

Criptovalute: Popolarità e Valore

Daniele Adduci, Informatico, and Davide Brinati, Economista and Stefano Daraio, Statistico
(833222, d.adduci@campus.unimib.it) (771458, d.brinati@campus.unimib.it) (718443, s.daraio@campus.unimib.it)

Università degli studi di Milano Bicocca

Dopo i recenti sviluppi nel corso degli ultimi anni sta diventando chiaro come le criptovalute stiano tendendo ad abbandonare un ruolo marginale all'interno della società per ricoprire uno attivo. Le valute digitali non sono emesse o regolate da un istituto bancario e questo le rende libere da ogni tipo di controllo diretto da parte delle banche centrali nazionali, e il loro valore è determinato fondamentalmente da un tema di domanda e offerta. Quando la domanda aumenta più rapidamente dell'offerta, la valuta aumenterà di prezzo. E' possibile quindi spiegare, interpretare e quindi prevedere il variare del valore delle criptocurrency?

0 INTRODUCTION

Le criptovalute sono a tutti gli effetti delle monete virtuali che permettono di compiere operazioni in sicurezza ed anonimato.

Esse, possono essere considerate come degli strumenti digitali che basano il proprio funzionamento sui principi della crittografia.

Tra le principali criptovalute in circolazione vale la pena di ricordare sicuramente Bitcoin, Ethereum, i veri e propri giganti. La loro fetta di mercato, però, ha iniziato a guardare con sospetto l'arrivo di altri player tra cui Ripple e Litecoin, solo per citarne qualcuna.

Le criptovalute sono state pensate con lo scopo di eliminare l'intermediario finanziario che emette denaro contante, e per questo sono svincolate da qualsiasi tipo di autorità. Questa caratteristica continua a costare caro alle valute virtuali: esse stentano a ricevere l'approvazione di numerosi governi ed istituti bancari, che non sembrano fidarsi della loro autonomia e dell'anonimato che garantiscono.

Il funzionamento delle criptovalute si basa sulle attività di mining [Fig.1], processi in cui i computer risolvono problemi matematici generando in cambio tali monete virtuali. Fondamentale è anche il concetto di Blockchain, registro pubblico di tutte le transazioni in criptovalute.

Le criptovalute non possono essere comprese usando i fattori macroeconomici che di solito influenzano i mercati azionari, valutari e delle materie prime. Ma, esisterebbero altre strategie comuni, in grado di interpretare al meglio gli andamenti dei criptomercati.

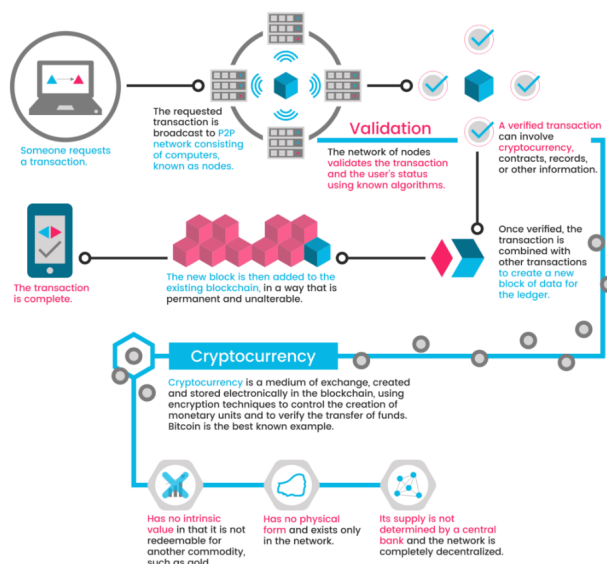


Fig. 1. come i miners creano valuta e confermano le transazioni.

1 OBIETTIVO

Per analizzare la variazione delle principali criptocurrency si è deciso di studiarne varie possibili cause. Come primo elemento abbiamo considerato quello che per gli investitori è noto come **""effetto momentum""** ovvero la tendenza di un determinato mercato a crescere qualora sia salito per un certo lasso di tempo e decrescere qualora sia mediamente sceso. Il secondo aspetto considerato è stato il più imprevedibile **"effetto virale"**.

