criptomoedas e posterior interrupção dos serviços da corretora de criptoativos FTX. Em 20/12/2022 foram observadas 240 mil relações, porém não foi identificado um fator único de impacto no mercado nas datas próximas ao evento. Esses movimentos estão indicados pelas setas na Figura 4(1)(2).

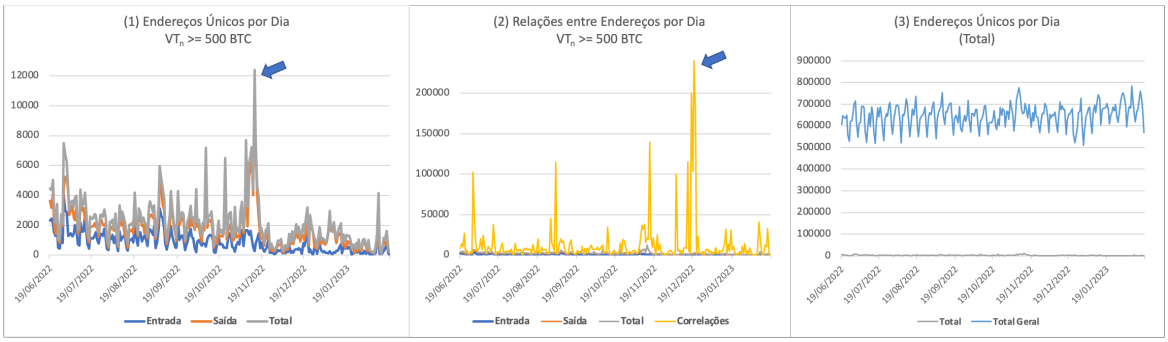

Figura 4. Consolidação dos dados obtidos para cada dia no período analisado. (1) Endereços únicos de entrada, saída e total. (2) Total de relações de endereços (E x S) existentes por transação. (3) Valor total de endereços por dia, sem restringir o valor mínimo da transação.

Para a construção da rede, dois critérios adicionais foram considerados, sendo o primeiro a exclusão das transações do tipo *coinbase*, uma vez que elas não possuem endereço de origem (i = 0) e não representam uma transferência de ativos BTC. Já o segundo, dentro de uma transação T, foram excluídas a relações onde $E_i = E_j$ uma vez que elas representam o troco de uma transação em forma de *loop* e não uma movimentação de valores.

4.2. Análise de Dados

Para realizar o processamento das métricas de centralidade nas redes de transações diárias, foi utilizada a ferramenta *NetworkX* [Hagberg et al., 2008], que é uma biblioteca que permite criar, manipular e estudar estruturas de redes complexas. O *NetworkX* permite a criação de diferentes tipos de grafos baseados nos nodos e arestas. Para o estudo foi considerado o formato *DiGraph*, que representa um grafo direcional sem arestas paralelas. Para análise de grafos individuais e análise visual dos resultados, foi utilizada a ferramenta *Gephi* [Bastian et al., 2009].

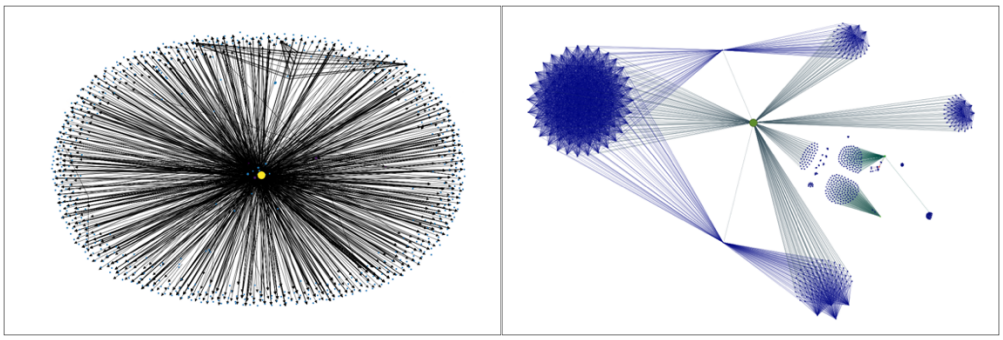

Figura 5. Dígrafo da rede complexa de transações em 11 Fev 2023, contendo 649 nodos e 3916 arestas. Em destaque os nodos mais centrais pela métrica de *Betweenness*. (1) Imagem gerada pelo *NetworkX* com *matplotlib*. (2) Imagem gerada pelo *Gephi*.

