

Social Media e Rewards: un'analisi del subreddit ethTrader

Tesi di laurea triennale - Informatica umanistica



UNIVERSITÀ DI PISA

Dipartimento di Filologia, letteratura e linguistica

Relatori:

Prof. Barbara Guidi

Dott. Andrea Michienzi (correlatore)

Presentata da: Francesco Maggio

2 febbraio 2023

Indice

1	Introduzione	5
2	Stato dell'arte: Blockchain e Blockchain Online Social Media	9
2.1	La blockchain	9
2.1.1	Origine e riferimenti storici	9
2.1.2	Struttura e caratteristiche	10
2.1.3	Algoritmi di consenso	11
2.2	Bitcoin	14
2.2.1	Ideali, scopi e caratteristiche alla base	15
2.2.2	Funzionamento della rete	15
2.2.3	Le transazioni	17
2.2.4	La validazione	19
2.3	Ethereum	20
2.3.1	Ether	20
2.3.2	Portafogli e conti	21
2.3.3	Dapps	21
2.3.4	Smart contracts	22
2.3.5	Token	23

2.3.6	Exchange	25
2.4	Blockchain Online Social Media	26
3	Reddit e i Community Points	28
3.1	Reddit	28
3.1.1	Karma	29
3.2	Reddit Community Points	30
3.2.1	Distribuzione dei Reddit Community Points	30
3.3	ethTrader e Donuts	32
3.3.1	ethTrader	32
3.3.2	Donuts	32
3.4	Scopo della tesi	34
4	Implementazione	36
4.1	Strumenti	36
4.1.1	Pushshift	36
4.1.2	Python	37
4.1.3	JSON	38
4.1.4	Spacy	40
4.2	Fasi dell'implementazione	42
4.2.1	Raccolta dei dati	42
4.2.2	Organizzazione dei dati	43
4.2.3	Generazione dei grafici	45
5	Analisi e grafici	46
5.1	Dataset	46
5.2	Distribuzione dei dati	50

5.2.1	<i>Comments e Submissions</i>	50
5.2.2	Parole (Wordcloud)	55
5.3	Autori e numero di pubblicazioni	60
5.4	Sentiment analysis	61
6	Conclusioni	66
6.1	Possibili lavori futuri	67

Elenco delle figure

2.1	La struttura di una blockchain.	10
2.2	Rappresentazione schematica di una transazione Bitcoin. . . .	18
2.3	Riassunto schematico delle differenze tra token fungibili e non fungibili.	24
5.1	Distribuzione annuale.	50
5.2	Distribuzione mensile.	51
5.3	Prezzo Donut to USD.	53
5.4	Prezzo Ethereum to USD.	54
5.5	Wordcloud dell'intero subreddit.	57
5.6	Wordcloud per l'anno 2018.	57
5.7	Wordcloud per l'anno 2019.	58
5.8	Wordcloud per l'anno 2020.	58
5.9	Wordcloud per l'anno 2021.	59
5.10	Wordcloud per l'anno 2022.	59
5.11	Grafico <i>Submissions</i> /Autori.	60
5.12	Positività/Negatività per <i>Comments</i> e <i>Submissions</i>	62
5.13	Soggettività per <i>Comments</i> e <i>Submissions</i>	63

Capitolo 1

Introduzione

Ai giorni d'oggi, i Social Media sono utilizzati da persone di ogni età e ceto sociale per le più disparate mansioni, come tenersi aggiornati su ciò che accade nel mondo, mantenere contatti con amici e familiari (fisicamente lontani e non), promuovere la propria attività lavorativa o addirittura direttamente lavorare, e molte altre.

Nell'ottobre del 2022, gli utenti dei Social Media erano 4.74 miliardi, ovvero circa il 59,3% della popolazione mondiale. Degli utenti presenti su Internet, inoltre, più di 9 su 10 utilizzano mensilmente dei Social Media.[4].

Insieme al numero di utenti, aumenta anche il tempo passato sui Social Media: sempre nel 2022, mediamente in una giornata il consumatore medio passa 2,5 ore sui Social Media, visitando 7,2 piattaforme. Il 28% di questi si dichiara inoltre preoccupato per il troppo tempo speso sui Social Media[5]. Notiamo tra i fattori rilevanti anche la pandemia di COVID-19 come evento recente che ha sicuramente aumentato la necessità, ma anche la volontà, di digitalizzarsi per molte persone; durante il lockdown, tantissimi utenti hanno

utilizzato i Social Media per poter rimanere in contatto con le persone a loro care. Viene da sé che dopo essere stati abituati a una comodità simile, che per molti versi semplifica la vita, anche una volta finito il periodo di lockdown la maggior parte delle persone siano rimaste attaccate alle proprie abitudini nel lockdown, soprattutto quelle digitali.

Nella società di oggi, quindi, si può tranquillamente dire che i Social Media rivestono un ruolo di primaria importanza. Non sono però esenti da alcune problematiche importanti: prima fra tutte, nonché quella più mediaticamente diffusa, la questione della privacy e della sicurezza. I Social Media sono strutture estremamente complesse e che raccolgono un grande quantitativo di dati sui propri utenti, dunque accade spesso che essi siano veicoli di trasmissione di informazioni che violano la privacy degli utenti o che mettono a repentaglio la loro sicurezza. Per evitare ciò, un utente dovrebbe stare costantemente attento a cosa autorizza e pubblica, ma per via del modo spesso sbrigativo e disorganizzato in cui gli utenti si approcciano di solito ai Social ciò raramente viene fatto in maniera attenta e precisa. Non aiuta il fatto che è sempre più diffuso il fenomeno di Social Media Bingeing, ovvero l'utilizzo dei Social media per un periodo di tempo superiore al consueto. Altri problemi legati ai Social Media possono essere la diffusione di *fake news*, ovvero notizie false o fuorvianti, l'aumento delle problematiche relative all'ansia e depressione, la possibilità di incappare in truffe sempre più articolate.

Le piattaforme Social più diffuse presentano generalmente un'architettura centralizzata, dunque nelle mani di un'autorità centrale che le possiede, e i dati sono raccolti generalmente su server di proprietà dell'autorità centrale stessa. Questo potrebbe essere un problema perché una struttura simile

difficilmente può gestire i problemi di privacy o della diffusione di *fake news* prima citati, anzi spesso è proprio la natura centralizzata dei Social che permette a queste problematiche di diffondersi: basti pensare al fatto che in un Social centralizzato, un'autorità centrale possiede i dati dei suoi utenti, e potrebbe di conseguenza farne qualsiasi cosa, come venderli a terzi o utilizzarli per i propri scopi. In cerca di un'eventuale soluzione ai questi problemi, si è fatta sempre più forte la possibilità di integrare la tecnologia blockchain nei Social Media. È così che sono nati i BOSM (vedi Capitolo 2.4), Blockchain Online Social Media, ovvero dei Social Media che si distaccano dalla classica visione centralizzata e potenzialmente pericolosa, e che abbracciano il concetto della decentralizzazione per restituire il potere ai propri utenti. Ciò permette cose non possibili sui classici Social Media, come la rimozione quasi completa della censura, la possibilità di contrastare le *fake news* attraverso dei meccanismi in cui gli utenti decidono quale contenuto è valido e quale no, e soprattutto l'aggiunta di sistemi di ricompensa che assegnano premi in criptovaluta agli utenti in base al loro contributo. Questi ultimi sono molto interessanti in quanto una piattaforma come un Social Network che si basa sulla blockchain integra anche una componente economica basata sul concetto di token, con tutti i vantaggi che ne vengono.

Negli ultimi anni, sono nate anche piattaforme di Social Media del tutto decentralizzate come Steemit, ma anche le classiche piattaforme centralizzate, ben diffuse e conosciute, hanno pensato di introdurre dei concetti di decentralizzazione, come ad esempio Reddit con i Reddit Community Points: questi utilizzano la tecnologia della blockchain per gestire un sistema di rewarding che vorrebbe spingere gli utenti a pubblicare contenuti di qualità. L'idea

di base dietro ai Reddit Community Points è che essi fungono nello stesso momento da bene per accedere a funzioni esclusive e da indicatore di reputazione nella community, oltre alle implicazioni economiche della loro esistenza come token sulla blockchain (i Reddit Community Points sono approfonditi al Capitolo 3.2).

A partire da queste considerazioni iniziali abbiamo deciso di attuare una ricerca approfondita su Reddit, e in particolare su un subreddit molto legato all'idea della decentralizzazione e ai suoi progetti: **ethTrader**. Di questo subreddit abbiamo analizzato le tendenze, studiato i trend ricorrenti, la sua attività, ricercato le parole più utilizzate e osservato il sentimento degli utenti. Abbiamo scelto proprio questo subreddit in quanto ritenevamo importante lo studio di un subreddit che avesse già introdotto un meccanismo di Reddit Community Points: ethTrader è stato il primo subreddit a farlo, e per noi questo è stato un ottimo punto di partenza in quanto abbiamo potuto considerare la presenza dei Reddit Community Points (chiamati Donuts in questo specifico subreddit) nelle nostre analisi, anche quelle linguistiche.

Il resto del documento è organizzato nel seguente modo: una presentazione dello stato dell'arte, con panoramica sulla tecnologia blockchain, in particolare Bitcoin ed Ethereum, e una descrizione dei Blockchain Online Social Media. A seguire, la spiegazione dell'oggetto della ricerca, ovvero Reddit e i Reddit Community Points, con annesso focus sul subreddit ethTrader e sui Donuts. Presentiamo poi l'implementazione, specificando strumenti e stesura delle fasi di lavoro. Subito dopo mostriamo le analisi svolte e i grafici ottenuti, con alla fine un resoconto sul lavoro effettuato, una riflessione sulle analisi e una parentesi su eventuali sviluppi futuri.

Capitolo 2

Stato dell'arte: Blockchain e Blockchain Online Social Media

In questo Capitolo presentiamo la tecnologia blockchain, soffermandoci in particolare su quella di Bitcoin, la prima ad avere un successo planetario, e su quella di Ethereum, quella più interessante per noi in quanto è l'argomento d'interesse del subreddit su cui abbiamo fatto la ricerca. Presenteremo anche i *Blockchain Online Social Media*, una nuova tipologia di Social Media che sfrutta la tecnologia blockchain per implementare parte del servizio.

2.1 La blockchain

2.1.1 Origine e riferimenti storici

La prima descrizione della blockchain risale al 1991, a opera di Stuart Haber e W. Scott Stornetta, nella pubblicazione "How to time-stamp a digital

document”[14]. L’anno in cui però la blockchain inizia a suscitare interesse in maniera più estesa è il 2008, quando viene pubblicato il whitepaper della prima moneta decentralizzata, Bitcoin[11], sotto lo pseudonimo tutt’oggi fittizio di Satoshi Nakamoto. A oggi, la blockchain è utilizzata come struttura alla base di moltissime piattaforme e su moltissimi campi, come ad esempio la finanza, ma anche i Social Network.

2.1.2 Struttura e caratteristiche

Una blockchain è formata come una catena di blocchi, contenenti delle transazioni e collegati fra di loro in modo sicuro attraverso la crittografia. Ogni singolo blocco è composto da due parti: un body, il quale racchiude varie transazioni, e un header, contenente invece dei campi di gestione del blocco stesso; tra questi è presente un hash crittografico del blocco precedente, in modo da poter creare la catena collegando ogni blocco al precedente.

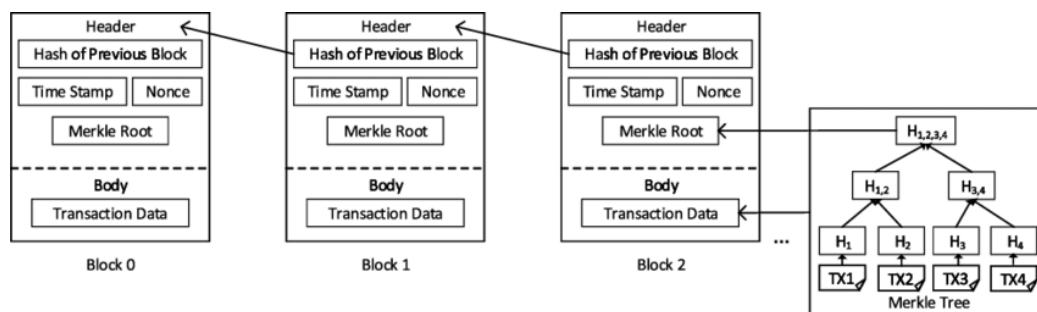


Figura 2.1: La struttura di una blockchain.

