



Università degli Studi di Cagliari

LM-91 “Data science, Business Analytics e Innovazione”



Stefano Mattu 11/82/00249

Michela Concas 11/82/00209

Giorgia Pili 11/82/00246

Massimo Loi 11/82/00315

Indice

1. Introduzione.....	3
2. Descrizione dei dataset.....	6
3. Analisi esplorativa.....	6
4. Analisi del network	
4.1 Primo quesito.....	8
4.2 Secondo quesito.....	18
4.3 Terzo quesito.....	23
4.4 Quarto quesito.....	26
5. Conclusioni e limiti del progetto.....	28
6. Bibliografia.....	29

Introduzione

Il presente progetto si pone come obiettivo l'analisi, tramite l'utilizzo della Network Analysis, di un canale Discord riguardante Algorand, una criptovaluta basata sull'utilizzo della blockchain. Abbiamo concentrato l'attenzione sul canale “Governance”, che riporta la seguente descrizione: “devoted to the discussion of onchain voting in support of new community rewards and incentives” (“dedicato alla discussione del voto in catena a sostegno di nuovi premi e incentivi per la comunità”).



Algorand è una criptovaluta basata su blockchain che mira ad essere sicura, scalabile e decentralizzata, utilizza un algoritmo di consenso detto Proof-of-Stake Puro (Pure Proof of Stake). Algorand ricorre alla casualità per risolvere il dilemma della decentralizzazione della blockchain e utilizza PPoS basato sul problema dei generali bizantini. Un consenso dei generali bizantini è un protocollo di comunicazione che consente agli utenti del sistema distribuito di raggiungere un consenso in presenza di malintenzionati. La decentralizzazione in Algorand è ottenuta selezionando a caso i convalidatori tra tutti i detentori di token. L'algoritmo PPoS sceglie automaticamente i nodi per il gruppo successivo che riceve l'idoneità necessaria ad aggiungere blocchi. Chiunque può essere selezionato dal sistema e in questo modo si mantiene la decentralizzazione della piattaforma. Ciò garantisce la sicurezza e nessuno sa chi saranno i prossimi convalidatori. L'algoritmo PPoS usa lo staking per consentire a ogni detentore di token di assumere potenzialmente il ruolo di convalidatore di blocchi.

ALGO è il token di utilità nativo della rete Algorand. La piattaforma utilizza il token per governare la crescita dell'ecosistema di Algorand e distribuire premi di governance a chi si impegna a votare nell'ambito della governance. Il totale massimo di ALGO in circolazione è 10 miliardi, l'offerta attuale in circolazione è 7.682.398.812,93 ALGO (Luglio 2023). Il prezzo attuale è 0,0942 € per ALGO (Luglio 2023), è inferiore del 96,80% rispetto al massimo storico di 2,94 € (Settembre 2021). Di seguito il grafico spiega l'andamento del prezzo per token dal 2019 a oggi (in €).



Creata nel 2017 da Silvio Micali, professore del MIT di Boston, la rete di prova Algorand è stata lanciata nell'aprile 2019, la versione operativa nel giugno 2019. Nel marzo 2020 è stato annunciato che le Isole Marshall produrranno la loro moneta digitale, chiamata *Sovereign*, utilizzando come base la blockchain Algorand. A settembre 2021 anche El Salvador ha annunciato di essere al lavoro per realizzare una propria moneta digitale su rete Algorand. Nel marzo 2021 la Società Italiana degli Autori ed Editori (SIAE) ha caricato oltre 4 milioni di non-fungible token sulla blockchain Algorand, rappresentanti il copyright dei lavori prodotti dai membri SIAE. Il 2021 segna un anno di crescita per tutto l'ecosistema di Algorand, nascono e si affermando diversi protocolli nel mondo della finanza decentralizzata (DeFi) tra cui Folks Finance, AlgoFi, Yieldly, VenueOne e xBacked. Nel Maggio del 2022 diventa sponsor ufficiale e partner tecnico della FIFA, per sviluppare la loro strategia di gestione dei loro asset digitali e portare eventualmente sul mercato soluzioni basate su blockchain. Il 13 dicembre 2022, Algorand annuncia di essere stata selezionata da Banca d'Italia per la realizzazione della prima piattaforma su blockchain per garanzie bancarie digitali in Italia.

Il grafico seguente illustra l'andamento del Market-cap dalla creazione nel 2019 a oggi (Luglio 2023) in euro €.



Attraverso l'utilizzo dei grafi l'analisi comprenderà quattro diversi obiettivi:

1. Individuare i cluster di utenti attivi nella chat sia dal punto di vista generale ma anche in una prospettiva temporale. Sono stati infatti presi in considerazione 5 semestri in quanto i dati disponibili si riferiscono ad un arco temporale compreso tra il 24/11/2020 ed il 19/04/2023.
2. Individuare i topic predominanti osservando se gli autori scrivono messaggi riguardanti le stesse tematiche e verificare se, in base agli stessi, si creano delle comunità. Realizzando dei grafi i cui ***nodi sono gli autori e gli edges sono i topic***. Anche in questo caso i messaggi sono stati analizzati sotto un punto di vista temporale, prendendo in considerazione l'intero periodo ed i semestri.
3. Analizzare le similarità tra topic discussi all'interno del canale evidenziando le keywords in comune, realizzando un grafo i cui ***nodi sono i topic e gli edges le keywords***.
4. Confrontare le attività sul canale con l'**andamento azionario** della criptovaluta, nell'arco temporale, valutando se le due forze si influenzano a vicenda o meno.

Descrizione del Dataset

Partendo da un file json, sul quale si è effettuata una pulizia preliminare, abbiamo creato 3 file csv con il fine di avere dei dati in ordine e di condurre la nostra analisi. Durante la pulizia preliminare si è verificata la presenza di valori duplicati e degli Na (ovvero valori mancanti) e, una volta individuati, questi sono stati rimossi. I documenti finali risultano essere così composti:

- **Messages.csv** abbiamo informazioni relative ai messaggi quali: “Message_id” id del messaggio postato, “Type” variabile binaria che descrive la tipologia del messaggio, “Time_stamp” dato temporale in cui il messaggio è stato postato, “isPinned” variabile binaria che indica se il messaggio è stato fissato in alto o meno nella conversazione, “Content” contenuto testuale del messaggio e “Author_id” id dell'autore del messaggio.
- **Authors.csv** abbiamo informazioni relative all'id dell'autore, il nickname dello stesso e se il soggetto è un bot o meno codificato in una variabile binaria (“Author_id”, “Nickname”, “IsBot”).
- **Mentions.csv** abbiamo informazioni relative alle menzioni: “Message_id”, “Time_stamp”, Author_id, “Author_nick”, Mention_id, “Mention_nickname”, “Mention_isBot”.

Analisi esplorativa

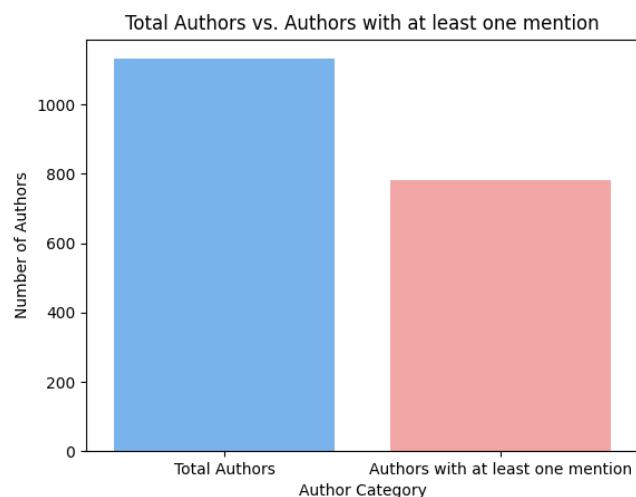
Creati i file csv, in una prima analisi esplorativa abbiamo voluto indagare le statistiche di base preliminare prima di effettuare l'analisi del network vera e propria.