Per analizzare il primo sono state chiamate le API del sito Poloniex relativo agli andamenti delle maggiori criptovalute, per il secondo invece l'idea è di utilizzare il social network Twitter, tramite il quale considerare specifici tweets presi sia da dati storici che in tempo reale e la popolarità di specifiche parole chiave tramite google trends.

2 ARCHITETTURA DATI

Per poter analizzare gli effetti visti precedentemente è necessario definire una struttura dati che ci permetta di ottenere tutte le informazioni necessarie [Fig. 2]. Avremo tre macroaree rappresentanti diversi flussi dati che andremo poi ad utilizzare per le analisi :

- 1) Twitter : nel primo flusso vengono estratti i tweet storici ed in tempo reale per le principali criptocurrency, nel primo caso tramite scraping mentre nel secondo utilizzando le API ufficiali di twitter e l'architettura producer-consumer di kafka. Tutti i dati vengono infine salvati su MongoDB.
- 2) Poloniex : connettendoci tramite API a questa piattaforma exchange di criptovalute è stato possibile scaricare una serie di informazioni sull'andamento delle principali criptovalute.
- 3) Google Trends : tramite il servizio gratuito offerto da google è possibile avere un indice sulle ricerche effettuate per una serie di parole chiave.

2.1 TWITTER

Per svolgere le nostre analisi abbiamo pensato di far ricorso ad uno dei social network maggiormente utilizzati e che ci permettesse di scaricare le informazioni necessarie, ovvero Twitter. Su questo social infatti sono pubblicati tweet dai maggiori esponenti politici e della nostra società che possono influenzare l'opinione pubblica e non solo. Per accedere e poter quindi scaricare questi dati, il social, rilascia delle API che permettono all'utente di "catturare" il 3% del traffico totale della rete in quel momento. Abbiamo quindi salvato lo storico dei tweet tramite uno script di scraping, e utilizzando l'architettura di Kafka abbiamo effettuato lo streaming dei tweets per alcuni giorni, e li abbiamo raccolti in MongoDB.

2.1.1 STREAMING KAFKA E MONGODB

Per poter fornire un architettura adeguata alla velocità richiesta dal progetto ci siamo serviti, per il recupero dei dati, dell'ecosistema Hadoop, ed in particolare di Kafka. Questa piattaforma si serve di:

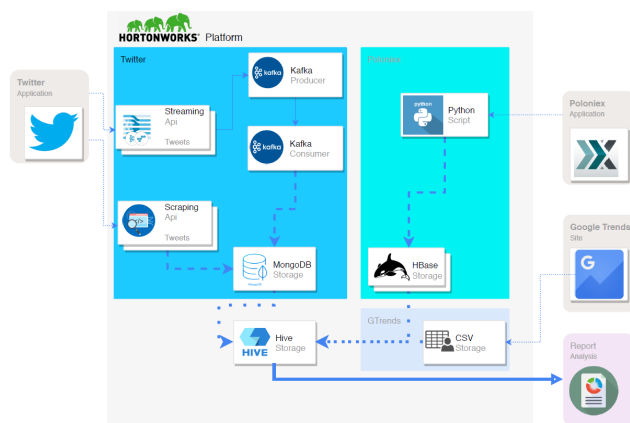


Fig. 2. schema rappresentante il flusso dati.

- 1) un Producer, per filtrare e catturare i tweets in streaming e "pubblicarli" su un topic precedentemente definito.
- 2) un Consumer che, effettuando l'iscrizione al topic creato, riesce a reperire ed elaborare i messaggi inviati.

Un esempio di tale struttura è illustrato dalla Fig 3. L'accesso alle API di Twitter è stato reso possibile da tweepy, una libreria python che permette di impostare il collegamento grazie a quattro token (consumer key, consumer secret, access token, access secret) resi disponibili direttamente da Twitter dopo aver creato un'applicazione. Per quanto riguarda lo storage dei dati recuperati dal Consumer abbiamo pensato di immagazzinarli in MongoDB ogni volta che il Consumer riusciva a recuperare un tweet dal topic; questo in quanto la natura delle informazioni ottenute si prestano ad essere salvate all'interno di un database documentale avendo alcuni campi un numero non definito di valori (ad esempio il campo che identifica tutti gli hashtag del tweet), e avendo la possibilità teorica di salvare dati multimediali come immagini. Abbiamo scaricato i tweets del 22-23 Settembre 2018, ovvero circa 5500 documenti, in prospettiva di lasciare in esecuzione la raccolta dei tweet per un periodo significativo successivo.