No exemplo da Figura 5, a análise visual permite uma melhor identificação das características da rede identificando por exemplo comunidades isoladas que foram incluídas nos resultados.

Para análise da rede, foram utilizadas quatro métricas de centralidades conhecidas: *Degree*, *Betweenness*, *Closeness* e *Eigenvector*. Para cada dia, foram computados os 5 nodos (endereços) mais centrais por cada uma das métricas, conforme a Tabela 1.

			DEGREE CENTRALITY									
Data	Nodos	Arestas	DN1	D1	DN2	D2	DN3	D3	DN4	D4	DN5	D5
19/06/22	4466	5556	3PBCQSPUrVn	0.0394176931	396kMsJ6Ppun	0.0360582306	3ETfLHAAtVST	0.0315789473	19mLB2hCG8V	0.0313549832	bc1qm34lsc65	0.0279955207
20/06/22	4343	12855	19sgxbbNcWg	0.0877475817	bc1quhznutcgv	0.0806080147	1LNHExQfnW8	0.0490557346	bc1qm34lsc65	0.0435283279	1Kr6QSydW9b	0.0423767848
21/06/22	5045	8626	3HLqQPZryAcc	0.1825931800	bc1qm34lsc65	0.0693893735	3QSPK5fyNPw	0.0418318794	bc1qcrkt7ndc3	0.0344964314	1Kr6QSydW9b	0.0340999206
22/06/22	2997	25357	337KKMeRifq	0.0687583444	bc1qf6845d7d	0.0677570093	bc1qrks09j3kh	0.0670894526	bc1qc8lywzn8	0.0670894526	3CSiUBSxpB3K	0.0670894526
23/06/22	2165	2830	bc1qm34lsc65	0.1025878003	3Bq8dkxz6Vz6	0.0563770794	35HzAbGmncK	0.0563770794	1Kr6QSydW9b	0.0512939001	3AUfBSmBVo	0.0485212569

Tabela 1. Amostra do resultado da análise de *Degree Centrality*. DNx representa o nodo (endereço público) em formato *hash* RIPEMD-160 enquanto Dx representa o valor de centralidade correspondente para aquele endereço naquele dia. O valor de x varia de 1 a 5, sendo 1 o nodo de maior centralidade.

A Tabela 1 demonstra como os resultados das métricas de centralidade foram calculados para cada dia, incluindo os 5 nodos mais relevantes. Através do exemplo, é possível identificar uma repetição de alguns nodos que permanecem centrais em diferentes datas.

5. Resultados

Um nodo com alta centralidade pode indicar sua importância ou influência na rede [Newman, 2003]. Os resultados foram analisados considerando, primeiro, quantas vezes o mesmo endereço resulta entre os cinco mais centrais para cada uma das métricas utilizadas.

	DEGREE		BETWEENESS		CLOSENESS		EIGENVECTOR		GLOBAL	
	Endereço	Soma	Endereço	Soma	Endereço	Soma	Endereço	Soma	Endereço	Soma
1	NODE A	157	NODE A	147	NODE A	73	NODE D	55	NODE A	381
2	NODE B	146	NODE B	94	NODE B	44	NODE F	25	NODE B	288
3	NODE C	101	NODE C	32	NODE C	24	NODE W	14	NODE C	167
4	NODE H	21	NODE E	19	NODE G	10	NODE X	14	NODE D	73
5	NODE E	18	NODE G	18	NODE I	10	NODE C	10	NODE E	42
6	NODE K	17	NODE D	11	NODE S	9	NODE Y	9	NODE F	35
7	NODE L	12	NODE O	11	NODE J	8	NODE Z	8	NODE G	35
8	NODE M	12	NODE P	9	NODE T	8	NODE 1	8	NODE H	23
9	NODE N	11	NODE Q	8	NODE U	6	NODE 2	7	NODE I	22
10	NODE I	10	NODE R	8	NODE V	6	NODE J	6	NODE J	19

Tabela 2. Os dez nodos (endereços) mais recorrentes nos resultados de cada métrica de centralidade analisada e a consolidação global de todas as métricas. Os endereços foram substituídos do formato *hash* RIPEMD-160 para "NODE A-Z" para facilitar a leitura. A escala de cores é usada visualização dos nodos repetidos entre as métricas e o global.