Il trilemma della blockchain

Una blockchain dovrebbe avere tre caratteristiche principali:

- *Decentralizzazione*: Non c'è una persona o un'organizzazione al comando, chiunque può partecipare e il controllo non appartiene a un'unità ma è completamente distribuito, dunque tutti i partecipanti hanno accesso agli stessi dati.
- *Sicurezza*: La rete deve essere resistente a eventuali attacchi di entità malintenzionate. Queste ultime potrebbero prendere il controllo della rete, tramite ad esempio il *51% attack*¹, e rendere così la catena inutile.
- *Scalabilità*: La rete deve supportare un numero sempre maggiore di transazioni al secondo. Questa si vede sempre più importante dal momento in cui questa tecnologia deve essere utilizzata per servire un numero molto ampio di utenti.

Il trilemma della blockchain consiste nella difficile coesistenza di queste tre caratteristiche; l'aumento di una di esse porta solitamente all'indebolimento di una delle altre. A oggi, non si può dire di avere una soluzione, ma in poco tempo sono stati fatti molti passi a proposito, e il futuro sembra sempre più roseo.

2.1.3 Algoritmi di consenso

Ruolo principale nell'ecosistema della blockchain lo svolge l'algoritmo di consenso adattato. Questo può essere definito come un meccanismo che permette

¹*51% attack*: un tipo di attacco sulla rete blockchain che si verifica quando una singola entità (o un'organizzazione) arriva in possesso della maggior parte dell'*hash rate* (potenza di calcolo). In questo caso, tale entità sarebbe in grado di rompere l'equilibrio del sistema, ad esempio modificando l'ordine delle transazioni o impedendo che alcune di queste, a sua discrezione, siano confermate.

a utenti e/o dispositivi di coordinarsi in un contesto distribuito, e il suo fine ultimo è quello di aggiungere nuovi blocchi alla catena. L'algoritmo di consenso è molto importante perché l'ambiente della blockchain è fondamentalmente composto da utenti sconosciuti e che non si fidano gli uni degli altri, ma nonostante ciò è importante che tutti gli agenti del sistema possano concordare su una singola fonte di verità. I due algoritmi di consenso principali sono Proof of Work (PoW) e Proof of Stake (PoS).

Proof of Work

L'algoritmo di Proof of Work, per confermare le transazioni e quindi produrre nuovi blocchi della catena, incentiva i miner a competere tra di loro nell'elaborazione attraverso una ricompensa offerta in cambio della loro potenza computazionale: questo è possibile grazie a un problema, ovvero un algoritmo crittografico chiamato appunto Proof of Work. Quest'ultimo è computazionalmente costoso ed è strutturato in modo che l'unica modalità possibile di risolverlo sia andare a tentativi (spiegato più in dettaglio nella sezione 2.2.4, sulla validazione in Bitcoin). Una volta risolto il problema, il nodo ottiene il diritto di inserire il blocco sulla blockchain, e di conseguenza ricevere una ricompensa per incentivare la continuazione del suo lavoro; così facendo, un utente può essere sicuro che tutte le proprie transazioni vengano incluse nella blockchain. La transazione validata rimane poi sulla blockchain pubblica, in modo che chiunque possa vederla.

Un problema di questo sistema, sorto fin da subito, è quello che vede gli utenti con hardware più potenti (e quindi costosi) sempre come i vincitori della ricompensa, in quanto i loro computer vinceranno generalmente la gara

alla soluzione dell'algoritmo. Per ovviare a questo problema sono nate le cosiddette mining pool, ovvero degli spazi che consentono ai minatori di lavorare in modo cooperativo per estrarre blocchi di criptovalute, per poi spartirsi il ricavato. La Proof of Work richiede inoltre una grande quantità di energia elettrica per funzionare, ed è per questo che è stato proposto l'algoritmo di Proof of Stake come alternativa.

Proof of Stake

L'algoritmo di Proof of Stake (PoS) convalida le transazioni e raggiunge il consenso in maniera diversa dall'algoritmo di Proof of Work: mentre quest'ultimo premia il suo minatore più veloce, nell'algoritmo di Proof of Stake il minatore che crea il blocco successivo conquista la transazione proporzionalmente a quanto ha contribuito a minarla. Non c'è quindi un premio fisso, ma semplicemente il guadagno consiste nella commissione della transazione. Questo tipo di modello funziona grazie a dei validatori (*validators*), utenti che garantiscono la validità delle operazioni effettuato impegnano una quota delle proprie criptovalute, "bloccandola" sulla rete, in un processo che prende il nome di *staking*. La dimensione dello *stake* determina la possibilità che un nodo venga selezionato come prossimo validatore (più è alto lo *stake*, più è alta la possibilità), ma questo non è l'unico fattore che influenza la scelta del validatore: ogni criptovaluta che utilizza un algoritmo Proof of Stake ha infatti un proprio set di regole e metodi in modo da poter avere quella che pensa essere la migliore combinazione possibile per la rete. Nel momento in cui i validatori identificano un blocco ritenuto valido, lo verificano piazzando una scommessa su di esso e, se il blocco viene effettivamente aggiunto, i

verificatori coinvolti ricevono una ricompensa in proporzione al valore scommesso. Questo tipo di modello funziona quindi con l'utente che deve inserire le proprie monete (solitamente con un requisito minimo di partecipazione) in un portafoglio specifico, il quale congela le monete e le utilizza per fare "staking" sulla rete; la vincita della ricompensa per quell'utente sarà quindi connessa alla percentuale totale di monete possedute dallo stesso. La scelta del vincitore è comunque casuale in base all'importo puntato, rendendo inutile cercare di alterare la rete, ad esempio elaborando transazioni illegali.

Una differenza importante rispetto alla Proof of Work, in cui la ricompensa per i minatori è nuova valuta appositamente creata, è che nella Proof of Stake la ricompensa per i validatori viene trattenuta dalle transazioni validate. La rete infatti revisiona il lavoro dei validatori accertandosi della loro buona fede, e solo dopo rende possibile il ritiro dello stake depositato e la ricompensa. Un altro vantaggio che la Proof of Stake ha rispetto alla Proof of Work è la possibilità di essere utilizzata per pagamenti elettronici in tempo reale, cosa non possibile per la Proof of Work in quanto la complessità dei suoi problemi matematici rende necessario del tempo prima della loro risoluzione e di conseguenza della creazione del blocco.

2.2 Bitcoin

Bitcoin è una moneta virtuale creata da uno o più persone sotto lo pseudonimo di Satoshi Nakamoto. Essa si basa sulla coesistenza di un network peer-to-peer, ovvero un tipo di rete nella quale i vari nodi si scambiano informazioni in modo paritario e senza la presenza di un server centralizzato,

e l'uso della crittografia per validare e rendere sicure le transazioni. In questo Capitolo andremo a spiegare le caratteristiche alla base di Bitcoin, il funzionamento della sua rete, le transazioni e la validazione di esse.

2.2.1 Ideali, scopi e caratteristiche alla base

La pubblicazione del whitepaper di Bitcoin citato nella Sezione 2.1.1 rientra proprio nel periodo della crisi finanziaria globale del 2007-2009. In questo periodo, la fiducia nelle banche e nei sistemi centralizzati era ovviamente molto bassa, e il Bitcoin si propone come un'alternativa completamente decentralizzata e che veramente può eliminare fonti di controllo e centralizzazione. Il Bitcoin è stata addirittura definita "una moneta anticapitalista": oltre alla questione della perdita di potere per le banche, il sistema di Bitcoin è stato concepito per non generare più di 21 milioni di monete. Questo lo rende un sistema deflattivo piuttosto che inflattivo come le tradizionali valute fiat².

2.2.2 Funzionamento della rete

La rete di Bitcoin ha un approccio peer-to-peer (P2P) decentralizzato in cui la correttezza delle transazioni può essere rapidamente e immediatamente confermata da qualsiasi partecipante alla rete, in qualsiasi momento. Si basa sulla blockchain come registro pubblico e condiviso; dal momento che tutte le transazioni confermate sono incluse nella blockchain, i vari portafogli Bitcoin della rete possono calcolare il loro bilancio disponibile e nuove transazioni possono essere verificate, controllando che chi spende abbia sufficiente

²Valuta fiat: valuta nazionale non ancorata al prezzo di una materia prima, come oro o argento.

disponibilità attraverso la crittografia. Per partecipare attivamente alla rete è necessario avere una coppia di chiavi, una pubblica e una privata, fondamentali per l'accesso alla moneta (non c'è infatti possibilità di recupero di Bitcoin in caso di perdita della chiave privata). Se da una parte la chiave privata è nota solo all'utente di riferimento ed è usata per firmare le sue transazioni, la chiave pubblica è invece nota a chiunque, identifica pubblicamente e univocamente l'utente e può essere utilizzata anche per verificare la firma. Ogni transazione Bitcoin è un trasferimento di criptovaluta da chiave pubblica a chiave pubblica, ma la criptovaluta ricevuta può essere spesa solamente tramite l'utilizzo della chiave privata. C'è da notare che, nonostante la chiave pubblica identifichi univocamente un utente, si può comunque dire che il sistema di Bitcoin garantisce uno pseudo-anonimato, dal momento che questo identificatore non ha nulla a che vedere con la vera identità del proprietario.

Di seguito un breve riassunto delle fasi per il mining di un blocco:

1. *Transazione*: Il processo inizia quando un utente vuole inviare delle criptovalute a un'altra persona. Egli manda una transazione con i dati delle chiavi, in attesa che la rete esegua e confermi.
2. *Compilazione*: La transazione in sospeso, insieme alle altre sulla rete, viene raccolta e raggruppate in un blocco di transazione dai nodi.
3. *Formazione*: Ogni minatore seleziona le transazioni che desidera includere e crea il proprio blocco, il quale andrà tra i candidati (e non tra i blocchi validi, non avendo ancora una Proof of Work valida)

4. *Proof of Work*: Ogni minatore cerca di eseguire la Proof of Work, cioè di trovare una firma (hash) valida, per il proprio blocco.
5. *Trasmissione*: Quando un nodo riesce a trovare un hash di output valido per un blocco, trasmette tale blocco insieme alla firma agli altri nodi della rete, in modo che essi possano convalidarlo. In questo momento il miner riceve la ricompensa stabilita per il mining (se non sono stati emessi 21 milioni di Bitcoin) e le commissioni di mining che gli utenti hanno messo nelle transazioni che compongono il blocco (indipendentemente dal fatto che siano stati emessi 21 milioni di Bitcoin).
6. *Verifica*: Gli altri nodi della rete validano e verificano che il blocco e l'hash rispettino le condizioni di sistema e che siano legittimi.
7. *Conferma*: Una volta che il blocco viene aggiunto alla blockchain, tutti gli altri che vengono aggiunti su di esso conterranno come conferma. Ora che il blocco è stato generato, tutti i nodi di mining devono avviare il processo formando un nuovo blocco di transazioni.

2.2.3 Le transazioni

Una transazione è un trasferimento di valore tra due parti. Nella rete Bitcoin, queste sono fondamentalmente l'invio di criptovaluta tra le varie persone che usano la rete. In realtà, però, le transazioni non sono altro che messaggi programmati, firmati digitalmente con la crittografia e inviati alla rete per la convalida. Tutte le transazioni sono pubbliche e possono essere ritrovate

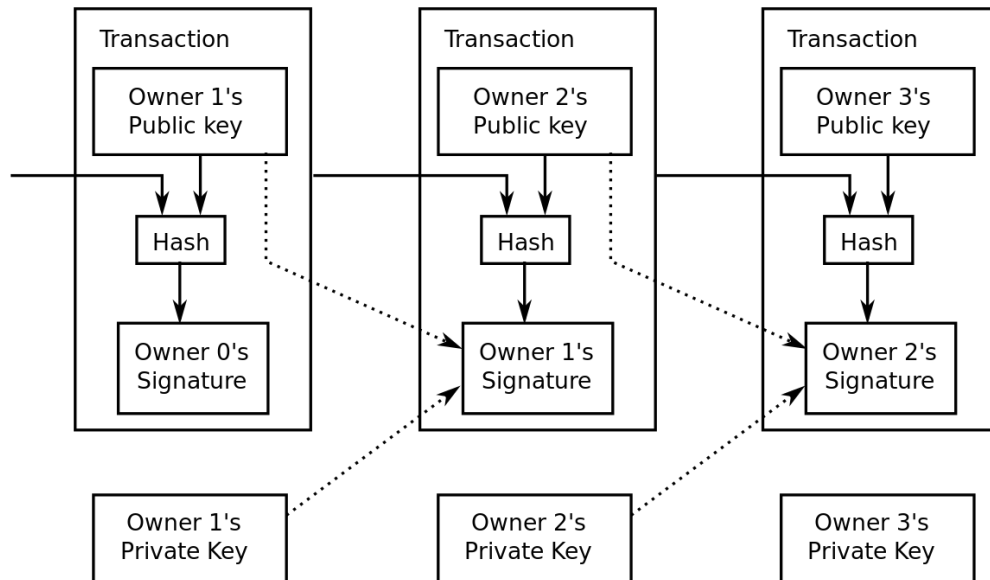


Figura 2.2: Rappresentazione schematica di una transazione Bitcoin.

e verificate sulla blockchain. Gli elementi che compongono una transazione Bitcoin sono i seguenti:

- *Entrate (input)*: Queste si riferiscono all'uscita di una transazione passata ancora non utilizzata in nessun'altra transazione. Contengono l'indirizzo dove sono stati originariamente ricevuti i Bitcoin.
- *Uscite (output)*: Queste contengono gli indirizzi beneficiari del trasferimento e l'importo inviato a ciascuno.
- *Identificatore*: Anche detto *TXid*, è un valore unico che permette di identificare univocamente una transazione all'interno della blockchain.