Il primo terminale mostra la sessione di login su un server VM. L'utente 'vmadmin' si connette a 'maria_dev@127.0.0.1' e lancia lo script 'python3 producer_v3.py'. Lo script configura il Producer e lo Streamer, e inizia lo streaming. Il secondo terminale mostra l'output dello script: 'New tweets sent', 'Looking for new tweets', e 'Storing...' ripetute.

```

maria_dev@sandbox:~$ login as: vmadmin
Using keyboard-interactive authentication.
Password:
Last login: Sat Sep 22 13:35:37 2018 from 149.132.31.93
[vmadmin@dsvm03 ~]$ ssh maria_dev@127.0.0.1 -p 2222
maria_dev@127.0.0.1's password:
Last login: Sat Sep 22 13:35:48 2018 from 172.17.0.1
[maria_dev@sandbox ~]$ python3 producer_v3.py

Getting config.yaml as configuration file

Setting Producer
Authenticating...
Setting Streamer
Stream flowing...

maria_dev@sandbox:~$
New tweets sent
New tweets sent
Looking for new tweets
New tweets sent
Storing...
New tweets sent
New tweets sent
New tweets sent
New tweets sent
Storing...
New tweets sent
Storing...
New tweets sent
New tweets sent
New tweets sent
Looking for new tweets
New tweets sent
New tweets sent
Storing...
New tweets sent
Storing...

```

Fig. 3. Esempio di Streaming : Producer-Consumer.

2.1.2 DATI STORICI

Uno dei principali limiti nello scaricare dati in streaming è quello di non avere la possibilità di recuperare le informazioni passate.

Per sopperire a questa restrizione abbiamo utilizzato uno script in python che ci permetta di fare scraping su twitter.

Questa tecnica di estrazione dati presenta vantaggi e svantaggi.

I vantaggi sono la possibilità di scaricare un numero potenzialmente indefinito di tweet utilizzando inoltre un range temporale personalizzabile.

Gli svantaggi invece sono riconducibili alla natura di tale metodologia che non essendo una fonte dati ufficiale autorizzata da Twitter (infatti non è necessario l'utilizzo di alcun token per poterlo effettuare) può portare a comportamenti inattesi in fase di ricerca; le parole cercate sono inoltre all'interno del tweet ma non sono necessariamente un hashtag e, con questa metodologia, non è garantita la correttezza e coerenza dei dati scaricati. Come è possibile vedere nella Table 1. ci sono meno informazioni di quelle ottenute tramite streaming.

Ci teniamo a sottolineare come non siano presenti informazioni sensibili, e le finalità di estrazione di questi dati sono solo a scopi accademici.

Table 1. Confronto variabili Streaming vs Scraping

	Streaming	Scraping
id	Y	Y
screen_name	Y	N
user_name	Y	N
hashtag_indices	Y	N
hashtag_text	Y	N
followers	Y	N
retweet	Y	Y
created_at	Y	Y
text	Y	Y

2.2 POLONIEX

Poloniex come già accennato è una piattaforma di scambio di criptovaluta negli Stati Uniti. La società è stata fondata nel gennaio 2014 da Tristan D'Agosta e ha sede a Wilmington, nel Delaware.

Tramite le API pubbliche fornite da questa piattaforma è possibile avere informazioni aggiornate ogni 5 minuti riguardo le valutazioni delle principali criptovalute e altri indicatori inerenti (tutte le variabili sono presenti nella Table 2.)

Abbiamo arricchito il modello dati con la creazione di variabili presentanti il suffisso `_usd` rappresentanti il valore in dollari delle altcoin (ovvero le criptovalute diverse da bitcoin). Questo si è reso necessario in quanto la conversione di quest'ultime non viene restituito sul dollaro ma sul bitcoin.