Partendo da “Messages.csv”, abbiamo calcolato il numero totale di messaggi estratti dal canale Discord, pari a **13.944**; da “Authors.csv”, il numero totale di autori che attivamente hanno postato almeno un messaggio nell'arco temporale citato, pari a **1.132**, di cui un solo autore è stato identificato come bot. Dei messaggi, **4** sono stati fissati in alto, ad evidenza che quei messaggi identificano

contenuti di un'importanza maggiore rispetto agli altri all'interno delle conversazioni: gli autori di questi messaggi sono '***Fabrice | Algorand Foundation***', '***Shai Halevi***', '***Adri***'. Sono dei nomi che ritroveremo poi nell'analisi e che suggeriscono la centralità degli utenti nella rete?

I messaggi scritti da autori verificati sono il **99.97%** del totale, quelli scritti dal bot il **0.03%**, cioè 4 soli messaggi: la proporzione non è significativa affinché si possa tenere in considerazione la presenza del bot all'interno della rete e studiare così la possibile influenza sugli altri utenti. Si è per cui deciso di eliminare i messaggi postati dal bot.

Da "Mentions.csv" abbiamo scelto di analizzare quanti degli autori totali sono stati menzionati almeno una volta da altri utenti all'interno delle conversazioni: di questi il **68.99%** del totale, pari a **781** users, sono stati menzionati da altri.



Primo quesito

Il primo quesito che ci siamo posti è quello di verificare se all'interno della chat si creano dei cluster, ovvero dei gruppi di soggetti, i quali si menzionano tra di loro escludendo gli altri, partendo dalla realizzazione di un dizionario in cui si sono raccolti i nickname dell'autore, del soggetto menzionato e il rispettivo numero delle menzioni. Al fine di raggiungere il nostro obiettivo, è stato realizzato un grafo con il **Kamada Kawai Layout** i cui nodi sono gli autori che vengono collegati ai soggetti menzionati, in base ad un peso che fa riferimento al numero delle menzioni. Questo layout si basa sull'algoritmo di Kamada e Kawai, dove i nodi che occupano una posizione centrale sono quelli direttamente collegati con la maggior parte degli altri nodi. I periferici, invece, sono quei nodi che presentano un'interconnessione minore, per cui la distanza media con tutti gli altri, considerata la lunghezza dei percorsi minimi, è maggiore. In particolare, la distanza viene calcolata utilizzando la *shortest_path_length*. Di conseguenza, i nodi centrali sono caratterizzati da un grado elevato rispetto ai nodi più periferici che presentano invece un minor numero di laterali adiacenti.

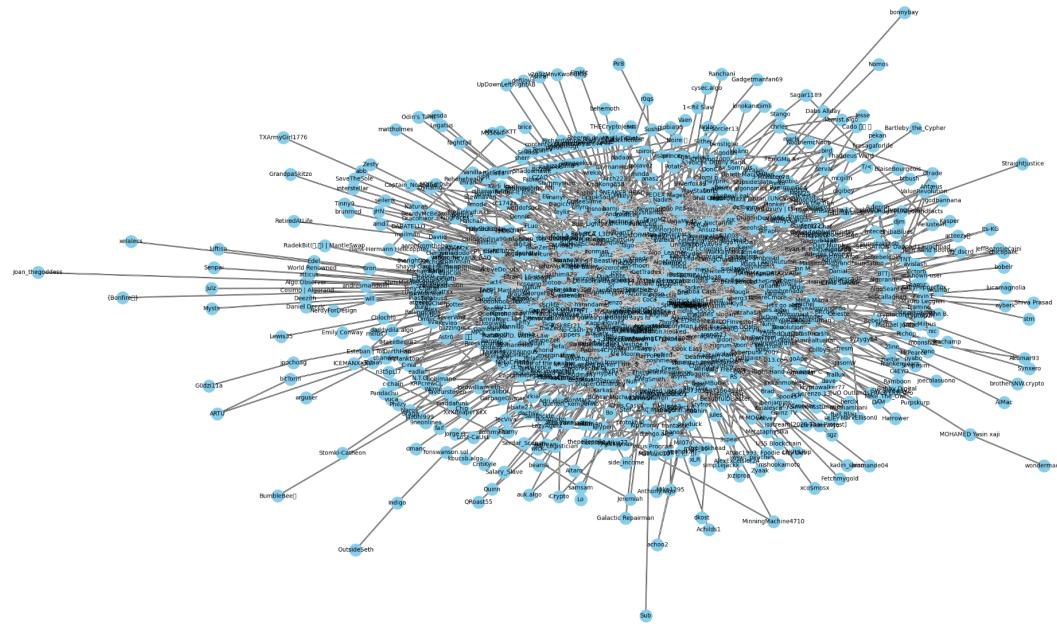
Prima di realizzare il primo grafo, abbiamo ritenuto importante verificare quali sono gli utenti maggiormente menzionati e risultano essere i seguenti:

1. nullun viene menzionato 239 volte
2. Mf3jgh7k viene menzionato 239 volte
3. Deleted User viene menzionato 194 volte
4. Jeremy.algo viene menzionato 176 volte
5. averagezen viene menzionato 169 volte
6. amalgam viene menzionato 126 volte
7. patrick.algo | nerd viene menzionato 108 volte
8. ipaleka viene menzionato 108 volte

Da questa prima stampa, non sembra che i soggetti **Fabrice | Algorand Foundation, Shai Halevi, Adri** ricoprono un ruolo di centralità nella rete, ma questo è stato ulteriormente indagato nei grafi realizzati. Possiamo però affermare, dopo esserci iscritti personalmente alla chat, che questi utenti

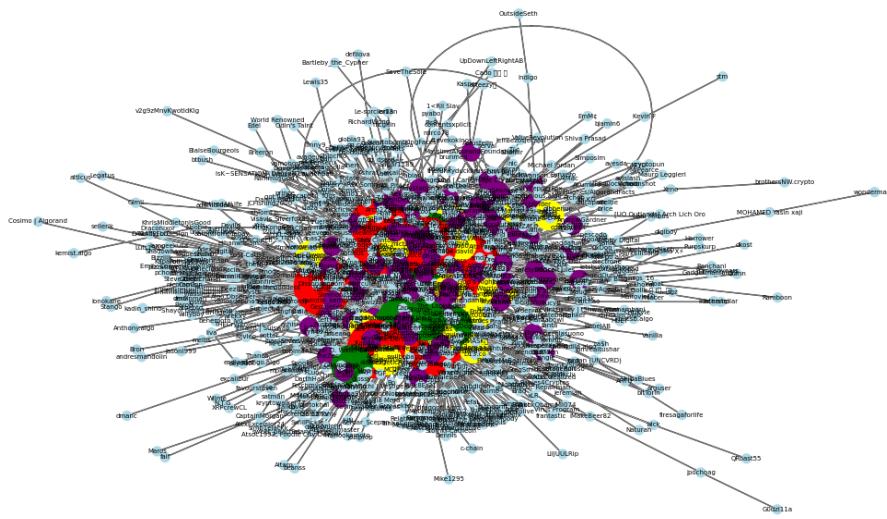
ricoprono un ruolo di membri dell'Algorand Foundation. L'Algorand Foundation è un'organizzazione senza scopo di lucro focalizzata sulla governance del protocollo, sulle dinamiche dei token e sul supporto dello sviluppo di base dell'ecosistema Algorand.