Na Tabela 2, a Soma indica a quantidade de vezes que cada endereço foi computado como um dos os 5 mais centrais em cada dia dentro do período analisado. A análise realizada através das métricas de *Degree* e *Betweenness* indicaram uma maior centralização da rede de transações onde uma quantidade menor de endereços atuam como *hub* de conectividade. Outro ponto observado foi que os três endereços de maior recorrência são os mesmos entre as métricas de *Degree*, *Betweenness* e *Closeness*. Esses resultados podem indicar uma alta centralização da rede, onde os nodos de maior grau também são considerados os principais caminhos e pontos de distribuição de transações.

Já a análise através do *Closeness* e *Eigenvector* observou em uma maior distribuição dos nodos centrais, entretanto, para todas as métricas de centralidade foi gerado o mesmo padrão gráfico, considerando todos os resultados e não, apenas, os dez primeiros, conforme pode ser observado na Figura 6.

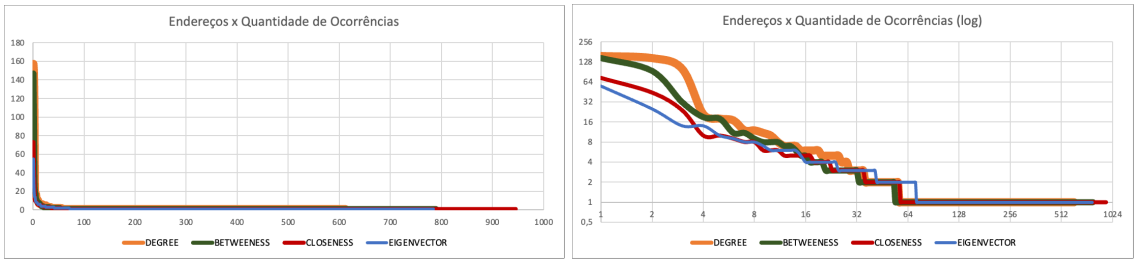

Figura 6. Total de ocorrências para cada endereço computado por cada métrica de centralidade em escala normal e logarítmica (log_2). O eixo X representa o número de endereços distintos registrados e o eixo Y representa a quantidade de vezes que cada endereço foi identificado como um dos mais centrais.

Ao consideramos a ocorrência dos endereços agregando os valores entre as diferentes métricas de centralidade, pode-se observar o mesmo padrão, que resulta em uma quantidade de endereços pequena com maior centralidade em todo o período observado.

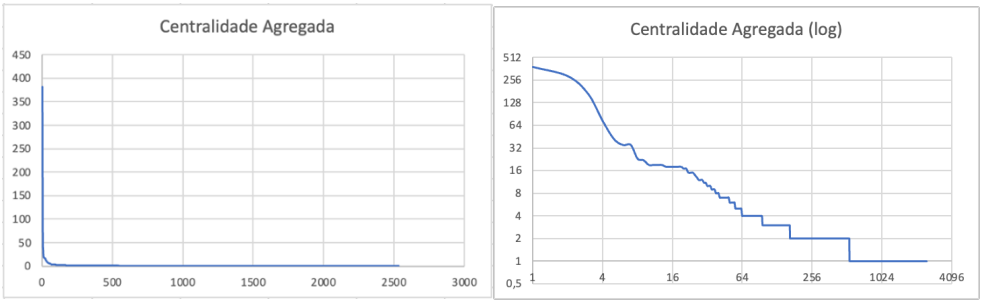

Figura 7. Global - Relação entre os endereços e o total de ocorrências de cada um. Gráfico em escala normal e logarítmica para melhor visualização da curva.