- *Tasso di commissione*: Anche detto *fee*, è un piccolo pagamento che i miners ricevono per l'elaborazione di una transazione.

Generalmente, prima di considerare davvero una transazione come verificata, si aspetta il mining di altri 6 blocchi, ma questo numero è arbitrario ed è più che altro una convenzione per assicurare il buon funzionamento del sistema.

2.2.4 La validazione

La validazione in Bitcoin funziona attraverso il meccanismo di consenso della Proof of Work (Capitolo 2.1.3); in particolare, i vari hardware della rete ricercano un *nonce*³, il quale, se concatenato alla sequenza di transazioni del blocco e all'hash del blocco precedente, genera una nuova immagine hash che inizia con almeno t zeri, dove t è un valore stabilito in modo decentralizzato. Il procedimento che genera t è il seguente: un algoritmo verifica quanti computer stanno provando a risolvere il puzzle crittografico in un dato momento, verificando il tempo impiegato per risolvere i puzzle precedenti, e alza o abbassa il numero di zeri richiesti (anche detto *difficulty*) di conseguenza, per mantenere il tempo tra due blocchi consecutivi il più costante possibile. L'unico modo possibile di trovare il *nonce* è per tentativi e, in media, ciò richiede l'applicazione della funzione hash 2^t volte. Nel caso in cui due nodi convalidino contemporaneamente lo stesso blocco, si crea ciò che è definito come un *fork* della blockchain, ovvero una diramazione della stessa in due rami differenti: viene poi mantenuto solo il ramo più lungo, per fare in modo che tutti i nodi concordino sulla stessa sequenza.

³*Nonce*: in crittografia, è un numero generalmente casuale (o pseudo-casuale) destinato all'autenticazione del trasferimento dati tra due o più parti.

2.3 Ethereum

Ethereum è una piattaforma software, descritta in un whitepaper[6] nel 2014 e lanciata nel 2015 da Vitalik Buterin, basata sulla tecnologia blockchain che si appoggia a svariati computer nel mondo intero. La peculiarità di questa piattaforma è che la sua rete permette di costruire e distribuire delle applicazioni decentralizzate, dette *Dapps* (*decentralised applications*); queste sono open source e decentralizzate per natura, e per la loro creazione vengono utilizzati gruppi di smart contracts. Ethereum è sempre stato basato sul protocollo Proof of Work, ma con il *Bellatrix update*, iniziato il 6 Settembre 2022, è nato Ethereum 2.0, basato invece sul protocollo Proof of Stake. In questo capitolo andremo a parlare della valuta nativa di Ethereum, l'Ether, ma anche delle novità che Ethereum ha portato in ambito criptovalute, come i portafogli, le Dapps e gli Smart Contracts. Andremo poi a soffermarci anche sui concetti di token ed exchange.

2.3.1 Ether

Ether (ETH) è la valuta nativa di Ethereum. Esso è stato progettato per essere una forma di pagamento per i partecipanti alla rete che eseguono gli smart contracts (Capitolo 2.3.4). Oltre a essere una normale criptomoneta utilizzabile come strumento di trading, dunque, è possibile utilizzare Ether per eseguire Dapps (Capitolo 2.3.3), in quanto le risorse computazionali necessarie per eseguire le operazioni di tali applicazioni sono tracciate e pagate con token Ether. Quando un utente invia Ether o utilizza un'applicazione Ethereum, egli paga comunque una commissione in Ether (detta *Gas*) sia per

utilizzare la rete, sia per ricompensare il miner che con il suo nodo elabora e verifica ciò che si sta facendo.

2.3.2 Portafogli e conti

I portafogli di Ethereum (*Ethereum wallets*) sono applicazioni che consentono di interagire con il proprio conto di Ethereum, inviando ad esempio transazioni, leggendo il proprio saldo, connettendosi alle applicazioni, e altro. I portafogli sono accessibili tramite le chiavi private, dunque perdere la propria chiave privata vuol dire perdere definitivamente l'accesso ai propri fondi. C'è da sottolineare che questi rimangono solo degli strumenti per gestire il conto, e non hanno custodia dei fondi stessi. Non sono infatti da confondersi con i conti di Ethereum (*Ethereum accounts*), i quali sono invece entità collegate a un indirizzo Ethereum che hanno un saldo e possono inviare transazioni: i portafogli non sono altro che prodotti per gestire questi conti.

2.3.3 Dapps

Le Dapps (*Decentralized applications*) sono una tipologia di applicazioni il cui funzionamento non dipende da punti di controllo centrali, ma funziona sulla base di una rete decentralizzata. La popolarità delle Dapps è aumentata molto con la presentazione di Ethereum e il suo linguaggio Solidity. La differenza principale tra un'applicazione classica e una Dapp sta nel backend: la logica decentralizzata delle Dapps vuole che il backend sia correlato a uno smart contract eseguito su una blockchain (Ethereum ad esempio). Dal momento che gli smart contracts sono pubblici e consultabili, viene garantito un livello elevato di trasparenza e sicurezza, poiché è impossibile che una

Dapp faccia qualcosa non specificata nello smart contract. Le Dapps sono applicazioni Turing complete, dunque sono in grado di eseguire qualsiasi azione una volta fornite le risorse necessarie, e godono di una resistenza alla censura totale: nessuna entità sulla rete può infatti impedire agli utenti di distribuire e interagire con le Dapps.

2.3.4 Smart contracts

Gli smart contracts sono dei software basati sulla blockchain, costituiti da un codice crittografico e usati per automatizzare l'esecuzione di un accordo in modo che tutti i partecipanti possano essere immediatamente e direttamente certi dell'esito; lo smart contract offre garanzie solide per via della sua essenza, trasparente e pubblica, sempre consultabile. Gli utenti possono essere certi che uno smart contract farà sempre solo e soltanto quello che dice di fare. Nel caso di Ethereum, uno smart contract ha un indirizzo e può inviare e ricevere ether, oltre a eventuali parametri. Gli account utente possono dunque facilmente interagire con lo smart contract, ad esempio inviando transazioni che svolgono le funzioni definite sullo smart contract stesse. Caratteristica molto importante degli smart contract è che, una volta registrati sulla blockchain, non possono essere modificati né cancellati, dunque eventuali errori nel processo di creazione possono portare a uno spreco di ether usato come gas nel migliore dei casi, oppure a risultati indesiderati nel peggiore.

Una delle applicazioni più importanti degli smart contracts è quella delle Decentralized Autonomous Organizations, anche dette DAO.

DAO

Una DAO[7] (Decentralized Autonomous Organization) è un'organizzazione autonoma decentralizzata, basata sulle possibilità di regolare le transazioni tramite smart contracts e di associare particolari diritti a specifici token appositamente emessi. I diritti attribuiti a questi token permettono di legittimare l'invio di transazioni verso uno smart contract. Essendo quindi una DAO un ente gestito da una rete di pari, i detentori di token prendono decisioni sulla gestione in maniera autonoma e senza che nessuno di essi abbia qualifiche superiori agli altri. La struttura stessa della DAO è fatta in modo da autogovernarsi, rendendo totalmente inutile e superflua l'esistenza di un organo amministrativo.

2.3.5 Token

Con il termine token ci si riferisce solitamente a un criptoasset gestito tramite la blockchain di un'altra criptovaluta, di cui sfrutta uno smart contract (vedi Capitolo 2.3.4) per essere distribuito. Questi sono saliti alla ribalta grazie a Ethereum e al suo ecosistema. Proprio Ethereum, inoltre, ha introdotto degli standard per questi token, in modo da fissare delle regole di riferimento che facilitino la cooperazione, ma anche la compatibilità tra le varie soluzioni su blockchain: gli standard più importanti sono Fungible Token ERC-20 e Non-Fungible Token ERC-721.

Fungible Token ERC-20

L'ERC-20 è il token standard più diffuso. Esso è detto fungibile perché sempre frazionabile e tutti i token hanno lo stesso valore, dunque può essere sempre scambiato facilmente. Questo standard è solitamente utilizzato per criptovalute di progetti che non hanno una propria blockchain e per gli utility token, ovvero token che sono al servizio di uno scopo specifico. Sono spesso utilizzati ad esempio dalle aziende per suscitare interesse nei loro prodotti.

Non-Fungible Token ERC-721

L'ERC-721 è un token standard creato per i token che non devono essere fungibili né divisibili, ma unici. L'idea di base è che possedere un bene digitale basato su questo standard vuol dire averne la proprietà registrata su blockchain. I token generati con questo standard sono anche detti NFT (Non-Fungible Token). Non esistendo token ERC-721 equivalenti, ogni NFT ha quindi un valore e una storia diversi dagli altri.

Parametri	Token Fungibili	Token Non Fungibili
Scambiabilità	Possono essere scambiati con token dello stesso tipo	Non possono essere scambiati con token dello stesso tipo
Uniformità	Sono identici	Sono unici
Frazionabilità	Possono essere suddivisi in unità più piccole	Non possono essere suddivisi

Figura 2.3: Riassunto schematico delle differenze tra token fungibili e non fungibili.

Altri tipi di token

Oltre ai due principali, esistono altri standard di token: tra questi:

- ERC-1155, uno standard a metà tra fungibile e non fungibile, ovvero un token che può essere a scelta fungibile, non fungibile o semi-fungibile.
- ERC-1400 ed ERC-3643, due security token, ovvero rappresentanti la proprietà, totale o frazionaria, sugli interessi generati da un asset o da un'entità, differenziati tra di loro per il modo in cui i proprietari eseguono le transazioni: il primo richiede l'uso di chiavi crittografiche generate fuori dalla blockchain, il secondo invece si basa su un sistema automatizzato, su blockchain.

2.3.6 Exchange

La diffusione di Ethereum e tutti i progetti che a esso sono collegati o si appoggiano ha portato a una grande crescita in quantità del mondo criptomonetario: si è fatta presto sentire quindi la necessità di poter gestire i propri fondi in maniera pratica e sicura. Al giorno d'oggi, Criptovalute e token possono essere quotati in quelli che si dicono exchange, ovvero piattaforme di scambio in cui è possibile comprare e vendere criptomonete di qualsiasi tipo.

Gli exchange possono essere utilizzati per acquistare criptomonete attraverso denaro reale, per vendere criptomonete e per scambiare una tipologia di criptomoneta con un'altra. Queste attività non sono esenti da tasse, difatti l'exchange si riserva il diritto di porre un costo in commissioni per ogni tipologia di transazione effettuata su di esso. In genere, questa tassa serve per incentivare altri utenti a fornire liquidità al servizio.