Table 2. Elenco campi API Poloniex e informazione se sono stati calcolati o direttamente importati

	Campo_creato
key	Y
currency	N
timestamp	N
data	N
time	N
high	N
low	N
open	N
close	N
volume	N
quote volume	N
weightedAverage	N
price_usd	N
high_usd	Y
low_usd	Y
open_usd	Y
close_usd	Y
volume_usd	Y
quote_volume_usd	Y
weighted_average_usd	Y

2.2.1 HBASE

Per salvare le informazioni scaricate tramite API pubbliche da Poloniex abbiamo scelto di utilizzare Apache HBase per alcune sue utili ed importanti caratteristiche :

- 1) Schema-less : molto utile considerando che abbiamo un diverso numero di campi valorizzati per ogni riga, ad esempio i campi di conversione da currency a USD sono validi per le sole criptovalute diverse dal bitcoin; permette inoltre una migliore flessibilità, rispetto ad un database relazionale, al cambiare della struttura dati.
- 2) Efficienza in letture/scritture casuali : L'accesso random si riferisce alla possibilità di accedere ai dati in modo non ordinato/sequenziale, ad esempio nel caso di join con le informazioni corrispondenti ai vari tweet.
- 3) Efficienza nell'accesso realtime : Fondamentale inoltre che il database sia predisposto ad un accesso in realtime sia in fase di scrittura avendo un tempo di aggiornamento molto frequente (cinque minuti) che in fase di lettura, per permettere uno sviluppo di analisi su dati attuali.

2.3 GOOGLE TRENDS

Google Trends è un tool gratuito di Google che mostra l'interesse raccolto da determinati argomenti / parole chiave nel corso del tempo.

Come riportato nella guida ufficiale di Google Trends, "Le notizie di tendenza sono basate sulla tecnologia del Knowledge Graph nella Ricerca Google, su Google News e su YouTube per rilevare gli argomenti di tendenza su queste tre piattaforme". E ancora: "Il Knowledge Graph consente alla nostra tecnologia di collegare le ricerche a

cose e luoghi reali. L'algoritmo relativo alle notizie di tendenza raggruppa gli argomenti che sono di tendenza nello stesso momento su Google News, nella Ricerca Google e su YouTube, dopodiché posiziona le notizie in base al picco relativo nel volume di ricerca e al volume assoluto delle ricerche".

Abbiamo quindi sfruttato l'algoritmo di Google per ottenere informazioni riguardo l'interesse nel tempo delle principali criptovalute, per fare ciò abbiamo scaricato i file CSV risultanti dalle ricerche delle seguenti parole chiave ("Bitcoin", "Ripple", "Ethereum").

2.4 HIVE

Abbiamo infine pensato di utilizzare Hive come centro per connettere le diverse fonti di dati precedentemente costruite e poterle interrogare per avere le informazioni desiderate. Inoltre tale soluzione permette di scrivere query grazie una interrogazione semplificata tramite linguaggio HQL.

I campi utilizzati per far comunicare i diversi database e le diverse tabelle sono i seguenti :

- 1) Data : la prima variabile di riferimento è la data, le diverse informazioni in nostro possesso presentano però delle tempistiche di aggiornamento diverse (i Tweet ad esempio possono essere creati in ogni momento mentre l'aggiornamento sulle quotazioni avviene ogni 5 minuti). Abbiamo però deciso di tenere le informazioni con i loro riferimenti di datatime originari, in modo da permettere a seconda dell'analisi necessaria di valutare quale sia la scelta migliore e creare altre tabelle/viste o semplici interrogazioni che considerano i tempi in modo diverso. Per la nostra finalità abbiamo deciso di considerare le informazioni raggruppate con una granularità giornaliera (questo si è reso necessario per valorizzare le informazioni ottenute tramite Google Trends che presentavano tale valore massimo di dettaglio), e quindi non c'è stato bisogno di fare ulteriori operazioni di modifica.
- 2) Valuta : La seconda chiave necessaria è la criptovaluta di riferimento, questa informazione è stata creata in ogni tabella di riferimento in modo da agevolare le operazioni di join.
- 3) Chiave concatenata : Per consentire la storizzazione delle quotazioni è stato necessario sfruttare le informazioni relative a data e valuta. Al fine di ottimizzare il funzionamento dell'architettura abbiamo scelto di concatenare valuta e timestamp espresso in formato linux in modo da permettere una migliore scalabilità in fase di scrittura.