Il risultato ottenuto è il seguente:

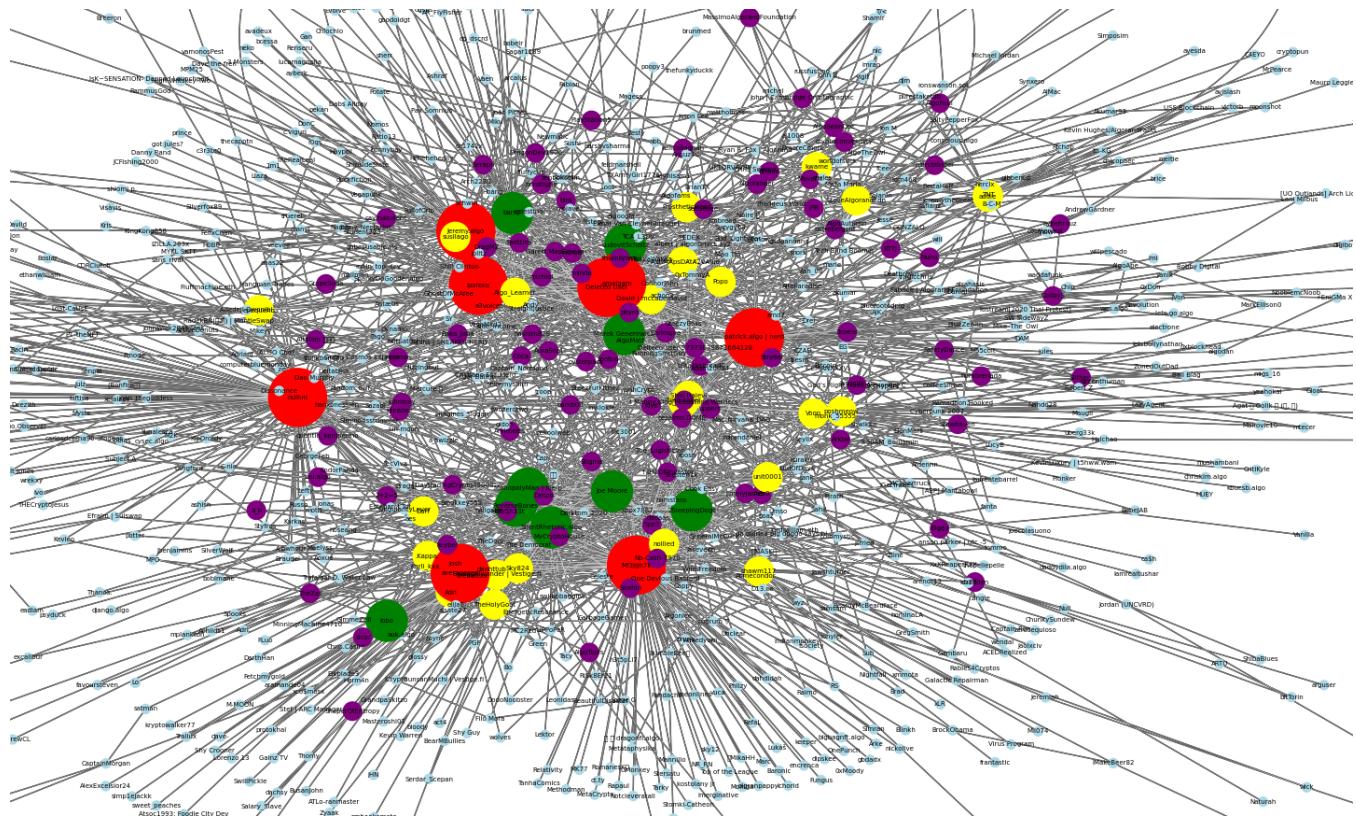


Da questo primo grafo è difficile avere la certezza che i soggetti più menzionati assumono una posizione centrale, dunque per approfondire l'analisi sulle menzioni, sono stati modificati i colori dei nodi in base al numero delle citazioni.

È stato utilizzato il rosso per gli utenti che ricevono più di 100 menzioni, verde per gli utenti che ne ricevono più di 50, giallo per quelli menzionati più di 20 volte, viola per quelli menzionati più di 5 volte e infine celeste per gli utenti menzionati più di una volta.



In questo grafo abbiamo una panoramica generale dei nodi e possiamo già notare come i soggetti meno menzionati si collocano all'estremità del grafo, man mano che il numero delle menzioni aumenta di conseguenza vi è un avvicinamento dei nodi al centro e l'incremento dell'ampiezza degli stessi.



Eseguendo uno zoom notiamo che dai soggetti maggiormente menzionati (rossi) partono numerosi edges, in quanto, come precedentemente affermato, i soggetti vengono collegati in base al numero

delle menzioni. A sostegno di quanto affermato prima, i nodi rossi sono riferiti agli utenti identificati precedentemente come maggiormente menzionati. Per quanto riguarda i membri dell'Algorand Foundation, i loro nodi non sono visibili, quindi pur ricoprendo un ruolo importante all'interno del canale Discord, non vengono citati dagli utenti nei loro messaggi.

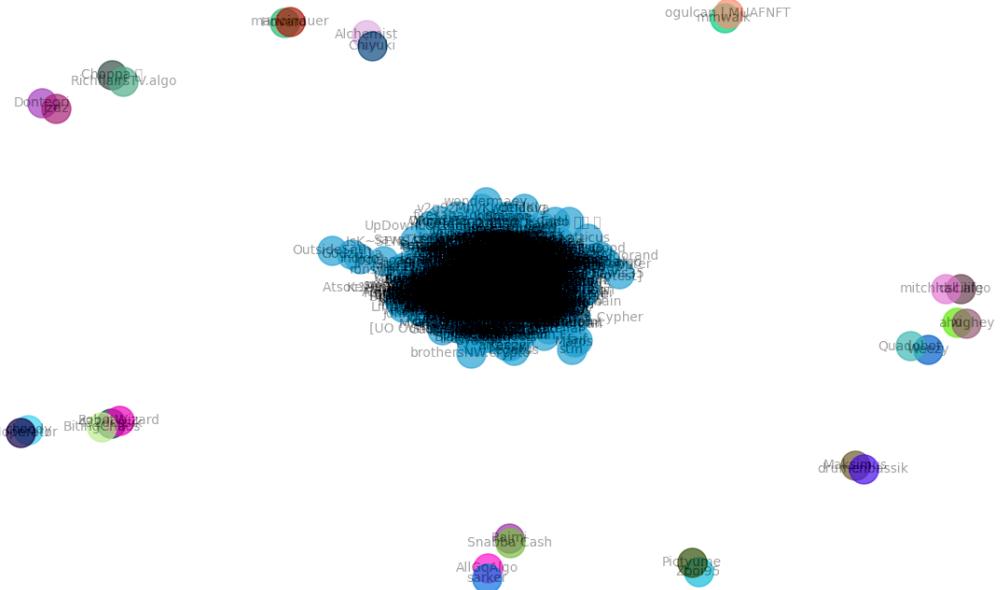
Il risultato ottenuto non ci consente ancora di capire quali sono i cluster che effettivamente si formano, per questo abbiamo deciso di utilizzare il **Girvan-Newman algorithm**, uno degli algoritmi di *hierarchical community detection*. Tale algoritmo rileva ed esegue un'analisi sulla struttura della comunità, inoltre elimina progressivamente i lati che presentano il più alto livello di *betweenness*.