2.4 Blockchain Online Social Media

I Blockchain Online Social Media (BOSMs) [8] sono per il Web3⁴ ciò che i Social Media sono per l'attuale Web2: questo a motivo dei vari vantaggi che i Social attuali non riescono a offrire in alcun modo, e probabilmente non potranno mai fare proprio per via della loro natura. Le piattaforme di social media centralizzate sono spesso al centro dell'attenzione per via della loro gestione di vari argomenti molto accesi, che possono risultare in questioni di censura, di politica, di privacy, e molte altre; questioni che sono sempre molto alimentate, e non sempre una piattaforma centralizzata è in grado di gestirle al meglio e senza davvero nuocere a nessuno. I BOSM, per via della loro essenza decentralizzata, non potranno mai essere additati come la causa di questioni simili, in quanto essi offrono ai loro utenti trasparenza e sicurezza massime, oltre a un maggiore controllo della piattaforma stessa. Si può dire che in un certo senso la piattaforma venga costruita dai suoi utenti. L'idea di base dietro ai BOSM è l'utilizzo della tecnologia blockchain non solo per dare garanzie agli utenti, ma anche per incentivarli a produrre contenuti di valore. Essendo infatti l'ecosistema del Social in mano agli utenti, idealmente i migliori contenuti verranno premiati in criptovaluta, poiché è nell'interesse di tutti gli utenti trovare dei contenuti di valore, o comunque interessanti. Ma i vantaggi dei BOSM non finiscono qui: un'altra piaga che sta spopolando sempre di più sui Social, ma anche in generale sul Web, è quella delle fake news. Con un meccanismo funzionante di rewarding e

⁴Web3: il presunto prossimo grande passo che farà il mondo del Web, funzionando grazie alle blockchain e togliendo così alle grandi compagnie la possibilità di controllare il Web stesso, massimizzando di conseguenza la libertà d'uso degli utenti.

un'utenza attiva, queste non riusciranno mai ad avere scalpore, e inoltre non ci potrà essere qualche individuo o entità che manipola o lascia trasparire ciò che vuole, magari per fini maliziosi.

Negli ultimi anni, di pari passo con l'evoluzione nell'ambito della blockchain, anche il numero di BOSM è cresciuto notevolmente, nonostante a livello di numeri i Social Media tradizionali rimangano insuperabili. Citiamo Steemit[12][13] (basata sulla blockchain di Steem, specificamente progettata per lo sviluppo e la gestione dei social media e molto efficiente nell'ambito) come il più famoso, ma anche Minds, BitClout, Hive.blog e PeakD.

Capitolo 3

Reddit e i Community Points

Introduciamo in questo capitolo la piattaforma Reddit e in particolare i Reddit Community Points, una recente introduzione di Reddit che mira a evolvere il concetto di community dando il potere agli utenti stessi e abbracciando una visione decentralizzata dell'Internet. Analizzeremo, inoltre, il subreddit ethTrader e i Donuts, ovvero i suoi Community Points, spiegando anche perché questi sono stati per noi importanti e cosa abbiamo pensato di trarne.

3.1 Reddit

Reddit è una piattaforma Internet generalmente considerata come un Social Media, ma che si trova in realtà a metà tra un Social Media e un forum. Reddit è infatti diviso in *subreddit*, i quali sono fondamentalmente dei forum divisi per argomento e in cui gli utenti (i *redditors*) sono allo stesso tempo creatori di contenuti, consumatori e curatori. Seppure ogni subreddit disponga di uno o più moderatori, la piattaforma può considerarsi come autogestita;

ad esempio, è la community stessa a decidere cosa è importante e cosa no, attraverso un sistema di upvotes/downvotes che poi si riflette sulla visibilità dei post stessi. Con il tempo, Reddit si è evoluto sempre di più, e la direzione scelta è sempre stata verso una piattaforma *community-driven*¹, e a riprova di ciò possiamo notare le ultime introduzioni, prima tra tutte quella dei Reddit Community Points (approfonditi nella sezione successiva, 3.2.1), il cui scopo è proprio quello di consentire agli utenti di esprimere le proprie opinioni e di avere un controllo, in un certo senso, decentralizzato.

Al 2022, Reddit conta più di 100.000 subreddit attivi, con un incremento su base annua del 14% per quanto riguarda i post e del 7% per quanto riguarda i commenti[3]. La sua diffusione in Italia non è ancora paragonabile a quella di altre piattaforme simili, come Facebook o Instagram, ma l'accrescimento delle community anche italiane non esclude che in un futuro Reddit possa diventare una piattaforma al pari di quelle citate per diffusione.

3.1.1 Karma

Il karma è un punteggio utilizzato su Reddit il cui scopo è quello di riflettere quanto il contributo di un singolo utente è stato apprezzato dalla community.

È possibile guadagnare karma in più modi su Reddit: un utente guadagna del karma quando un suo post o un suo commento di viene votato positivamente (*upvoted*), ma anche quando egli dona o riceve un award, ovvero dei "premi" che gli utenti di Reddit possono offrire ad altri per reagire alle loro pubblicazioni. Reddit stesso non vuole che gli utenti pubblicino contenuti

¹Community-driven: diretta dalla comunità.

interessanti solo per aumentare il proprio karma, ma piuttosto che il karma rispecchi l'impegno e la dedizione messi a disposizione della community.

3.2 Reddit Community Points

I Reddit Community Points sono fondamentalmente una misura della reputazione del singolo utente all'interno di una community e, in quanto tali, quelli guadagnati sono visualizzati accanto al nome utente in ogni suo contenuto pubblicato nel subreddit, e valgono solo per quel singolo subreddit. La cosa che li contraddistingue però dai classici punti reputazione che si trovano solitamente sui social è che i Community Points esistono su una blockchain come token (vedi Capitolo 2.3.5), e non sono quindi vincolati direttamente a Reddit; si può dire che appartengano direttamente all'utente. Ovviamente, l'utilizzo maggiore dei Community Points si può fare sul subreddit in cui si sono guadagnati, ad esempio comprando speciali badges, emojis o addirittura la possibilità di modificare la grafica del subreddit stesso, ma come token sulla blockchain possono anche essere convertiti in altre criptovalute e usate quindi per scopi speculativi. Una precedente ricerca[10] ci mostra che i Donut token (ovvero i Reddit Community Points del subreddit ethTrader) sono utilizzati dagli utenti più per azioni speculative che nell'ecosistema del subreddit.

3.2.1 Distribuzione dei Reddit Community Points

I Community Points sono distribuiti secondo il seguente schema:

- Contributori: 50%.

- Moderatori: 10%.
- Community Tank²: 40%.

Il modo in cui i Contributori vengono classificati per la distribuzione dei Community Points è attraverso il karma (vedi Capitolo 3.1.1) ottenuto nel singolo subreddit, che starà a indicare quindi quanto l'utente ha contribuito al subreddit stesso: questo è detto *Contribution Score*. I subreddit possono, a loro discrezione, applicare regole aggiuntive oltre a queste basiche (ad esempio, raddoppiando il karma ottenuto dai commenti oppure rendendo nullo il karma ottenuto da post con una specifica etichetta).

La distribuzione iniziale di Community Points per le community stabili e attive da molto tempo è di 50 milioni di punti; dopo questa, altri punti sono distribuiti ogni 4 settimane, con 5 milioni di punti distribuiti la prima volta, ridotti del 2.5% per le distribuzioni successive. Nel momento della distribuzione mensile, Reddit pubblica un post nel subreddit con la lista del *Contribution Score* di ogni utente, come file .csv³; fatto ciò, la community ha 1 settimana di tempo per discutere della lista e proporre eventuali cambiamenti, i quali verranno poi eventualmente discussi e valutati attraverso dei sondaggi, in cui i membri con più Community Points avranno più importanza. Passata la settimana, Reddit pubblica una seconda lista con i dati finali, e i Community Points sono distribuiti in base a questa⁴. Da notare che il numero di Community Points citato è quello base; ogni subreddit po-

²Community Tank: un fondo che supporta il progetto.

³Formato CSV: formato utilizzato per rappresentare dati in forma tabellare all'interno di un file di testo.

⁴<https://www.reddit.com/community-points/documentation/distribution-process>

trà poi apportare modifiche alla distribuzione dei propri Community Points attraverso le DAO (vedi Capitolo 2.3.4).

3.3 ethTrader e Donuts

In questo Capitolo andiamo a vedere cosa è ethTrader e in particolare cosa sono e come funzionano i Donuts, ovvero i Reddit Community Points del subreddit ethTrader.

3.3.1 ethTrader

ethTrader è un subreddit dedicato a Ethereum (vedi Capitolo 2.3), in cui gli utenti discutono liberamente dell’ecosistema Ethereum, con discussioni sul mercato, news sulla tecnologia, memes e quant’altro. Un utente moderatore in particolare, *EthTraderCommunity*, mantiene il subreddit informato con le news più recenti a proposito del mondo di Ethereum, e aiuta gli utenti a farsi sentire nella community grazie al post ricorrente *Daily Discussion*, in cui gli utenti possono utilizzare la sezione commenti per discutere di come la community possa venire incontro agli utenti stessi, suggerendo cambiamenti o altro. Anche i Donuts, essendo comunque un meccanismo strettamente legato all’attività degli utenti nonché parte integrante dell’ecosistema del subreddit, sono solitamente discussi in questo post.

3.3.2 Donuts

I Donuts sono i Reddit Community Points del subreddit ethTrader, e sono stati i primi Reddit Community Points a essere introdotti in un subreddit.

I Donuts sono dei token *ERC-20* (vedi Capitolo 2.3.5) gestiti da una *DAO* (vedi Capitolo 2.3.4), controllata interamente da smart contracts: questi e solo questi determinano le regole di come le parti coinvolte nella *DAO* devono comportarsi e cooperare.

I Donuts esistono sia sulla blockchain principale di Ethereum che su quella di Gnosis, inoltre è possibile scambiarle sugli Exchange (vedi Capitolo 2.3.6) Uniswap⁵ e Honeyswap⁶.

La distribuzione mensile dei Donut avviene secondo il seguente schema, diverso da quello base per i Reddit Community Points (che si trova nel Capitolo 3.2.1) perché modificato dalla DAO del subreddit:

- *Submissions*: 510K Donut per distribuzione.
- *Comments*: 340K Donut per distribuzione.
- *Bonus ricevuti per le Submissions attraverso tips*: 340K Donut per distribuzione.
- *Fondi per la community*: 255K Donut per distribuzione.
- *Tips in Donut & cura dei contenuti*: 170K Donut per distribuzione.
- *Incentivi Liquidity pool Uniswap*: 400K Donut per distribuzione.
- *Incentivi Liquidity pool Honeyswap*: 200K Donut per distribuzione.
- *Moderazione*: 85K Donut per distribuzione.

⁵<https://app.uniswap.org/tokens/ethereum/0xc0f9bd5fa5698b6505f643900ffa515ea5df54a9>

⁶<https://info.honeyswap.org/token/0x524b969793a64a602342d89bc2789d43a016b13a>

Esiste anche un meccanismo di *burning*⁷: sono bruciati e quindi rimossi per sempre dalla circolazione i Donut spesi per le membership premium del subreddit e i Donut spesi per l'ownership del banner del subreddit (il banner è gestito attraverso un sistema di tassa Harberger⁸).

3.4 Scopo della tesi

Ricerche precedenti su Social Media completamente decentralizzati[9] ci hanno mostrato una tendenza negli utenti a citare più spesso gli argomenti blockchain, criptovalute e la piattaforma su cui si sta pubblicando il contenuto (Steemit⁹ in caso della ricerca citata) in ambienti sottoposti a un meccanismo di rewarding, oltre ai classici contenuti come cibo, viaggi e sport. A partire da questo, la nostra ricerca ha voluto ottenere considerazioni simili sul piccolo ecosistema di ethTrader, una community interessante in quanto prima in assoluto ad aver introdotto il sistema di Reddit Community Points ed è comunque basata su un argomento molto vicino a quello del rewarding e dei Social Media decentralizzati: blockchain, Ethereum e trading sono i tre argomenti principali di discussione. Ci siamo interessati, ad esempio, alle

⁷*Token burning*: processo di rimozione di una porzione di un certo numero di monete dalla circolazione per controllarne il prezzo e mantenere bassa l'inflazione.

⁸Tassa Harberger: in economia, uno strumento di misurazione dell'eccesso di pressione o cuneo fiscale generato dall'introduzione di un'imposta, causando una possibile perdita di benessere sociale che non viene trasferito ad altri, ma perso del tutto. È un meccanismo sempre più utilizzato nell'ambito della blockchain in quanto aiuta la decentralizzazione.

⁹Steemit: sito di blog e social media basato su blockchain, in cui gli utenti ottengono una criptovaluta, STEEM, per pubblicare e curare i contenuti.

parole più diffuse all'interno del subreddit, ma anche al livello di positività e di soggettività delle pubblicazioni e dei commenti.