2.5 CONTROLLI

Per garantire il corretto funzionamento di tutte le componenti trattate abbiamo pensato ad alcuni semplici controlli che impediscano di salvare informazioni non corrette. Abbiamo notato ad esempio, che lo streaming su Twitter

non sempre restituiva risultati coerenti con la parola chiave specificata (sembra essere un problema lato Twitter), per evitare quindi di conservare informazioni non rilevanti abbiamo predisposto un controllo su Kafka che verifichi prima di inviare le informazioni su MongoDB che la parola ricercata sia presente nel corpo del tweet (Fig. 4). Un'altra verifica effettuata è stata quella di valorizzare a 0 i campi delle quotazioni delle altcoin per le quali non è stato possibile ottenere da Poloniex la corrispondente in Bitcoin su Dollaro (l'incidenza di tale problema è stata minima, si parla infatti di 2-3 casi su tutto il periodo). Inoltre in questo modo è possibile recuperare successivamente tali valori in Hive e decidere se valorizzarli (ad esempio utilizzando medie o altre statistiche, oppure eliminarli dalle analisi successive).

2.6 RIFLESSIONI E CRITICITÀ

Ipotizzando un aumento importante dei volumi abbiamo pensato a quali potessero essere le maggiori criticità e a come sarebbe possibile gestirle con l'architettura proposta.

- 1) HBase : l'aspetto maggiormente penalizzante in caso di aumento dei volumi è avere una tipologia di chiave rappresenta da un valore progressivo. Questo perché l'architettura è ottimizzata per leggere e scrivere casualmente gestendo in autonomia le dimensioni delle tabelle, suddividendo nei nodi a disposizione le diverse "Regioni". Nel caso la chiave fosse un numero progressivo, al crescere di esso i record tenderebbero ad essere scritti tutti nell'ultimo nodo (problema noto come "hotspotting"). Questo si può risol-

```
try:
    # Decodifico
    dec_message = json.loads(message.value.decode("utf-8"))
    # prendo il corpo del messaggio
    tweet_body = dec_message.get("text")
    # Verifico per quali valute è valido
    for coin in COINS:
        query = QUERY_DICT.get(coin)
        if query in tweet_body:
            # valorizzo il campo coin e scrivo sul DB
            dec_message["coin"] = coin
            print("Storing...")
            tweetdb[COLLECTION].insert_one(dec_message)
    except Exception as e:
        print("Couldn't store tweet\n")
        print("{}\n".format(e.details))
```

Fig. 4. controllo parola cercata nel corpo del tweet.

```
try:
    crypto_datum_insert["high_usd"] = float(crypto_datum["high"]) * btc_datum["high"]
    crypto_datum_insert["low_usd"] = float(crypto_datum["low"]) * btc_datum["low"]
    crypto_datum_insert["open_usd"] = float(crypto_datum["open"]) * btc_datum["open"]
    crypto_datum_insert["close_usd"] = float(crypto_datum["close"]) * btc_datum["close"]
    crypto_datum_insert["volume_usd"] = float(crypto_datum["volume"]) * btc_datum["volume"]
    crypto_datum_insert["quote_volume_usd"] = float(crypto_datum["quoteVolume"]) * btc_datum["quoteVolume"]
    crypto_datum_insert["weighted_average_usd"] = float(crypto_datum["weightedAverage"]) * btc_datum["weightedAverage"]
except Exception as e:
    crypto_datum_insert["high_usd"] = float(0)
    crypto_datum_insert["low_usd"] = float(0)
    crypto_datum_insert["open_usd"] = float(0)
    crypto_datum_insert["close_usd"] = float(0)
    crypto_datum_insert["volume_usd"] = float(0)
    crypto_datum_insert["quote_volume_usd"] = float(0)
    crypto_datum_insert["weighted_average_usd"] = float(0)
return crypto_datum_insert
```

Fig. 5. valorizzazione valori mancanti a 0.

vere generando una UUID a codice durante la fase di recupero dati, che permetterebbe la scrittura in un nodo casuale. Ispirati da OpenTSDB, che colloca il timestamp in mezzo ad altri valori nella rowkey, abbiamo optato per la concatenazione di valuta e timestamp, nonostante avessimo a disposizione un unico cluster.