La betweenness, essendo una misura di centralità, coglie la centralità di un nodo nello scambio di comunicazioni, ovvero misura il numero di volte in cui un nodo si trova nel percorso più breve tra 2 nodi, definendone così l'influenza nel flusso delle informazioni.

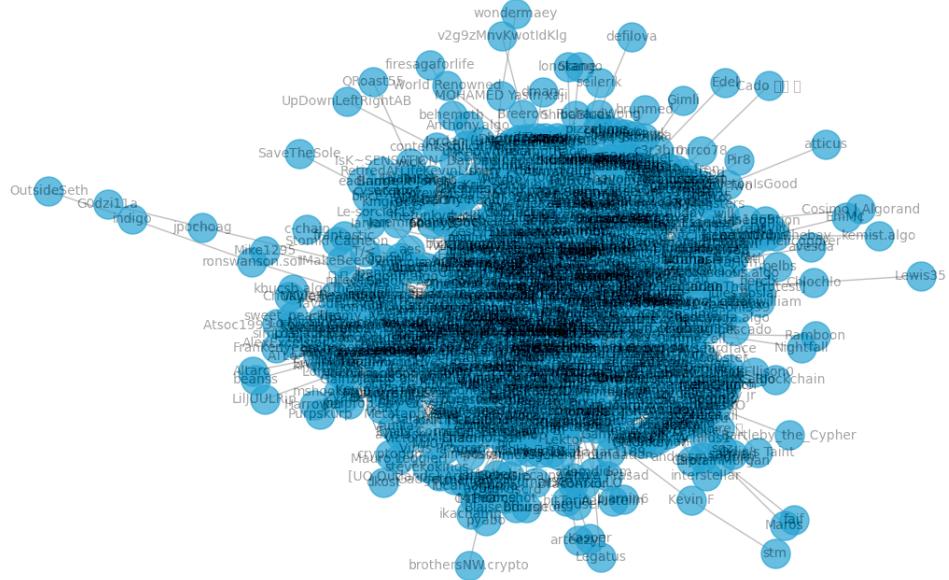
Abbiamo calcolato il nodo con il valore massimo di betweenness, il quale ha un risultato pari a **0.22** e si riferisce al soggetto **Mf3jgh7k**, il che sta ad indicare che l'utente si trova nel percorso più breve circa il 22% delle volte nel grafo. Considerando il numero di nodi presenti nei nostri grafi possiamo affermare che questo utente ha una certa importanza dell'interconnessione tra i nodi e svolge un ruolo importante nel viaggio delle informazioni. L'algoritmo si articola in quattro passaggi principali:

1. Calcola la centralità per ogni edge;
2. Rimuove gli edges con la più alta betweenness.
3. Calcola la betweenness per ogni edge rimanente.
4. Ripete i passaggi 2 - 3 fino a quando non ci sono più archi.

L'algoritmo **Girvan-Newman** ha rilevato la presenza di 31 comunità e utilizzando **Spring Layout** è emersa una comunità di utenti molto ampia posizionata al centro mentre le altre cluster sono composte solo da uno utente.



Zoom sul corpo centrale:



Eseguendo uno zoom, per focalizzare l'attenzione su questa ampia comunità, la maggior parte dei nodi appartengono alla stessa, quindi non soddisfatti del risultato, abbiamo deciso di utilizzare anche il ***Greedy Modularity Algorithm***.

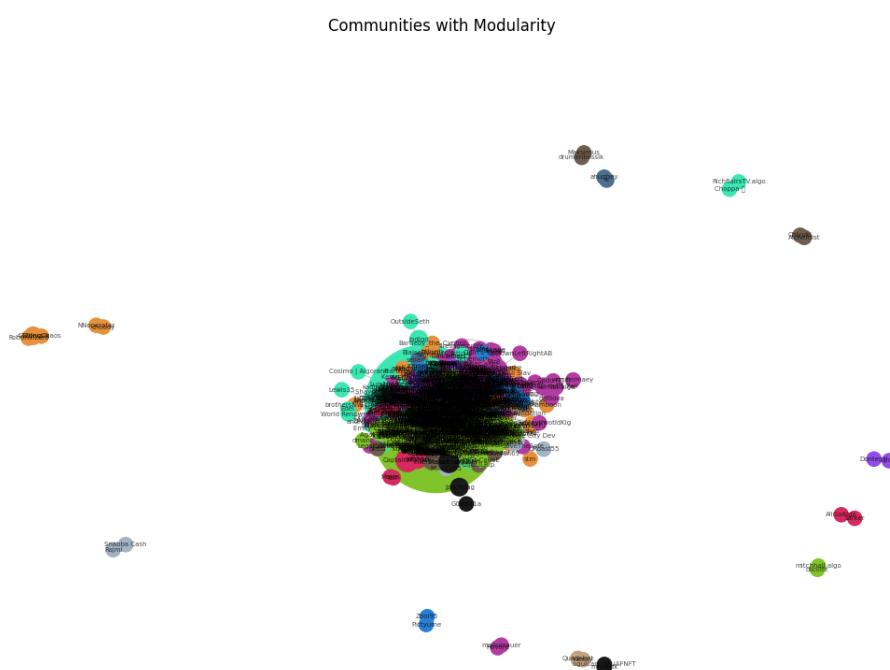
Questo algoritmo utilizza la massimizzazione della modularità per misurare le qualità dei partizionamenti comparando la frazione di archi all'interno della comunità con la frazione di archi che si avrebbe se venissero effettuate connessioni casuali tra i nodi, quantificando così la forza della comunità.

L'algoritmo si articola in quattro passaggi principali:

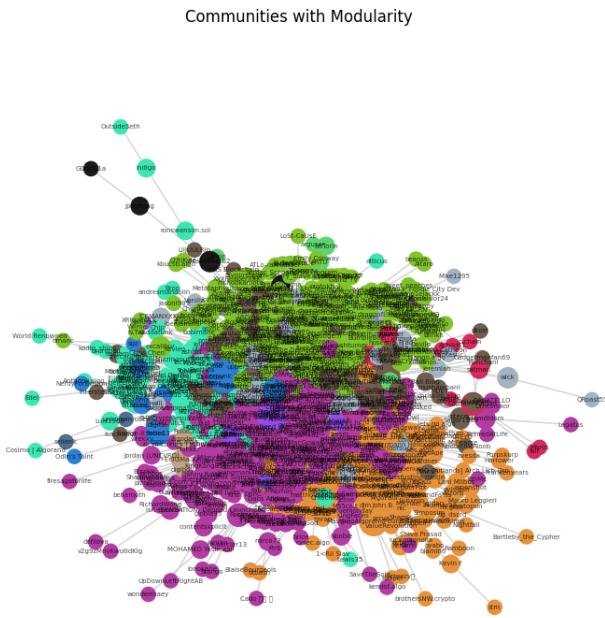
1. All'inizio, ogni nodo appartiene a una comunità diversa;
 2. La coppia di nodi/comunità che, uniti, aumentano di più la modularità, diventa parte della stessa comunità, la modularità viene calcolata per l'intera rete;
 3. Il passaggio 2 viene eseguito fino a quando non rimane una sola comunità;
 4. Viene scelta la partizione di rete con la maggiore modularità.