Per fare ciò, abbiamo seguito 3 passi principali:

- Abbiamo raccolto i dati sulle pubblicazioni e sui commenti del subreddit ethTrader a partire dal Gennaio del 2018 fino al Settembre del 2022.
- Abbiamo riorganizzato e analizzato i dati.
- Abbiamo generato e discusso dei grafici a partire dati analizzati in precedenza.

Seguito questo processo, con tutto il materiale alla mano, abbiamo potuto trarre le nostre conclusioni, talvolta anche appoggiandoci a materiale esterno che ha funzionato da base per la nostra ricerca.

Capitolo 4

Implementazione

Spieghiamo in questo Capitolo come è stata attuata la ricerca sugli oggetti presentati nel Capitolo 3. Ci soffermeremo, in particolare, sugli strumenti utilizzati, sulle fasi dell'implementazione e sul codice prodotto ai fini della ricerca.

4.1 Strumenti

4.1.1 Pushshift

Pushshift è un motore di ricerca di contenuti pubblicati su Reddit, le quali API completamente open-source rendono possibile esportazione, tramite codice, di tantissimi tipi di dati, inclusi *Comments* e *Submissions*. Le API sono disponibili direttamente dal sito¹.

¹<https://api.pushshift.io/>

L'utilizzo di Pushshift avviene tramite una richiesta HTTP, generata a partire dagli endpoint `/reddit/search/comment` (per i *Comments*) e `/reddit/search/submission` (per le *Submissions*): queste chiamate generali possono essere poi specializzate aggiungendo parametri specifici. Il risultato della chiamata sono dei dati in formato JSON, con i risultati raccolti nella chiave "data". Tra i parametri:

- *q*: cerca contenuto con parole specifiche.
- *ids*: cerca contenuto con uno specifico id.
- *author*: cerca contenuti pubblicato da uno autore specificato.
- *subreddit*: cerca contenuti in uno specifico subreddit.
- *before* e *after*: cercano contenuto prima o dopo una data specificata.
- *size*: il numero di risultati ritornati. Può arrivare al massimo a 500 e se non specificato è 25.

Alternative a Pushshift possono essere le API ufficiali di Reddit, le quali però limitano le possibili richieste giornaliere, oppure *PRAW* (*Python Reddit API Wrapper*), uno strumento ideato per facilitare l'accesso alle API di Reddit. Quest'ultimo richiede un'autenticazione per il suo utilizzo, a differenza di Pushshift, ed è quindi indicato per ricerche più specifiche.

4.1.2 Python

Python è un linguaggio di programmazione ad alto livello orientato a oggetti. Il suo miglior vanto è possedere una enorme quantità di librerie, che lo

rendono utilizzabile in moltissimi ambiti e per moltissimi scopi. Di seguito, dunque, le librerie utilizzate:

- *requests*, che permette di effettuare richieste http e https. Grazie a questa abbiamo potuto ottenere i dati necessari, i quali sono poi stati salvati e riutilizzati sotto forma di file JSON.
- *json*, pacchetto già incluso in Python e che permette di lavorare su dati in formato JSON (approfondito nel Capitolo 4.1.3).
- *matplotlib*, fondamentale per la generazione di visualizzazioni grafiche. Grazie a questa abbiamo potuto generare i vari grafici presentati nel Capitolo 5.
- *wordcloud*, per la generazione dei Wordcloud, ovvero "nuvole di parole": si tratta di immagini composte da un insieme di parole, la cui grandezza varia a seconda della loro frequenza. Questo strumento è utilizzato soprattutto per permettere di visualizzare in maniera efficace e immediata i concetti di maggiore importanza.
- *collections*, in particolare la sua classe *Counter*, una sottoclasse di *Dictionary*, utilizzata per contare elementi.
- *spacy*, libreria volta al *NLP* (*Natural Language Processing*).

4.1.3 JSON

JSON (*JavaScript Object Notation*) è un formato di serializzazione di dati, molto semplice e di immediata interpretazione, sia per l'umano che per la macchina.

JSON è basato su due strutture, le quali riprendono convenzioni tipiche dei linguaggi di programmazione, nonostante JSON sia completamente indipendente da essi:

- Un insieme di coppie nome/valore.
- Un elenco ordinato di valori.

Un file JSON inizia sempre con una coppia di parentesi graffe racchiudenti il corpo dell'intera struttura. I nomi JSON sono disposti sulla sinistra, e hanno la struttura di una stringa qualsiasi tra virgolette. In un oggetto JSON, ogni nome deve essere unico. I valori JSON sono invece disposti sulla destra, e possono essere uno dei seguenti 6 tipi di dato: *string*, *number*, *object*, *array*, *boolean*, *null* (o *empty*). Questi valori possono essere ricavati richiamando il rispettivo nome, in una relazione come quella chiave/valore dei linguaggi di programmazione. Le strutture sopra citate vanno poi ad assumere le seguente forme:

- *Oggetto*: una serie non ordinata di nomi/valori. Inizia con "}" (parentesi graffa aperta) e finisce con "}" (parentesi graffa chiusa). Ogni nome è seguito da ":" (due punti) e le coppie nome/valore sono separate da "," (virgola).
- *Array*: una raccolta ordinata di valori. Inizia con "[" (parentesi quadra aperta) e finisce con "]" (parentesi quadra chiusa). I valori sono separati da "," (virgola). È possibile fare riferimento all'i-esimo valore di un array specificando la posizione di quel valore.

- *Valore*: una stringa tra virgolette, un numero, un valore booleano (*true* o *false*), un valore *null*, un oggetto o un array. Queste strutture possono essere annidate.
- *Stringa*: una raccolta di zero o più caratteri Unicode², tra virgolette. Un singolo carattere è rappresentato come una stringa di caratteri di lunghezza uno.
- *Numero*: un valore numerico.
- *Whitespace*: un carattere di spaziatura, che può essere inserito in mezzo a qualsiasi coppia di token. Descrive quasi totalmente il linguaggio.

4.1.4 Spacy

Spacy è una libreria open source rilasciata sotto licenza MIT per l'elaborazione del linguaggio naturale (*NLP*). Scritta in Python e Cython, il suo principale utilizzo si trova nel campo delle applicazioni software, in quanto aiuta a costruire applicazioni che possano processare una grande quantità di testo. A oggi, Spacy supporta analisi basate sull'apprendimento profondo³, consentendo di impiegare modelli statistici addestrati utilizzando librerie per l'apprendimento automatico come TensorFlow e Scikit-learn. Tra le funzioni che Spacy possiede ci sono *named entity recognition*, *Part-of-speech tag-*

²Unicode: standard di codifica dei caratteri che assegna un numero univoco a ogni carattere usato per la scrittura di testi, in maniera indipendente dalla lingua, dalla piattaforma informatica e dal programma utilizzato.

³Apprendimento profondo (o *deep learning*): un insieme di tecniche basate su reti neurali artificiali organizzate in diversi strati, dove ogni strato calcola i valori per quello successivo ai fini di avere un'elaborazione completa dell'informazione.

ging, *dependency parsing*, segmentazione del testo, classificazione del testo, lemmatizzazione, analisi morfologica e altre, il tutto grazie a un sistema di addestramento e una tokenizzazione efficienti alla base.

A differenza di NLTK (*Natural Language Toolkit*), uno strumento molto simile a Spacy, quest'ultimo ha un approccio orientato agli oggetti: quando analizziamo del testo, Spacy ritorna un oggetto documento le cui parole e frasi sono oggetti anch'esse. NLTK, al contrario, ha un approccio orientato alle stringhe: esso riceve delle stringhe in input e ritorna stringhe o liste di stringhe come output. NLTK è stato creato con l'idea di aiutare a creare funzioni complesse o modelli, offrendo accesso a vari algoritmi e permettendo all'utente di personalizzare il suo approccio; Spacy, al contrario, offre un approccio poco personalizzabile ma efficiente e accurato.

Modelli

Spacy supporta tokenizzazione e addestramento per più di 70 lingue. Per quanto riguarda la lingua inglese, sono disponibili 3 modelli principali:

- *en_core_web_sm*: modello piccolo, il più rapido ma il meno accurato.
- *en_core_web_md*: modello medio, una via di mezzo, bilanciato a metà tra immediatezza e accuratezza.
- *en_core_web_lg*: modello grande, il più lento nel caricamento ma il più completo.

Le differenze tra questi modelli sono comunque poco rilevanti a livello di prestazioni⁴. I due modelli che per lo più abbiamo utilizzato sono stati quello

⁴<https://spacy.io/models/en>

piccolo, nei casi in cui era presente una grande quantità di dati, dal momento della poca differenza prestazionale tra i modelli, e quello grande, per i wordcloud, in quanto abbiamo immaginato che in un ambiente come quello di ethTrader i termini più utilizzati avessero potuto anche essere poco comuni nella lingua di tutti i giorni, dunque non volevamo rischiare di perdere qualche dato importante.

4.2 Fasi dell'implementazione

Descriviamo qui le fasi dell'implementazione, che si è articolata in 3 fasi principali: la raccolta iniziale dei dati, la successiva organizzazione degli stessi, e infine la generazione dei grafici a partire da tali dati.

4.2.1 Raccolta dei dati

Per prima cosa, abbiamo raccolto i dati necessari. Per farlo abbiamo utilizzato le API di Pushshift (spiegate nel dettaglio al Capitolo 4.1.1). In particolare, le richieste che abbiamo effettuato sono state le seguenti:

- `https://api.pushshift.io/reddit/search/submission/?subreddit=ethtrader&size=250&after=1514764801&before=str(before_time)`
- `https://api.pushshift.io/reddit/search/comment/?subreddit=ethtrader&size=250&after=1514764801&before=str(before_time)`

Un primo ostacolo davanti al quale ci siamo trovati è stata una limitazione delle API, le quali non permettevano di ricavare più di 250 *Comments/Submissions* alla volta. Per aggirare questa inconvenienza, abbiamo in-

nanzitutto fatto una prima richiesta con *before_time* uguale a 1664294400 (ovvero il timestamp epoch del 27 Settembre 2022 alle ore 16:00), la quale ci ha dato le ultime 250 submissions a partire da quella data. Abbiamo poi letto le informazioni di tale richiesta, e ne abbiamo costruito una seconda sostituendo il parametro *before_time* iniziale con il timestamp epoch dell'ultima *Submission* o *Comment* presente nella richiesta. Abbiamo poi impostato un procedimento analogo per il raccoglimento di tutti i dati necessari, fino a che non viene raggiunto il timestamp epoch 1514764801, corrispondente all'1 Gennaio del 2018. In questo modo, siamo riusciti ad avere tutti i *Comments* e tutte le *Submissions* del subreddit ethTrader dal 2018 fino al Settembre del 2022.

4.2.2 Organizzazione dei dati

A partire dal dataset originale che comprende *Comments* e *Submissions* (descritto in dettaglio al Capitolo 5.1), abbiamo scritto dei programmi per produrre altri output, in modo che questi fossero facilmente consultabili anche da noi. Tra questi, risultano:

- Un file JSON in cui sono elencati tutti gli autori che hanno pubblicato almeno una *Submission* sul subreddit, assieme al numero di *Submissions* da loro pubblicate, generato attraverso un programma in Python che conta il numero di volte in cui il campo *author* è uguale in più *Submissions*. Un esempio di dato presente in questo file è:

```
{  
  "name" : "Big_Beyotch",
```

```
"submissions" : "192"  
}
```

- Un file in cui sono presenti tutti gli output relativi alla sentiment analysis, sia quelli per anno sia quelli assoluti, grazie al quale abbiamo potuto poi generare i grafici sulla sentiment analysis e farne a proposito alcune considerazioni utili alla ricerca.
- Due file, uno per gli anni e uno per i mesi, in cui è scritto il numero di *Comments* e di *Submissions* pubblicate in quel determinato anno o mese.
- Un file che elenca tutto il testo nel subreddit (*Comments* + *Submissions*) suddiviso nei vari anni, a partire dal quale abbiamo ricavato altri file che elencano le parole singole e i bigrammi più comuni nel tempo. Importante in questo è stata l'implementazione di un sistema di rilevamento delle *stopwords*, ovvero le "parole vuote" del testo, che non hanno di per sé significato (e che spesso sono molto comuni nei testi, basti pensare alle congiunzioni come "e", "che", ecc...), le quali sono state escluse dal conteggio per il risultato finale. Questo è stato fatto grazie alla libreria *spacy*; nonostante anche la libreria *wordcloud* presentasse un modo per escludere le *stopwords*, abbiamo rilevato che quello di *spacy* risulta più efficace (queste due librerie, insieme alle altre utilizzate, sono presentate nello specifico al Capitolo 4.1.2).