- 2) Kafka : Abbiamo pensato a due possibili soluzioni. La prima consiste nell'avere più producer ciascuno con un topic ed una query specifica che mandi le informazioni al consumer corrispondente (quindi diverse coppie di producer-consumer). La seconda soluzione prevede un unico producer con un topic avente più parole chiave di ricerca, e la presenza di un consumer per ciascuna di esse.
- 3) MongoDB : una maggiore scalabilità potrebbe passare dal valutare un maggior numero di collection di dati, o valutare se creare DB diversi a seconda delle esigenze, per esempio, suddividere i dati streaming da quelli storici.

3 ANALISI E RISULTATI

Di seguito andremo ad analizzare alcuni risultati prevalentemente descrittivi ed esemplificativi che lasciano spazio ad ulteriore analisi più articolate e/o complesse. In particolare per analizzare l'effetto social non è stato possibile utilizzare le informazioni scaricate per i tweet storici in quanto siamo andati incontro ad un blocco da

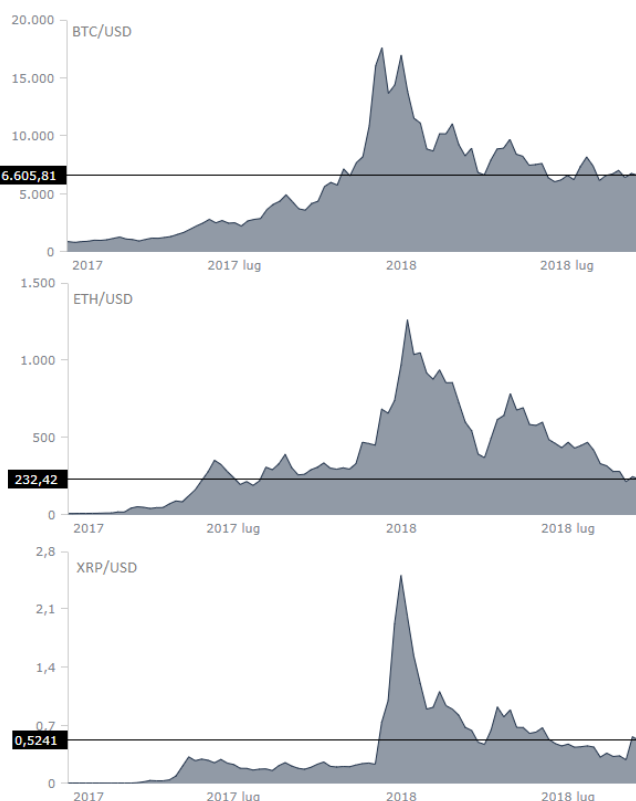


Fig. 6. andamento quotazioni BTC/USD, ETH/USD, XRP/USD.

parte del social network nei confronti dei nostri tentativi di scraping, questa per essere effettuata correttamente considerando l'enorme mole di tweet da scaricare e lo scarso numero di variabili selezionabili richiede molto tempo.

CONSIDERAZIONI TEMPI DIVERSI

3.1 EFFETTO MOMENTUM

Verifichiamo ora le principali proprietà statistiche delle serie temporali per i rendimenti di Bitcoin, Ripple ed Ethereum partendo dall'andamento dei prezzi per ogni criptovaluta [Fig.4]

La Fig.5 mostra le statistiche relative al Bitcoin, Ripple ed Ethereum considerando le osservazioni con cadenza giornaliera, settimanali e mensili confrontate con il mercato azionario [1]. Sia i rendimenti che la loro volatilità sono molto alti. Considerando la cadenza giornaliera, il rendimento medio è dello 0,52% e la deviazione standard è del 5,55%; con cadenza settimanale il rendimento medio è del 3,79% e la deviazione standard è del 16,64%; mensile, il rendimento medio è del 21,60% e la deviazione standard è del 69,46% con una percentuale che il rendimento sia maggiore di 0 del 60%. Sia le medie che le deviazioni standard sono di un ordine di grandezza superiore rispetto agli asset più tradizionali, (rappresentati in tabella ad esempio dalle azioni) come è facile supporre.