Abbiamo calcolato la modularità, la quale può assumere valori compresi tra 0 e 1, e si è ottenuto un risultato pari a **0.52**, valore che si trova in una situazione intermedia tra lo 0, che indica una frazione di nodi all'interno della comunità non migliore rispetto all'ipotesi casuale, e un valore pari a 1 che indica una struttura della comunità che ha la massima forza possibile.

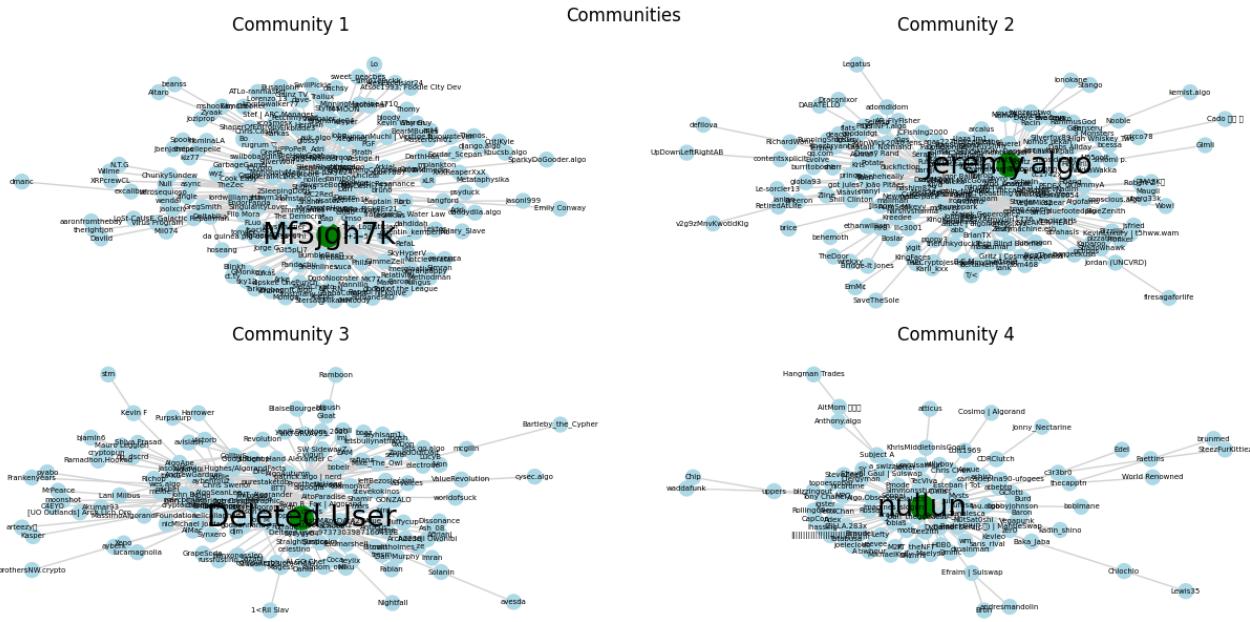
Il grafo ottenuto con questo algoritmo risulta essere il seguente:



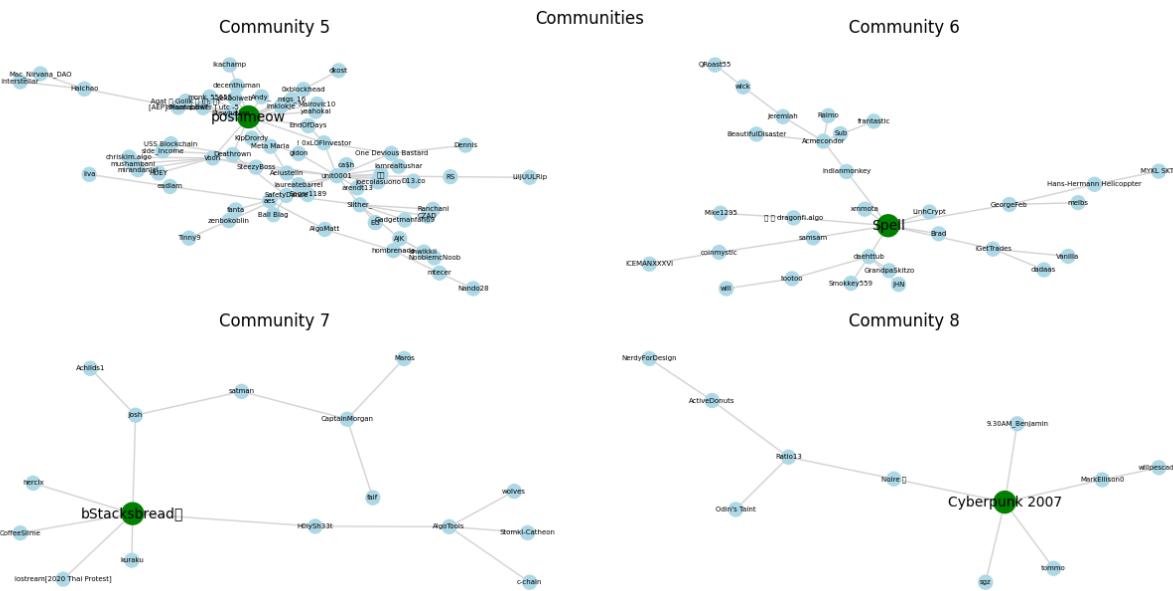
In questo caso non abbiamo un'unica comunità centrale ma cluster differenti che identificano soggetti che comunicano prevalentemente con gli stessi utenti.



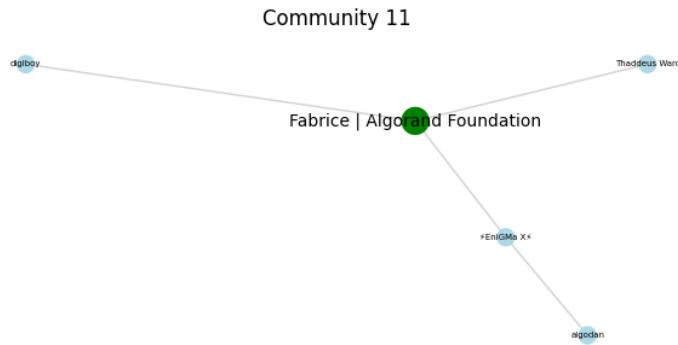
Avendo eseguito uno zoom abbiamo un'idea più chiara del fatto che la nuvola centrale presenta numerose comunità la cui dimensione varia. Anche in questo caso sono emerse 31 cluster e utilizzando una funzione abbiamo deciso di realizzare dei grafi, sempre attraverso l'utilizzo dello ***spring_layout***, che mostrino i quattro cluster più ampi e mettano in risalto il nodo più importante all'interno di ciascuno. L'importanza di ciascun nodo è stata determinata attraverso la funzione ***pagerank***, ovvero è stata definita in base alla struttura dei link in entrata.



Dalla visualizzazione delle comunità più grandi, abbiamo notato che i soggetti più importanti all'interno coincidono con i soggetti che ricevono più menzioni.