A partire da questi dati in forma analizzata, abbiamo potuto svolgere le nostre analisi che poi sono state rappresentate attraverso dei grafici.

4.2.3 Generazione dei grafici

Grazie al materiale citato nelle due sottosezioni precedenti e alla libreria di Python *matplotlib* siamo stati in grado di scrivere del codice per generare alcuni grafici, utili per la visualizzazione della ricerca effettuata. Un grafico per cui la rappresentazione è stata un po' difficoltosa è quello *Submissions/Autori* (figura 5.11): presentando esso infatti esso dei numeri molto elevati, è stato difficile rappresentarli graficamente rendendo possibile la visualizzazione e la consultazione. Abbiamo inoltre realizzato grafici sulla distribuzione di *Comments* e *Submissions* per anno e per mese, wordcloud per anno e per mese e grafici sulla sentiment analysis, in particolare sulla positività e sulla soggettività, anche questi per anno e per mese.

Capitolo 5

Analisi e grafici

In questo Capitolo presentiamo i dati ottenuti attraverso le API di Pushshift, insieme alle analisi eseguite su di essi. In particolare, ci siamo interessati ai *Comments* (i commenti) e alle *Submissions* (le singole pubblicazioni).

Sulla base dei dati di partenza, abbiamo creato delle rappresentazioni grafiche in modo da poterli analizzare adeguatamente, sia in maniera unica che confrontandole tra loro; ci siamo serviti inoltre di *CoinMarketCap*¹ per poter avere, a scopo di confronto, un grafico di andamento del prezzo del Donut (in corrispondenza con il dollaro americano).

5.1 Dataset

Grazie alle API di Pushshift, descritte nel Capitolo 4, siamo riusciti a ricavare un dataset composto da *Submissions* (Gennaio 2018 - Settembre 2022) e *Comments* (Aprile 2018 - Settembre 2022).

¹(<https://coinmarketcap.com/currencies/donut/>)

La singola *Submission* è rappresentata nel seguente modo:

- id: L'id univoco della *Submission*.
- title: Il titolo della *Submission*.
- text: Il testo contenuto nella *Submission*.
- total_awards_received: Il numero di Reddit awards che sono stati assegnati alla *Submission*.
- author: L'autore della *Submission*.
- num_comments: Il numero di commenti alla *Submission*.
- score: Il punteggio della *Submission*, ovvero il numero di upvotes meno il numero di downvotes.
- is_crosspostable: Se la *Submission* può essere direttamente ripubblicata (appunto, *crosspostata*) in un altro subreddit.
- num_crossposts: Il numero di volte che è stato fatto il *crosspost* di una *Submission*.
- is_video: Vero se la *Submission* è un video, falso altrimenti.
- over_18: Vero se la *Submission* è protetta da un limite d'età, falso altrimenti.
- time: La data e l'ora in cui la *Submission* è stata pubblicata.
- epochTime: La data e l'ora in cui la *Submission* è stata pubblicata, in formato epoch.

Di seguito un esempio di *Submission* rappresentata in JSON nel nostro dataset:

```
{
  "id": "mhyzul",
  "title": "Anyone else surprised ETH is up and BTC is down.",
  "text": "Something is changing",
  "total_awards_received": 0,
  "author": "whiteninja123",
  "num_comments": 15,
  "score": 1,
  "is_crosspostable": true,
  "num_crossposts": 0,
  "is_video": false,
  "over_18": false,
  "time": "2021-04-01 18:26:25",
  "epochTime": "1617301585"
}
```

Il singolo *Comment* è rappresentato nel seguente modo:

- id: L'id univoco del *Comment*.
- submission_id: L'id univoco della *Submission* a cui il *Comment* si riferisce.
- body: Il contenuto scritto del *Comment*.
- author: L'autore del *Comment*.

- time: La data e l'ora in cui il *Comment* è stato pubblicato.
- epochTime: La data e l'ora in cui il *Comment* è stato pubblicato, in formato epoch.

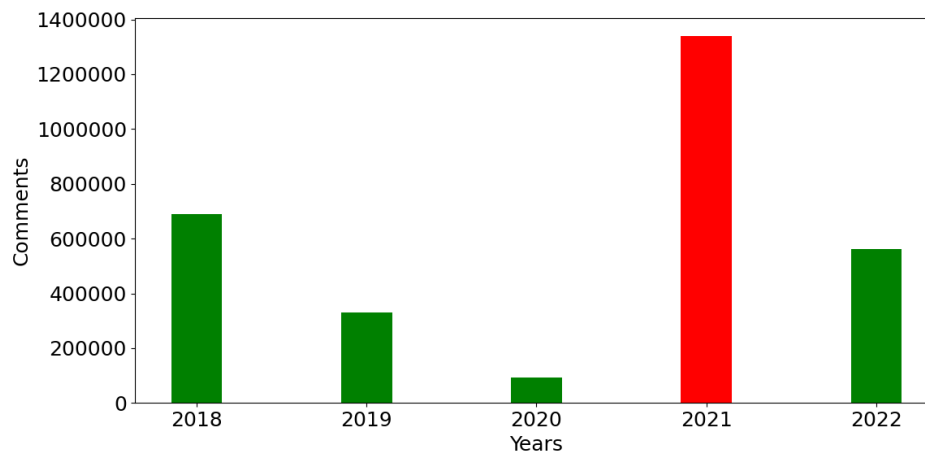
Di seguito un esempio di *Comment* rappresentato in JSON ricavato dal nostro dataset:

```
{  
  "id": "g70molj",  
  "submission_id": "t3_j1pn3y",  
  "body": "No one is surprised",  
  "author": "drunkshihtzu2",  
  "time": "2020-09-29 03:37:39",  
  "epochTime": "1601350659"  
}
```

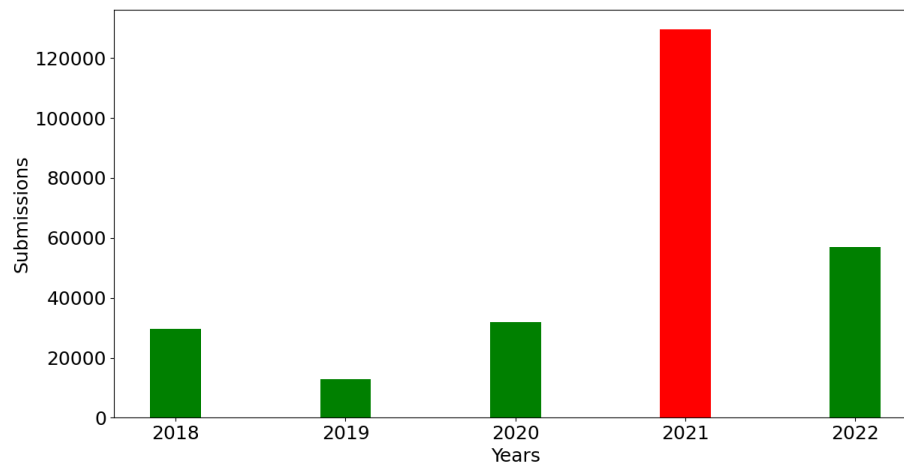
Abbiamo ottenuto alla fine un totale di 3.393.161 *Submissions* pubblicate da 172.294 autori differenti e 21.080.867 *Comments*, pubblicati da 599.286 autori differenti.

5.2 Distribuzione dei dati

5.2.1 *Comments* e *Submissions*

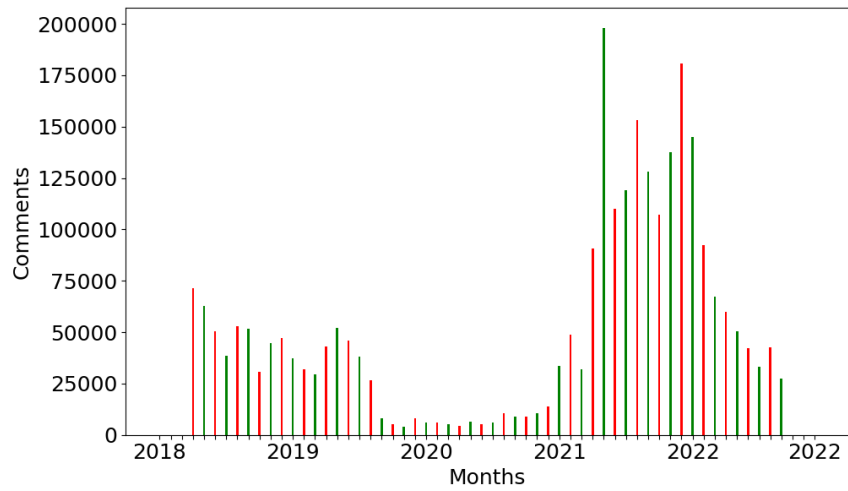


(a) Distribuzione *Comments* per anno.

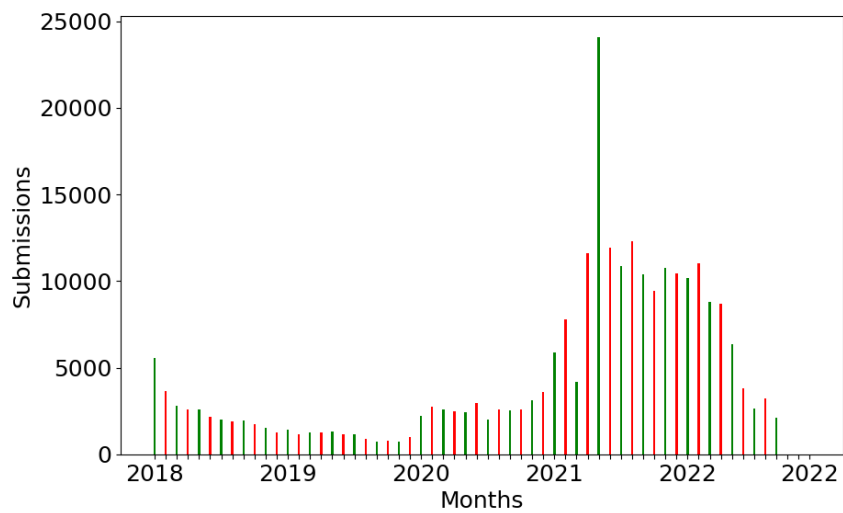


(b) Distribuzione *Submissions* per anno.

Figura 5.1: Distribuzione annuale.



(a) Distribuzione *Comments* per mese.



(b) Distribuzione *Submissions* per mese.

Figura 5.2: Distribuzione mensile.

Come prima analisi, ci siamo soffermati sulla distribuzione mensile e an-

nuale dei *Comments* e delle *Submissions*: in figura 5.1a mostriamo la distribuzione dei *Comments* per anno e in figura 5.1b la distribuzione delle *Submissions* per anno, mentre in figura 5.2a e 5.2b mostriamo la distribuzione rispettivamente dei *Comments* e delle *Submissions* per mese.

Dai grafici, notiamo innanzitutto che il picco di attività nel subreddit è avvenuto nel 2021, in particolar modo intorno al mese di maggio. Prima di questo picco, però, notiamo un'attività in diminuzione a partire dal 2018 e che raggiunge il suo minimo nel 2020. A seguito del grande incremento nel 2021, possiamo osservare un trend nuovamente in discesa: l'attività del 2022 è infatti paragonabile a quella del 2018.

Il confronto più chiaro di queste analisi viene dal rapporto con il grafico Donut/USD (vedi Figura 5.3) ottenuto da CoinMarketCap, che ci mostra una sorta di coerenza con l'attività del subreddit. Da questi grafici, possiamo notare un incremento di attività nel subreddit che va di pari passo con un incremento di valore del Donut. Questo fenomeno è indice del fatto che l'affluenza nel subreddit ha contribuito alla crescita del valore della moneta, e viceversa.



Figura 5.3: Prezzo Donut to USD.