O estudo também permite observar a distribuição da centralidade entre os endereços. Através dos resultados, é possível verificar que entre 91% a 94% dos endereços classificados como centrais pelas métricas utilizadas dentro do período de análise foram observados apenas uma vez. Conforme observado na Tabela 3, de 2539 endereços distintos classificados pelas métricas como centrais, 1998 (78,7%) tiveram apenas uma ocorrência enquanto 375 (14,8%) tiveram duas. Com isso, apenas 6,5% dos endereços foram observados 3 ou mais vezes.

DEGREE	Ocorrências	1	2	3	4	5	6	7	8	10	11	12	17	18	21	101	146	157
	Soma	556	21	7	3	6	5	3	1	1	1	2	1	1	1	1	1	1

BETWEENNESS	Ocorrências	1	2	3	4	5	6	7	8	9	11	18	19	32	94	147
	Soma	736	21	12	4	2	1	2	3	1	2	1	1	1	1	1

CLOSENESS	Ocorrências	1	2	3	4	5	6	8	9	10	24	44	73
	Soma	888	22	13	5	6	3	2	1	2	1	1	1

EIGENVECTOR	Ocorrências	1	2	3	4	5	6	7	8	9	10	14	25	55	180
	Soma	713	30	17	9	1	5	1	2	1	1	2	1	1	1

AGREGADO	Ocorrências	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	17	18	19	22	23	35	42	73	167	180	288	381
	Soma	1998	375	70	32	8	6	9	3	3	3	2	3	1	1	3	2	6	4	1	1	2	1	1	1	1	1	1

Tabela 3. Cada métrica de centralidade apresenta a quantidade de endereços N que obtiveram um determinado número de ocorrências. Tomando como exemplo a tabela para a métrica DEGREE, podemos observar que 556 endereços foram classificados como centrais apenas uma vez, enquanto 21 endereços foram classificados como centrais duas vezes. O AGREGADO representa a análise baseada no total de resultados para cada métrica.

Esse resultado observado pode indicar o baixo reaproveitamento de endereços e a geração de novas chaves para os mesmos usuários e carteiras, conforme apontado em estudos de referência [Kondor et al., 2014] [Ebrahimi e Babveyh, 2018] [Chan et al.,2020] [Di Francesco Maesa et al., 2017].

5.1. Análise dos Endereços (Nodos) Mais Centrais

Os nodos de maior relevância para as métricas de centralidade, que obtiveram uma maior ocorrência global, conforme listado Tabela 2, foram analisados de forma individual a fim de identificar suas propriedades e funções. Embora existam limitações para rastrear as atividades na *blockchain* do Bitcoin, técnicas como o uso de diagramas aluviais permitem uma representação visual da relação entre as carteiras [Remy et al., 2017]. Além disso, algumas ferramentas conhecidas e disponíveis on-line, na Internet, catalogam os principais endereços. Para identificação destes, foi utilizado o site https://oxt.me (*O X T - THE BLOCKCHAIN BY THE PEOPLE FOR THE PEOPLE*). Para análise dos diagramas aluviais foi utilizado o site https://kycp.org (*Know Your Coin Privacy*).

	GLOBAL			
	Endereço	Soma	Hash	OXT
1	NODE A	381	1Kr6QSydW9bFQG1mXiPNNu6WpJGmUa9i1g	BITFINEX (WALLET B)
2	NODE B	288	bc1qm34lsc65zpw79lxes69zkqmk6ee3ewf0j77s3h	BINANCE (HOT WALLET)
3	NODE C	167	19iqYbeATe4RxghQZJnYVFU4mjUUu76EA6	RENBTC
4	NODE D	73	bc1quq29mutxkgxmjfdr7ayj3zd9ad0ld5mrhh89l2	GEMINI (WALLET A)
5	NODE E	42	38MfJkxPkpwsZxkyfESwjdaKBdXrV8DpWr	COINBASE
6	NODE F	35	bc1qmtl499lclce4gvjcgn38gehmmkh2kekywkc273gh	GEMINI (CUSTODY)
7	NODE G	35	3E2adcep2NRRpriLnWn1AvW3AHKqBx2mMr	KRAKEN (WALLET C)
8	NODE H	23	bc1qmgj3w0aw5455y9s4zfhts2kxm4qstwyjx5f907	PRIME TRUST (Unconfirmed)
9	NODE I	22	bc1qfu6su3qz4tn0et634mv7p090a0cgameq6rdvuc	SWISSBORG (WALLET B)
10	NODE J	19	19aaLsPkiJuFZck7U4mryKFiUg633UJDhm	BINANCE (WALLET U)

Tabela 4. Resultados globais apresentados na Figura 7, incluindo o nome da entidade responsável pelo endereço com base no site OTX.