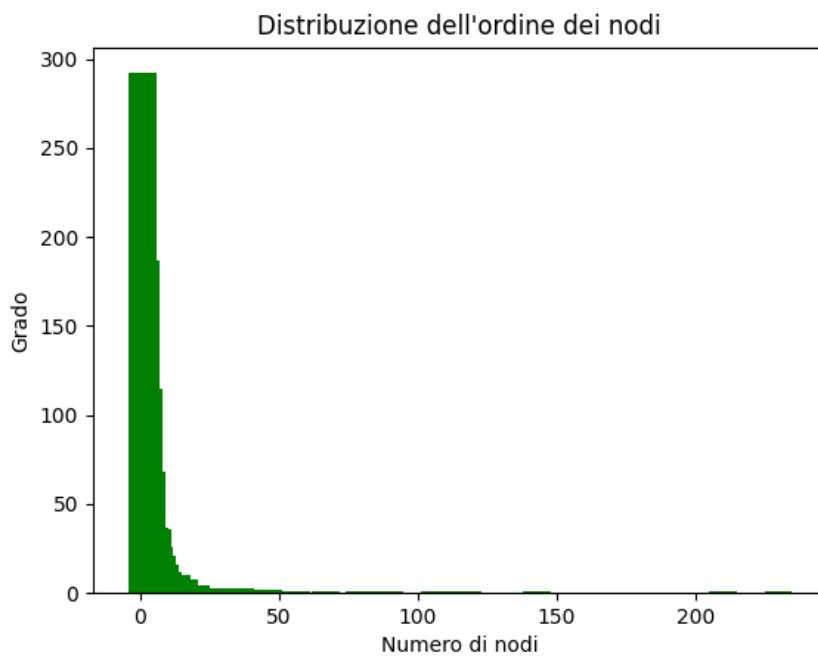


Analizzando le successive 4 comunità notiamo che la numerosità all'interno diminuisce, quindi nella chat vi sono soggetti che scambiano messaggi con pochi utenti. Di queste nessun autore rilevante risulta essere tra i soggetti più menzionati.



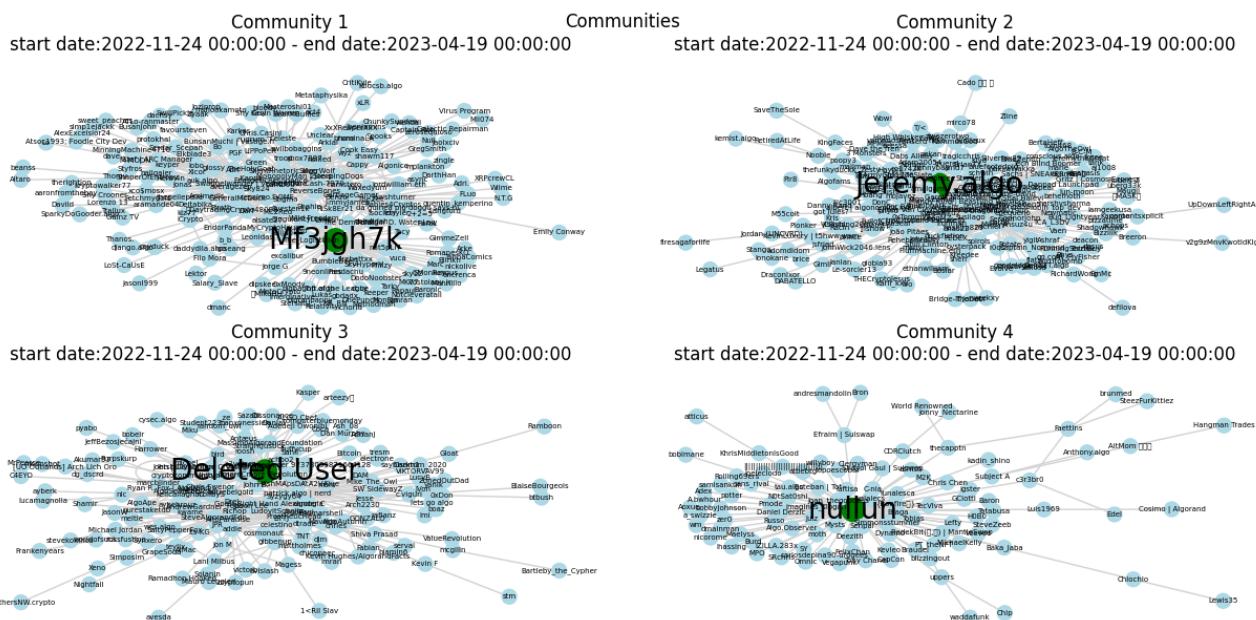
Ponendo l'attenzione sulla comunità numero 11, abbiamo notato che il nodo più importante è rappresentato da ***Fabrice | Algorand Foundation***, il membro della foundation crea una comunità con altri quattro soggetti: ‘agodan’, ‘>EniGMA X<’, ‘Thaddeus Word’, ‘dglboy’.

Nella Cluster generica abbiamo infine calcolato la distribuzione dal grado dei nodi del grafo G, utilizzando la libreria NetworkX e stampato un diagramma a barre.



Dal grafo possiamo vedere che solo circa 20 nodi hanno un grado elevato, che raggiunge un valore prossimo a 300. La restante parte dei nodi presenta un grado molto basso.

Terminata l'analisi generale siamo passati ad un'analisi temporale, con l'intenzione di capire se la formazione dei cluster fosse diversa in base al periodo di riferimento e se gli utenti più menzionati subissero delle modifiche in base al periodo temporale. Utilizzando anche stavolta l'algoritmo di **Greedy Modularity**, abbiamo sia osservato le comunità a livello generale per ogni semestre, sia riposto l'attenzione sulle 4 community più popolate. Notiamo che nel tempo la composizione delle community non presenta grosse differenze, infatti, come si può vedere dal grafico sotto, riferito all'ultimo semestre, i 4 grafi hanno gli stessi nodi centrali, ovvero gli utenti più menzionati: “Mf3jgh7k”, “jeremy.algo”, “Deleted User” e “nullun” che sono presenti anche nel grafo riferito al primo semestre visibile nel codice.



Da questa analisi possiamo affermare che, nel corso degli anni, i soggetti mantengono le stesse abitudini, comunicando prevalentemente sempre con gli stessi soggetti e che, all'interno di ciascuna comunità, i punti di riferimento (i nodi più importanti) sono sempre gli stessi, per questo la composizione dei cluster non varia.