Notiamo come, sia per i *Comments* che per le *Submissions*, gli anni meno attivi siano stati il 2019 e il 2020, mentre il 2021 superi tutti gli altri per un grosso margine. Questo è avvenuto in quanto il 2021 è stato un anno



Figura 5.4: Prezzo Ethereum to USD.

d'oro per le criptovalute in generale, e per Ethereum nello specifico, fatto che si può riscontrare anche dal grafico del prezzo di Ethereum (vedi 5.4); durante quest'anno, infatti, il suo prezzo ha superato i \$4000 a Luglio[1] e ha raggiunto il suo tutt'ora attuale *all-time-high* di \$4815 a Novembre.[2] Considerando che l'argomento del subreddit è proprio il trading di Ethereum, possiamo aspettarci che questa grande affluenza venga dal fatto che molta gente si è probabilmente rivolta a ethTrader per cercare consigli e consulenze finanziarie. Osserviamo infine un doppio calo nel 2022, sia per quanto riguarda l'interesse nel subreddit, sia per quanto riguarda il prezzo del Donut e di Ethereum, coerentemente con quanto affermato in precedenza.

5.2.2 Parole (Wordcloud)

In seguito, abbiamo svolto un'analisi riguardante le parole più utilizzate nel subreddit in generale (prendendo dati dall'unione di tutti i *Comments* e di tutte le *Submissions*), sia suddividendo il dataset per anno che in totale, e abbiamo scelto il wordcloud come metodo di rappresentazione grafica dei dati ottenuti.

Le figure del 2018 (5.6, 2019 (5.7) e 2020 (5.8) appaiono tra loro abbastanza simili, mostrando termini generici legati soprattutto alla blockchain e all'ecosistema di Ethereum, come "ETH", "Ethereum" e "Bitcoin". Il 2020, però, si distingue per la presenza maggiore del termine "DeFi" (*Decentralized finance*, ovvero finanza decentralizzata), ormai comunissimo nel settore della blockchain. La figura 5.9, riferita all'anno 2021, presenta alcuni termini tecnici del mondo del trading come "buy" e "dip" (ovvero i "tuffi", che nel mondo degli investimenti rappresentano una caduta di prezzo), da cui possiamo dedurre un aumento dell'interesse dell'utenza verso i concetti specifici degli investimenti, più che verso il meccanismo che li circonda. La figura 5.10, ovvero il wordcloud dell'anno 2022, ci mostra alcune tendenze già osservabili nel 2021: è presente infatti una crescita d'utilizzo di alcuni termini importanti, primo fra tutti *NFT*, acronimo di *Non fungible tokens*, ovvero dei particolari token utilizzati come atto di proprietà e/o autenticità di un certo bene non per forza digitale, seppure abbiano trovato campo di applicazione in primis nell'arte digitale. Questa tecnologia ha riscosso molto successo negli ultimi due anni, e il fatto che sia stato un argomento molto discusso sul subreddit man mano che incrementava la sua popolarità ci fa capire verso quali direzioni si è mosso l'interesse dei più, seppure i reward

che Reddit offre per l'attività nel subreddit non abbiano nulla a che vedere con token non fungibili.

La figura 5.10, ovvero il wordcloud generale del subreddit, ci mostra nuovamente il termine *NFT*, seppure più piccolo rispetto alla sua apparizione negli anni 2021 e 2022, indizio del fatto che l'interesse accumulato negli ultimi 2 anni è stato tale da rendere possibile a questa tecnologia di imporsi come uno dei trend più importanti nell'attuale mondo delle criptovalute, e con un grande futuro per quanto riguarda la discussione sui rewards e sui Social Network. La stessa cosa vale per la parola *Metaverse*, ovvero il Metaverso, un nuovo modo di fruire Internet in maniera completamente immersiva, in cui si accede a spazi virtuali attraverso il proprio avatar. Questo termine compare nel wordcloud del 2022 e in quello generale, a riprova di quanto è recente l'immaginario del metaverso. Anche in questo possiamo vedere molto facilmente un'interesse verso il mondo dei rewards e dei Social Network: basti pensare al fatto che il Metaverso potrebbe essere semplicemente considerato come un Social Network più immersivo di quelli attualmente in circolazione, e che i rewards potrebbero essere delle personalizzazioni o funzioni aggiuntive a tale Social Network.

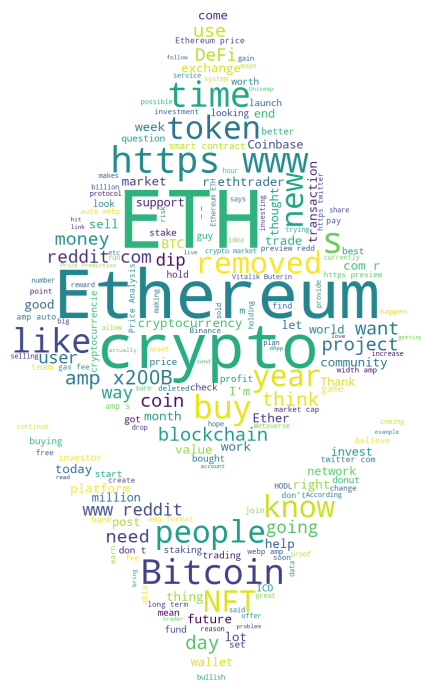


Figura 5.5: Wordcloud dell'intero subreddit.

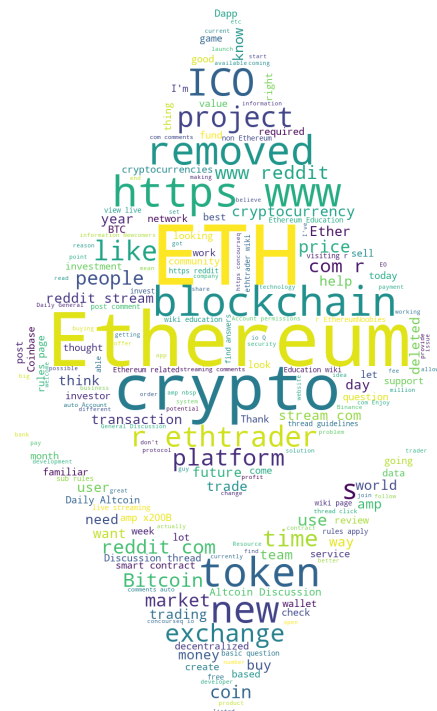


Figura 5.6: Wordcloud per l'anno 2018.

5.3 Autori e numero di pubblicazioni

Un'altra ricerca attuata è stata quella sui vari utenti attivi nel subreddit e sul numero di *Submissions* da loro pubblicate, con l'obiettivo di osservare questa distribuzione e trarne poi delle conclusioni.

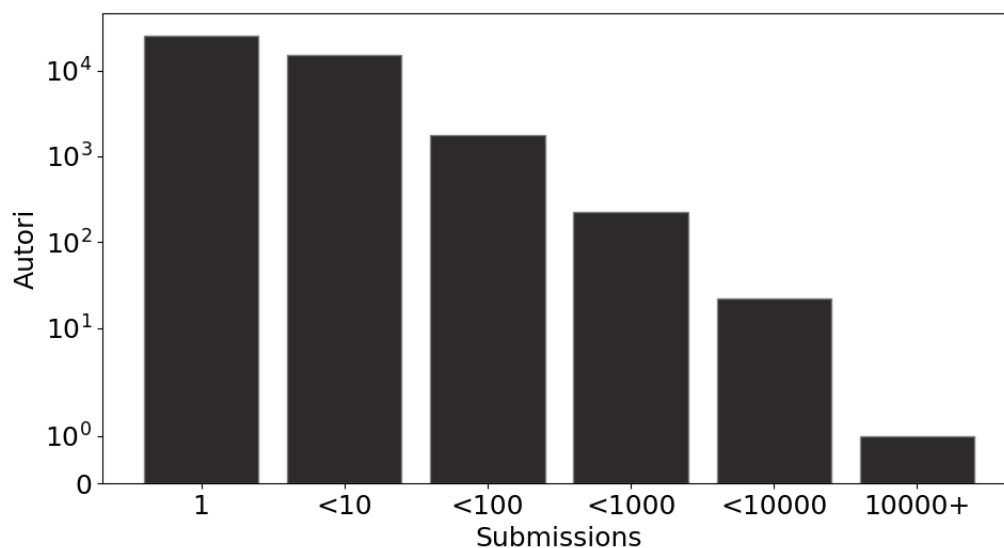


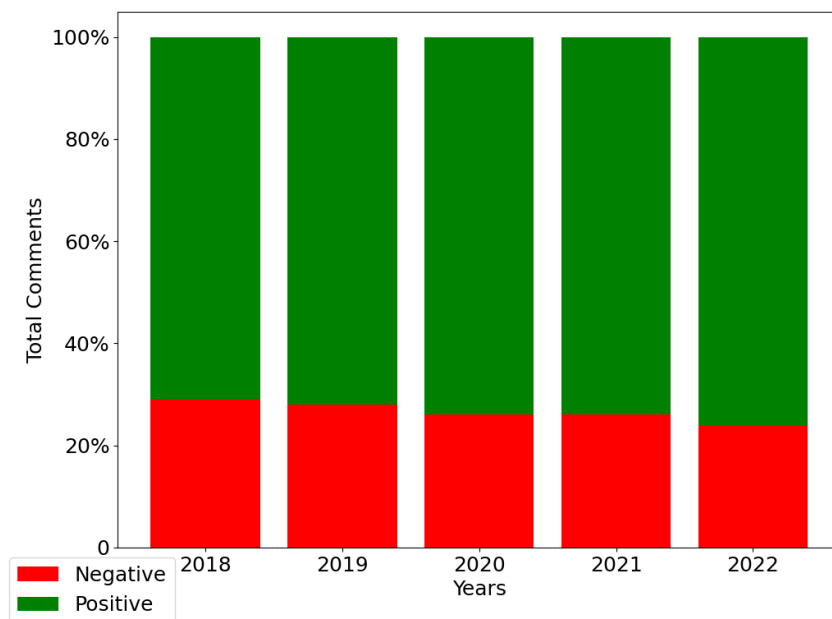
Figura 5.11: Grafico *Submissions*/Autori.

Come possiamo notare nella figura 5.11, la maggior parte degli autori ha pubblicato meno di 10 *Submissions*. Man mano che aumenta il numero di *Submissions*, sono sempre meno gli autori che ne hanno pubblicate; basti vedere quanto basso è il numero di autori che hanno pubblicato tra 1000 e 10000 *Submissions*. Inoltre, solo un autore ha pubblicato più di 10000 *Submissions*. Questo è probabilmente dovuto dal fatto che Reddit, essendo comunque un Social Network, è molto accessibile e spesso tende a essere utilizzato anche solo per fare una domanda o esprimere un'opinione in una

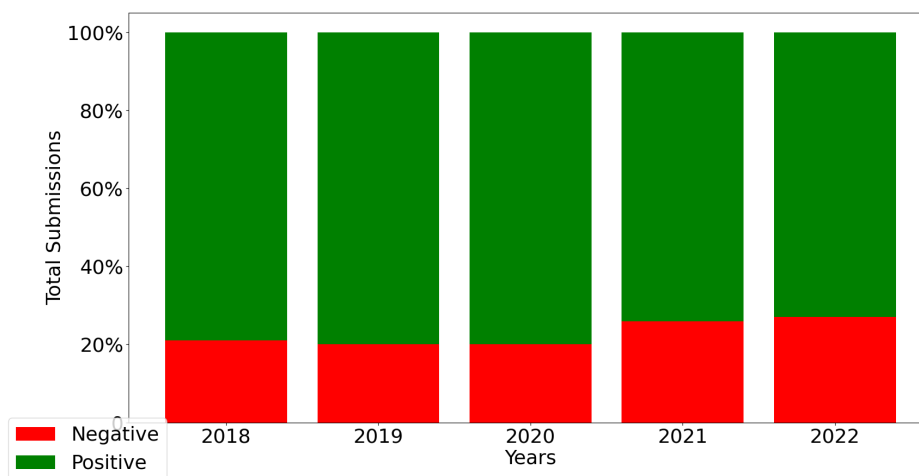
singola istanza, o comunque in poche occasioni. A partire da questo concetto possiamo notare che gli utenti più attivi e che più hanno pubblicato sul subreddit sono pochi rispetto al gran numero di utenti entrati in contatto con il subreddit, ma allo stesso tempo tengono alla community e provano a farla crescere, dando il loro contributo. Scendendo nel particolare, abbiamo scoperto che l'utente più attivo nella community (nonché unico ad aver pubblicato più di 10000 *Submissions*), *EthTraderCommunity* ha un account moderatore e aggiorna giornalmente gli utenti dello stato del subreddit, delle news sullo stesso e in generale sul mondo di Ethereum e delle criptovalute, e quant'altro. L'assegnazione dei *Donuts* varia tra gli utenti e i moderatori; questi ultimi hanno una quota fissa assegnata loro, indipendentemente dal karma ricevuto (metodo per cui invece dipende la quantità di *Donuts* ricevuti per gli utenti), quindi essi non sarebbero neanche tenuti a pubblicare più di un normale utente, se il loro obiettivo fosse solo l'acquisizione dei *Donuts*: possiamo quindi dire che questa vasta partecipazione è dovuta a una genuina volontà di aiutare e far crescere la comunità da parte dei moderatori.