Stabiliamo quindi che ci sono prove di momentum considerando ad esempio un orizzonte giornaliero. La Fig.6 documenta i risultati della serie temporale. Per i rendimenti settimanali sul Bitcoin, l'attuale rendimento positivo e statisticamente significativo prevede la settimana 1, 2, 3 e 4. Nello specifico, il rendimento settimanale per ogni settimana precedentemente definita, è quello di comprare un

Daily	Mean	SD	T-Statistics	Sharpe	Skewness	Kurtosis	% Return >0
Bitcoin	0.52%	5.55%	4.88	0.09	0.80	15.21	53.69
Ripple	0.59%	9.11%	2.74	0.06	6.11	99.05	46.09
Ethereum	0.81%	7.71%	3.41	0.11	0.24	15.71	49.33
Stock	0.05%	0.94%	2.44	0.06	-0.51	7.95	54.91

Weekly	Mean	SD	T-Statistics	Sharpe	Skewness	Kurtosis	% Return >0
Bitcoin	3.79%	16.64%	4.49	0.23	1.76	10.25	58.72
Ripple	6.26%	47.39%	2.11	0.13	7.62	76.26	46.27
Ethereum	6.88%	24.51%	3.43	0.28	1.71	7.24	54.00
Stock	0.26%	1.93%	2.59	0.13	-0.38	5.17	59.95

Monthly	Mean	SD	T-Statistics	Sharpe	Skewness	Kurtosis	% Return >0
Bitcoin	21.60%	69.46%	2.95	0.31	4.32	25.38	60.00
Ripple	36.20%	143.31%	1.94	0.25	3.82	18.69	40.68
Ethereum	30.26%	67.56%	2.65	0.53	1.24	3.95	54.29
Stock	1.08%	3.24%	3.12	0.33	-0.10	3.78	68.89

Fig. 7. statistiche descrittive e ritorno sul BTC, ETH, XRP

Weekly	R_{t+1} (1)	R_{t+2} (2)	R_{t+3} (3)	R_{t+4} (4)
R_t	0.19*** (3.73) [2.17]	0.22*** (4.52) [2.73]	0.21*** (4.26) [2.47]	0.09* (1.72) [1.40]
R-Squared	0.03	0.05	0.05	0.01

Fig. 8. serie storica sul momentum

Bitcoin alle 11:59:59 UTD Sunday e vendere alle 11:59:59 UTD una settimana più tardi.

3.2 EFFETTO VIRALE

Per valutare il così definito "Effetto virale" abbiamo quindi confrontato gli andamenti delle ricerche effettuate tramite Google Trends con i dati sulle quotazioni di 3 principali criptovalute, ovvero il Bitcoin (Fig. 7), Ripple (Fig. 8) ed Ethereum (Fig. 9). Il periodo di riferimento preso in esame è dal 10/11/2017 al 05/02/2018, intervallo in cui abbiamo visto esserci stata un'importante variazione delle quotazioni.

Possiamo vedere come ci sia un importante aumento dell'interesse in corrispondenza del momento iniziale di ascesa del valore delle criptovalute (questo è evidente per il Bitcoin ma ancora di più per il Ripple) che denota un'attenzione particolare in questa fase, quest'ultimo sembra diminuire fortemente non appena le quotazioni vanno incontro ad una lieve flessione o comunque un arresto dell'ascesa del valore. Sembra che per il Bitcoin ci sia un nuovo importante aumento dell'interesse con la forte diminuzione verificatasi nel periodo successivo tra il 15/12

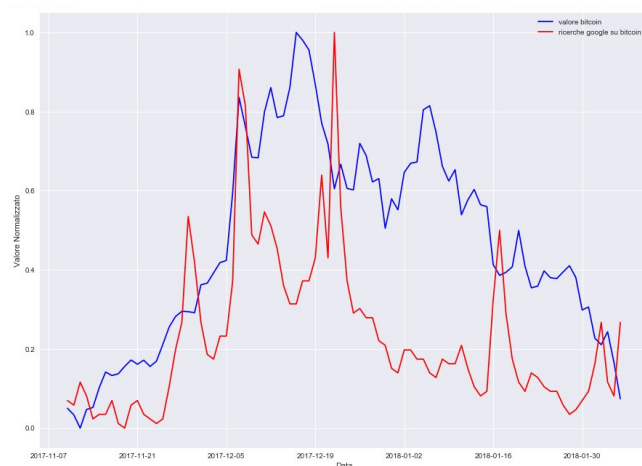


Fig. 9. confronto andamento quotazioni e ricerche per il Bitcoin.