Entre os endereços identificados, é possível verificar que eles pertencem de forma majoritária a corretoras de criptomoedas, incluindo também instituições de custódia e uma ponte (*bridge*) entre a *blockchain* do Bitcoin e outra rede. Esse resultado está alinhado com a importância dessas instituições e o papel que elas desempenham dentro do mercado e do ecossistema das criptomoedas.

6. Conclusão

O estudo desenvolvido nesse trabalho apresenta uma análise de centralidade na rede de transações da *blockchain* do Bitcoin no período de Jun/2022 até Fev/2023, totalizando 245 dias analisados. Para cada dia, foi construída a rede complexa formada pelas transações com valor mínimo de 500 BTC, assim como a relação entre todos os endereços de entrada e saída. Para cada dia, foram obtidos os cinco endereços mais centrais de acordo com as quatro métricas de centralidade analisadas, sendo os resultados avaliados de forma individual, comparados e consolidados.

Os resultados obtidos indicam uma forte paridade entre as métricas de *Degree*, *Betweenness* e *Closeness* dentro da rede de transações, uma vez que os três endereços mais reportados para essas métricas, em diferentes dias, foram os mesmos. Essa paridade indica uma forte centralização onde os nodos principais são ao mesmo tempo os mais conectados além de atuarem como uma ponte entre os demais nós. Além disso, considerando a métrica de *closeness*, é indicado que esses nodos são também centrais de forma geodésica. Uma remoção desses nodos, pode gerar uma redistribuição significativa da rede e gerar componentes desconectados.

Entre todas as métricas analisadas, 91% a 94% dos endereços foram identificados como centrais apenas uma vez, sendo que o baixo índice de repetição nos resultados pode estar relacionado ao uso de carteiras determinísticas que, como melhor prática, utilizam endereços únicos a cada transação. Considerando o agregado entre todas as quatro métricas de centralidade, 79% dos endereços foram considerados centrais apenas uma vez.

Foram analisados de forma individual os dez principais endereços que se mantém como centrais e importantes ao longo do período do estudo, sendo majoritariamente identificados como pertencente a corretoras de criptomoedas e instituições de custódia ou pontes (*bridges*), como a RENBTC que realiza a custódia dos Bitcoins quando migrados para seu protocolo próprio. Esse resultado indica uma forte centralização sobre essas instituições e atuam como pontos de distribuição de riqueza na rede.

Contudo, é possível que trabalhos futuros possam aperfeiçoar a metodologia apresentada, analisando outros cenários, como a correlação entre diferentes faixas de valores transacionais e o impacto resultante do peso atribuído às arestas. Ademais, outros trabalhos podem contribuir para o desenvolvimento de metodologias que permitam atribuir os pesos das arestas nas transações com múltiplos endereços de entrada e saída (E x S). Também se considera como trabalhos futuros relacionar as movimentações com transações escusas, tais como as provenientes de ações fraudulentas. Outro estudo que pode ser realizado é a correlação entre as métricas de centralidade e a variação de preço do Bitcoin.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e da FAPEMIG (PPM-00253-18).

Referências

Nakamoto, S. and Bitcoin, A. (2008). A peer-to-peer electronic cash system. Bitcoin.–URL: https://bitcoin.org/bitcoin.pdf, 4.

Silva, C., Ramos, B., Oliveira, S., & Piccoli, R. (2018). Caracterização da Rede Bitcoin: Uma Visão sobre a Evolução de Blocos, Transações, Endereços e Saldos de 2009 até 2017. In *Anais do I Workshop em Blockchain: Teoria, Tecnologias e Aplicações*. Porto Alegre: SBC.

Emery, J. A., & Latapy, M. (2021). Full Bitcoin blockchain data made easy. In *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 240-243).