5.4 Sentiment analysis

Abbiamo anche attuato una ricerca di sentiment analysis sul subreddit, in modo da avere un'idea generale del sentimento nel contenuto che gli utenti hanno messo in *Comments* e *Submissions*: in particolare, abbiamo analizzato gli aspetti di positività e soggettività.

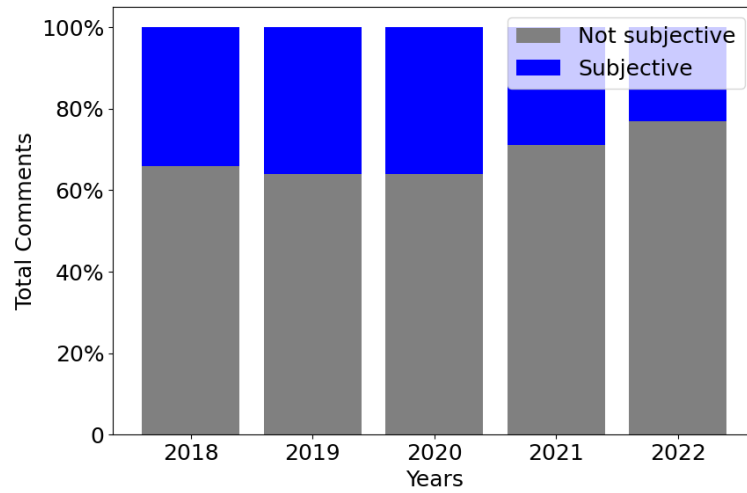


(a) Positività dei *Comments* per anno.

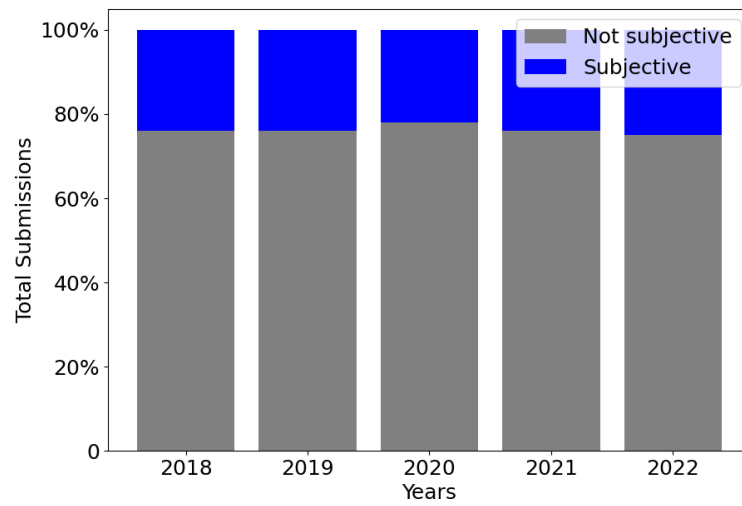


(b) Positività delle *Submissions* per anno.

Figura 5.12: Positività/Negatività per *Comments* e *Submissions*.



(a) Soggettività dei *Comments* per anno.



(b) Soggettività delle *Submissions* per anno.

Figura 5.13: Soggettività per *Comments* e *Submissions*.

Per quanto riguarda la positività e la negatività, notiamo che i *Comments* (figura 5.12a) hanno una tendenza crescente in positività, mentre le *Submissions* (figura 5.12b) tendono invece a essere man mano meno positive, seppur mantenendo comunque dei valori simili a quelli dei *Comments*. In valori assoluti, per ogni singolo anno la positività risulta sempre maggiore nelle *Submissions* rispetto che nei *Comments*. Possiamo immaginare che un *Comment* generico abbia più probabilità di essere negativo rispetto a una *Submission* generica, in quanto quest'ultima spesso si presenta come un dato di fatto o comunque un'affermazione che non si sbilancia troppo, al fine di generare una discussione, mentre un *Comment* è più un giudizio o un'opinione, e quindi può essere negativo in base all'opinione dell'utente commentatore. Questo concetto verrà poi ripreso anche per l'analisi della soggettività.

Riteniamo che la negatività riscontrata in queste analisi derivi dal fatto che insieme alla diffusione sempre più vasta degli argomenti criptovalute e blockchain è arrivata una ventata di negatività e poca fiducia verso questo mondo, considerato ancora da molti poco affidabile o addirittura truffaldino. Pur tenendo questo discorso in considerazione, possiamo comunque affermare tranquillamente che la maggior parte dei *Comments* e delle *Submissions* presenti nel subreddit è considerabile positivo, e ciò è coerente con il fatto che il subreddit è basato sul mondo di Ethereum e quindi possiamo aspettarci un utente medio interessato ed entusiasta sulla maggior parte degli argomenti trattati.

Possiamo inoltre osservare un grande distacco tra *Comments* e *Submissions* per quanto riguarda la soggettività (5.13a e (5.13b)), il cui motivo di fondo riteniamo essere la natura stessa dei *Comments*, dei commenti pub-

blicati appositamente per esprimere un'opinione e che quindi saranno mediamente più carichi di significato soggettivo che le *Submissions*. In valori assoluti, però, sia per i *Comments* che per le *Submissions* è maggiore il contenuto non soggettivo rispetto a quello soggettivo: d'altro canto, Ethereum stesso si basa proprio su convenzioni e standard uguali per tutti, e allo stesso modo l'analisi economica non è qualcosa di soggettivo. Troviamo dunque coerente una maggiore affluenza di contenuto oggettivo rispetto al contenuto soggettivo.

Come considerazione finale, possiamo però vedere che né per la positività né per la soggettività abbiamo dei picchi o dei crolli particolarmente evidenti; questo è coerente con il metodo attuale che il subreddit usa per distribuire i rewards, il quale favorisce la quantità alla qualità/personalità del contenuto, e ci mette davanti all'idea che, in un futuro, per evolvere e migliorare il sistema di rewarding potrebbe essere interessante tenere in conto anche l'impegno e la personalità che l'utente mette in ciò che pubblica.

Capitolo 6

Conclusioni

Possiamo dire che i risultati ottenuti sono stati soddisfacenti: siamo riusciti a ottenere una panoramica abbastanza vasta sul subreddit, e da questa abbiamo potuto fare molte considerazioni interessanti. Dal momento che il nostro interesse principale rimane comunque sulla decentralizzazione e su come i Social Network si rapportano a essa, certamente avere più informazioni sui Donuts sarebbe stato interessante; è vero anche che la nostra analisi ha coperto molti campi relativi ai contenuti del subreddit a partire dal 2018 fino ad arrivare al 2022, quindi le considerazioni hanno potuto anch'esse ampliarsi e comprendere vari ambiti.

I risultati ci hanno mostrato una certa coerenza nell'aumento di interesse e interazioni all'interno di ethTrader confrontato con ciò che accadeva nel mondo esterno e con le novità introdotte da Reddit a riguardo del mondo criptovalute. Possiamo infine dire che per questo microuniverso, alla fine dei conti, vale ciò che avevamo inizialmente pensato nell'osservazione della sentiment analysis (nel Capitolo 5.4): l'aggiunta di premi in criptovalute

secondo un principio come quello dei Reddit Community Points, o più specificamente dei Donuts, tende a valorizzare più la quantità che la qualità, spingendo gli utenti a postare di più, ma non favorendo in modo particolare i contenuti più meritevoli. Questo è però comprensibile se pensiamo a Reddit come una piattaforma che nasce comunque come centralizzata, e che sta solo muovendo dei passi verso la decentralizzazione, rimanendo comunque con la stessa struttura di base; un Social Media completamente decentralizzato (ne abbiamo parlato nel Capitolo 2.4) potrebbe invece certamente essere molto più libero per certi versi, e quindi sfruttare a dovere il concetto di premi in criptovaluta allo scopo di migliorare effettivamente la qualità dei contenuti e, per esteso, l'esperienza di navigazione degli utenti.

6.1 Possibili lavori futuri

Per quanto riguarda i possibili sviluppi di questa ricerca, il primo viene in relazione a ciò che abbiamo detto poco fa su riguardo dei Donuts: sarebbe interessante recuperare e analizzare i file .csv dei *Contribution Score* citati nel Capitolo 3.2.1, per poi magari attuare delle analisi temporale o anche dei confronti tra il file originale e il file pubblicato una settimana dopo (modificato grazie alla discussione della community). In secondo luogo, sarebbe da considerarsi anche una ricerca più approfondita sulla qualità del contenuto, magari scendendo nello specifico fino al singolo post per vedere quanto karma (e quindi quanta criptovaluta) ha portato all'utente, e quindi vedere quanto la community effettivamente apprezza e premia i vari tipi di contenuto che incontra sulla piattaforma.

Ulteriori analisi potrebbero essere interessanti da fare fuori dall'ecosistema di ethTrader, ad esempio in altri subreddit che hanno introdotto i Reddit Community Points, come CryptoCurrency (Moons) e FortniteBR (Bricks). Potrebbe essere una buona idea controllare anche l'andamento di questi subreddit e delle criptovalute a loro associati.

Infine, un tipo di ricerca possibile nel futuro potrebbe essere a riguardo dell'aspetto economico dei token offerti come ricompensa, da noi quasi del tutto ignorato se non per un confronto con l'attività del subreddit (nel Capitolo 5.2.1), che, per quanto dimostrativo, non si sofferma più di tanto sulle possibilità finanziarie del meccanismo. Queste ultime invece sono presenti e svolgono anzi un ruolo fondamentale, proprio a motivo della natura intrinseca delle criptovalute.

Bibliografia

- [1] Ethereum, the second-biggest cryptocurrency, soars above \$4,000 to hit a new record high. <https://www.euronews.com/next/2021/05/10/ethereum-the-second-biggest-cryptocurrency-soars-above-4-000-to-hit-a-new-record-high>. Updated: 19/07/2021.
- [2] Ethereum's price history (2013 – 2022, \$). <https://www.globaldata.com/data-insights/financial-services/ethereums-price-history/#:~:text=From%202019%20to%202021%2C%20Ether,%244%2C815%20on%20November%209%2C%202021>. Updated: 25/12/2022.
- [3] Revealing this year's reddit recap, where we highlight how redditors kept it real in 2022. <https://www.redditinc.com/blog/reddit-recap-2022-global>. Updated: 08/12/2022.
- [4] Global social media statistics. <https://datareportal.com/social-media-users#:~:text=Analysis%20from%20Kepios%20shows%20that,since%20this%20time%20last%20year>. By Kepios.
- [5] The 10 most important social media statistics for 2022. <https://blog.gwi.com/marketing/social-media-stats/>. By GWI.
- [6] Vitalik Buterin. Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform. Whitepaper, Massachusetts Institute of Technology, 2014.

- [7] Quinn DuPont. Experiments in algorithmic governance: A history and ethnography of “the dao,” a failed decentralized autonomous organization. *Bitcoin and beyond*, page 157–177, 2017.
- [8] Barbara Guidi. When blockchain meets online social networks. *Pervasive and Mobile Computing*, 62:101–131, 2020.
- [9] Kristina Kapanova & Barbara Guidi & Andrea Michienzi & Kevin Koidl. Evaluating posts on the steemit blockchain: Analysis on topics based on textual cues. *GoodTechs ’20: Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good*, page 163–168, 2020.
- [10] Barbara Guidi & Andrea Michienzi. The side effect of erc-20 standard in social media platforms. *International Conference on Social Informatics*, pages 114–127, 2022.
- [11] Satoshi Nakamoto. Bitcoin: A Peer-to-Peer Electronic Cash System. Whitepaper, Massachusetts Institute of Technology, 2008.
- [12] Barbara Guidi & Andrea Michienzi & Laura Ricci. Steem blockchain: Mining the inner structure of the graph. *IEEE Access*, 8:210251–210266, 2020.
- [13] Barbara Guidi & Andrea Michienzi & Laura Ricci. Steem blockchain: Mining the inner structure of the graph. *Mobile Networks and Applications*, 26:2099–2110, 2021.
- [14] Stuart Haber & W. Scott Stornetta. How to time-stamp a digital document. *Journal of Cryptology*, 3:99–111, 1991.