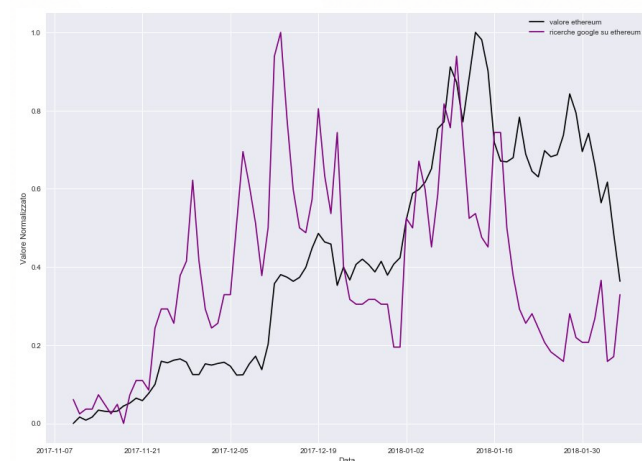


Fig. 10. confronto andamento quotazioni e ricerche per l'Ethereum.

e il 23/12 circa (oppure può essere visto come un effetto traslato di qualche giorno dell'aumento del valore immediatamente precedente a tale data), per tutte le valute abbiamo poi una generale diminuzione dell'interesse al diminuire del valore delle valute.

L'effetto spiegato sul Bitcoin è ben visibile nella Table.3 dove vediamo una forte correlazione positiva nel primo periodo ed una forte correlazione negativa nel secondo. Ci teniamo comunque a sottolineare come con la correlazione abbiamo voluto rappresentare la tendenza di una variabile a cambiare in funzione di un'altra e non un rapporto di causa-effetto.

Table 3. Confronto variabili Streaming vs Scraping

Periodo	Andamento	Corr
10/11/2017 - 16/12/2017	UP	0.83
15/12/2017 - 22/12/2017	DOWN	-0.85
22/12/2017 - 05/01/2018	UP	0.12
05/01/2018 - 05/02/2018	DOWN	-0.1

4 CONCLUSIONI E POSSIBILI EVOLUZIONI

Il lavoro effettuato lascia spazio a molte altre possibili analisi ed interpretazioni sull'andamento delle criptovalute. Avere altri eventi che possano in qualche modo essere collegati al tema delle criptovalute sicuramente potrebbe aiutare per spiegarne meglio l'andamento, ad esempio considerare altri indicatori macroeconomici. A livello di architettura le evoluzioni rispetto a quella proposta potrebbero riguardare un'implementazione tramite API di Google Trends e la possibilità di implementare soluzioni di analisi e reportistica real time rimanendo nell'ecosistema Hadoop.

5 REFERENCES

[1] Liu, Yukun and Tsyvinski, Aleh, Risks and Returns of Cryptocurrency (August 6, 2018).

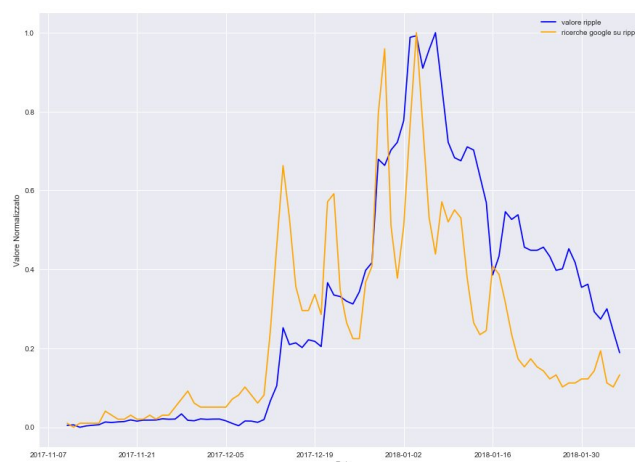


Fig. 11. confronto andamento quotazioni e ricerche per il Ripple.

THE AUTHORS