Silva, L. P. Q., de Araújo, A. P., Cota, D. O., Cota, G. O., & Antonio, A. D. A. (2021). Utilizando HMM para previsao de preço e estratégia de investimento em criptomoedas BitCoin. In *Anais do IV Workshop em Blockchain: Teoria, Tecnologias e Aplicações* (pp. 134-147). SBC.

Ho, K. H., Chiu, W. H., & Li, C. (2020). A Short-Term Cryptocurrency Price Movement Prediction Using Centrality Measures. In *2020 International Conference on Data Mining Workshops (ICDMW)* (pp. 369-376). IEEE.

Pereira, D. M., & Couto, R. S. (2022). Using Degree Centrality to Identify Market Manipulation on Bitcoin. In *Data Privacy Management, Cryptocurrencies and Blockchain Technology: ESORICS 2021 International Workshops, DPM 2021 and CBT 2021, Darmstadt, Germany, October 8, 2021, Revised Selected Papers* (pp. 208-223). Cham: Springer International Publishing.

Tao, B., Dai, H. N., Wu, J., Ho, I. W. H., Zheng, Z., & Cheang, C. F. (2021). Complex network analysis of the bitcoin transaction network. *IEEE Transactions on Circuits and Systems II: Express Briefs*, *69*(3), 1009-1013.

Kondor, D., Pósfai, M., Csabai, I., & Vattay, G. (2014). Do the rich get richer? An empirical analysis of the Bitcoin transaction network. *PloS one*, *9*(2), e86197.

Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, *74*(1), 47.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2), 167-256.

Ebrahimi, M. S., & Babveyh, A. (2018). Predicting User Performance and Bitcoin Price Using Block Chain Transaction Network. *arXiv preprint arXiv:1804.08044*.

Chan, W. K., Chin, J. J., & Goh, V. T. (2020, December). Evolution of Bitcoin addresses from security perspectives. In *2020 15th International Conference for Internet Technology and Secured Transactions (ICITST)* (pp. 1-6). IEEE.

Wuille, P. (2012). Hierarchical Deterministic Wallets. Bitcoin Improvement Proposal 32 (BIP0032). *Bitcoin Github, November*.

Palatinus, M., & Rusnak, P. (2014). Multi-account hierarchy for deterministic wallets. Bitcoin Improvement Proposal 44 (BIP0044). *Bitcoin Github, April*.

Hagberg, A., Swart, P., & S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX (No. LA-UR-08-05495; LA-UR-08-5495). *Los Alamos National Lab.(LANL), Los Alamos, NM (United States)*.

Remy, C., Rym, B., & Matthieu, L. (2018). Tracking bitcoin users activity using community detection on a network of weak signals. In *Complex Networks & Their Applications VI: Proceedings of Complex Networks 2017 (The Sixth International Conference on Complex Networks and Their Applications)* (pp. 166-177). Springer International Publishing.

Di Francesco Maesa, D., Marino, A., & Ricci, L. (2018). Data-driven analysis of bitcoin properties: exploiting the users graph. *International Journal of Data Science and Analytics*, *6*, 63-80.

Peshov, H., Todorovska, A., Marojevikj, J., Spirovska, E., Rusevski, I., Angelovski, G., ... & Trajanov, D. (2023, January). Using Centrality Measures to Extract Knowledge from Cryptocurrencies' Interdependencies Networks. In *ICT Innovations 2022. Reshaping the Future Towards a New Normal: 14th International Conference, ICT Innovations 2022, Skopje, Macedonia, September 29–October 1, 2022, Proceedings* (pp. 76-90). Cham: Springer Nature Switzerland.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *In Proceedings of the international AAAI conference on web and social media* (Vol. 3, No. 1, pp. 361-362).

Simoes, J. E., Ferreira, E., Menasche, D. S., & Campos, C. A. (2021). Blockchain privacy through merge avoidance and mixing services: a hardness and an impossibility result. *ACM SIGMETRICS Performance Evaluation Review, 48*(4), 8-11.

Ferrin, D. (2015). A preliminary field guide for bitcoin transaction patterns. In *Proc. Texas Bitcoin Conf.*.