Secondo quesito

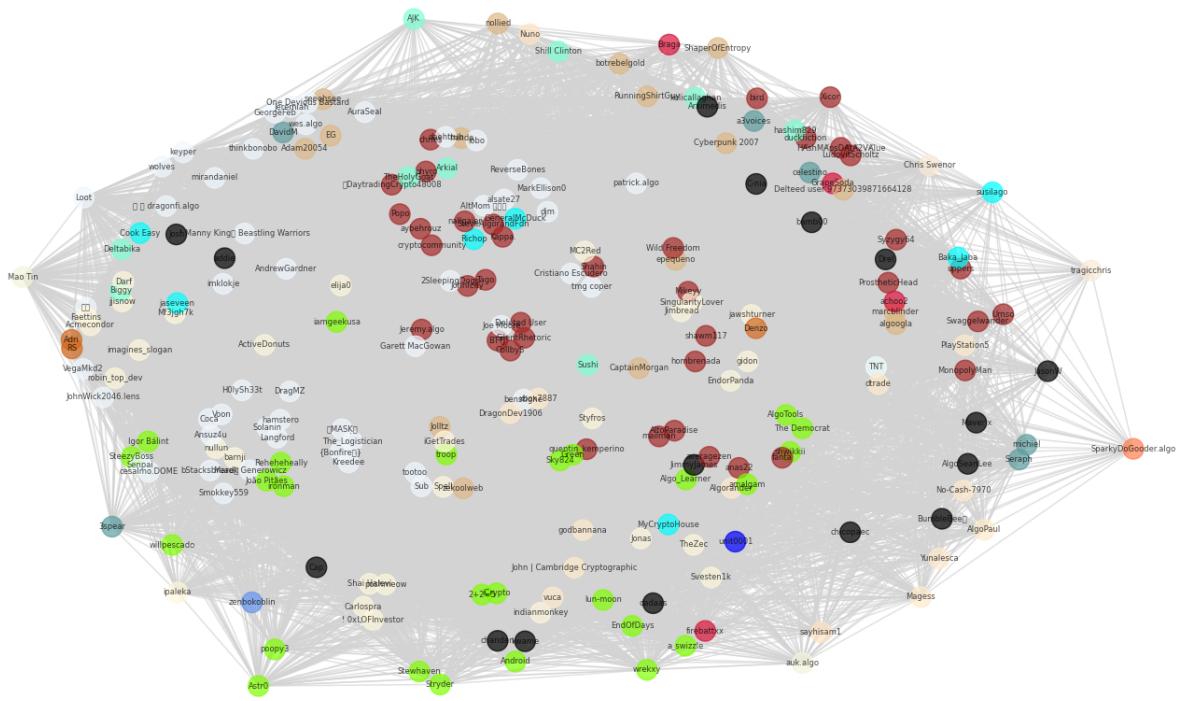
Nel secondo quesito ci siamo posti l'obiettivo di analizzare i messaggi che gli utenti si scambiano all'interno della chat, cercando di estrarre il topic principale, per capire quali sono gli argomenti maggiormente discussi da ciascun autore. Il metodo utilizzato è quello della **topic analysis**, ovvero una tecnica di *Natural Language Processing (NLP)* che ci permette di estrarre dal testo i token più ricorrenti e quindi di individuare il tema principale di ogni messaggio.

Abbiamo applicato una funzione per il preprocessing al testo dei messaggi, al fine di renderli adatti all'analisi. Le attività svolte dal preprocessing sono: tokenizzazione dei documenti, rimozione della punteggiatura, delle maiuscole, delle stopwords, delle parole con meno di due caratteri e infine lemmatizzazione, dunque le parole sono state riportate al loro lemma.

Una volta caricato il dataset, abbiamo creato un dizionario in cui le chiavi corrispondono agli autori e i valori a tutto il corpus dei messaggi scritto da ciascun autore e applicato la funzione di preprocessing ad ogni corpus. Abbiamo deciso di utilizzare i dati riferiti unicamente agli autori che nel periodo in esame avevano scritto almeno dieci messaggi, in modo da eliminare tutti quelli che erano ininfluenti ai fini dell'analisi.

Una volta applicata la topic analysis ai messaggi di ciascun autore, abbiamo creato una rete di utenti basata sulla similarità dei loro argomenti, in cui gli utenti sono i nodi di un grafo e ogni autore viene collegato ad un altro se, tra gli argomenti di cui scrivono nella chat, ce n'è almeno uno in comune.

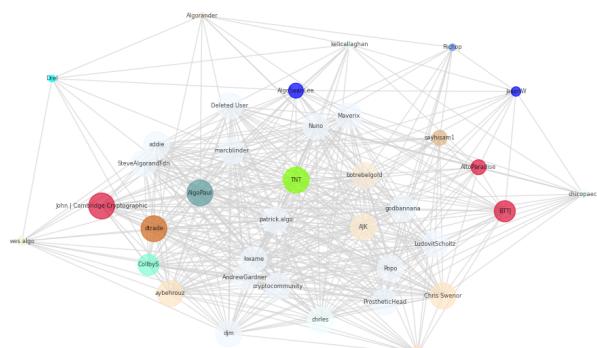
Prendendo come riferimento l'intero arco temporale, abbiamo ottenuto il grafo seguente utilizzando anche questa volta come layout il **Kamada Kawai Layout**.



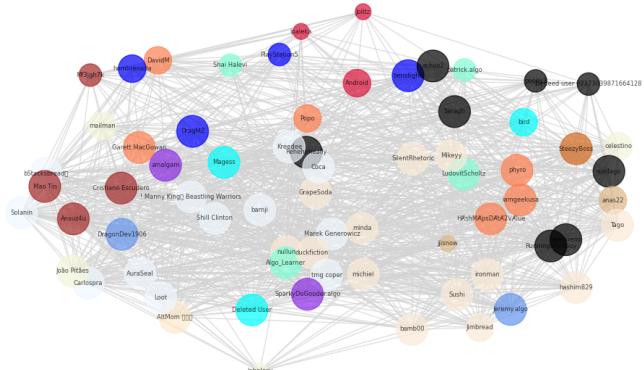
Da questo grafo si può notare che il grado medio è molto alto, ogni nodo infatti è collegato in media a oltre la metà degli altri nodi presenti nel grafo. Ciò significa che gli utenti all'interno della chat trattano più o meno gli stessi topic, e non ci sono gruppi isolati di autori che parlano di argomenti diversi rispetto agli altri utenti. Inoltre, abbiamo dato diverse colorazioni ai nodi in modo da evidenziare gli utenti che hanno lo stesso main topic, ossia quelli che scrivono principalmente messaggi riguardo allo stesso argomento.

Focalizzando l'attenzione sulla variazione temporale suddivisa in semestri abbiamo messo in risalto l'importanza dei nodi, variandone la dimensione, in base al numero di collegamenti (ovvero il grado). Attraverso lo strumento della *Word Cloud*, che evidenzia le parole ricorrenti in ogni semestre all'interno della chat, i risultati ottenuti per ciascun semestre sono i seguenti:

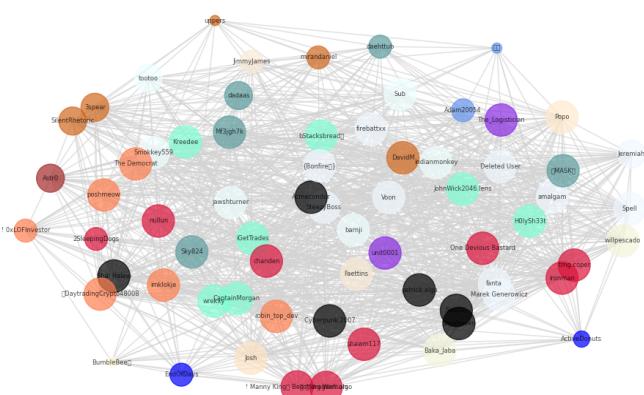
- #### ● Primo semestre



- Secondo semestre

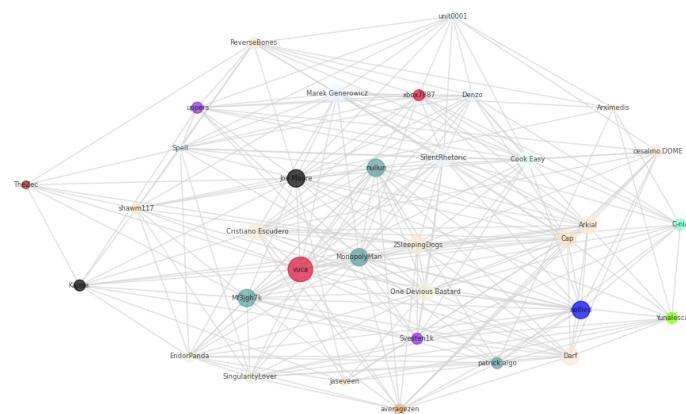


- ### ● Terzo semestre

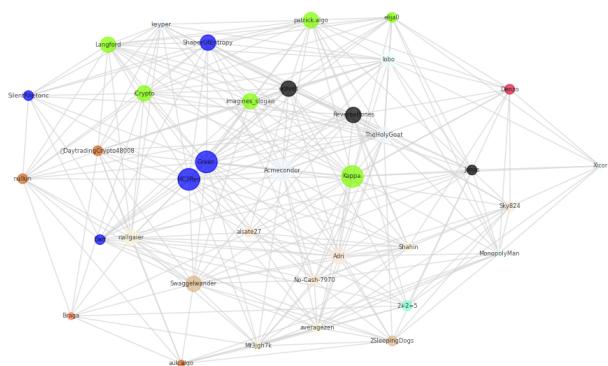


https://www.algorand.com/governance/vote/reward/period

- Quarto semestre



- Quinto semestre



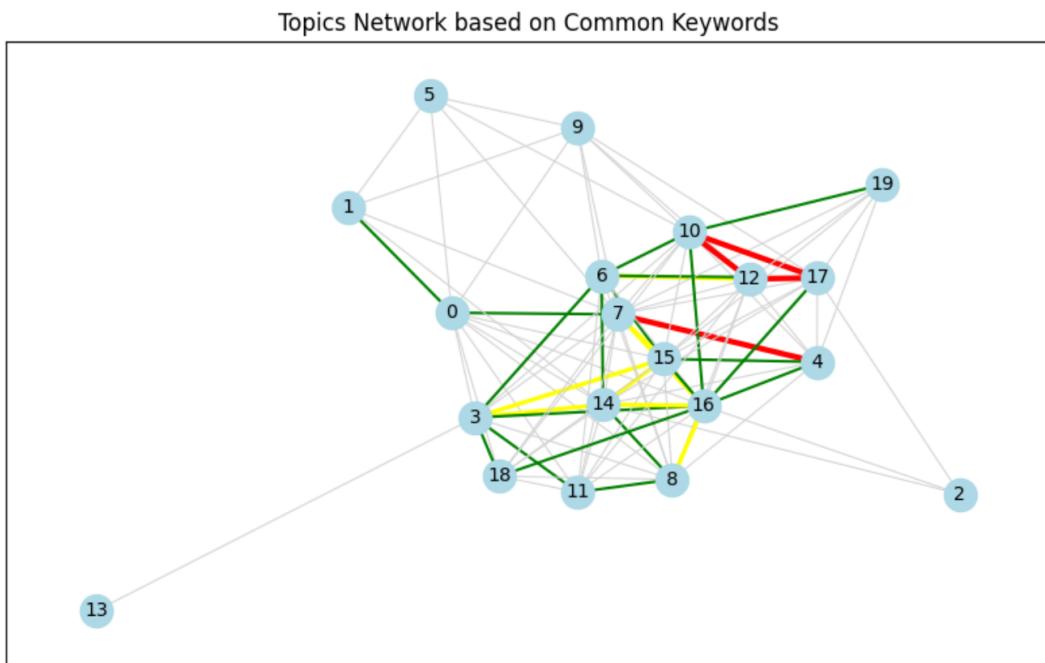
Possiamo notare come nell'arco dei semestri le parole più utilizzate sono abbastanza simili e i cambiamenti non sono molto significativi, come evidenziato dal Word Cloud, tra le keywords principali troviamo: Governance, Algo, reward, Algorand e wallet. Nel tempo varia principalmente il numero di nodi, che aumenta soprattutto nel secondo e nel terzo semestre, per poi diminuire nuovamente nei due semestri successivi, il che probabilmente è dovuto al fatto che nel 2021 il valore della moneta è aumentato e con esso il market cap di Algorand e questo potrebbe aver generato discussioni più vivaci all'interno della chat.

Durante ciascun semestre abbiamo calcolato diverse misure, due di centralità (eigenvector centrality e grado), transitivity e density:

- *Eigenvector centrality*, misura per ogni nodo la centralità, ossia l'importanza dello stesso, ma a differenza del grado, che indica il totale delle connessioni, questa considera anche **chi sono** i tuoi collegamenti e non solo il numero. Per tutti i semestri la Eigenvector centrality media assume un valore tra lo 0.1 e lo 0.2 e non rilevano nodi che presentino un valore molto più elevato, ciò indica che l'importanza dei nodi è relativamente uniforme e non spiccano autori più centrali di altri da questo punto di vista;
- *Transitivity*, rapporta il numero di triadi chiuse al numero di sentieri di lunghezza due, maggiore è il suo valore tanto più i nodi del grafo tendono a creare connessioni tra di loro. Per tutti i semestri si registrano valori più alti di 0.5, quindi tutti i grafi hanno una transitività relativamente alta;
- *Density*, indica il rapporto tra numero di link nel grafo e numero massimo di collegamenti possibili. Anche la density assume valori relativamente elevati in tutti semestri, questo indica che il numero di archi effettivi, ossia di connessioni, è alto rispetto al numero totale di connessioni possibili.

Terzo quesito

Nel terzo quesito abbiamo riposto l'attenzione sui topic, con l'obiettivo di trovare delle similarità, impostando la fase di preprocessing come nel punto precedente, usando quindi la LDA. Per favorire la visualizzazione del grafo, abbiamo impostato un numero massimo di topic pari a 20 e per ciascuno abbiamo estratto le 10 keywords più importanti. Usando quindi lo **Spring Layout** abbiamo stampato il grafo i cui nodi sono i topic e gli archi sono dati dalla comunanza di keywords tra topic mettendo in evidenza il peso degli edges secondo il numero di elementi in comune tra nodi. Il colore degli edge è stato assegnato in base al numero di keywords comuni, rosso maggiori di 3, giallo uguali a 3, verde uguali a 2 e grigio pari ad uno.



Osservando il grafo vediamo che i collegamenti tra i topic sono numerosi e ciò significa che condividono diverse keywords tranne pochi casi, ad esempio, il nodo corrispondente al Topic 13 risulta essere isolato rispetto a tutti gli altri, è infatti collegato solo al nodo 3 in quanto hanno in comune solo una keyword.

Per rendere più chiaro il contenuto di ciascun Topic, abbiamo qui riportato le keywords utilizzate per la costruzione del grafo.

TOPIC 0: ['time', 'finance', 'really', 'governance', 'like', 'fine', 'anything', 'bring', 'work', 'way'],

TOPIC 1: ['yeah', 'sure', 'hey', 'better', 'work', 'dao', 'like', 'browser', 'folk', 'stuff'],

TOPIC 2: ['thanks', 'connect', 'wait', 'commitment', 'window', 'appreciate', 'worked', 'open', 'update', 'wallet'],

TOPIC 3: ['algorand', 'http', 'forum', 'governance', 'foundation', 'please', 'community', 'org', 'voting', 'doc'],

TOPIC 4: ['reward', 'period', 'apr', 'balance', 'committed', 'asset', 'committee', 'vote', 'amount', 'eligible'],

TOPIC 5: ['platform', 'good', 'yet', 'bit', 'know', 'even', 'like', 'feel', 'given', 'question'],

TOPIC 6: ['would', 'network', 'algo', 'algorand', 'protocol', 'round', 'foundation', 'think', 'come', 'gon'],

TOPIC 7: ['algos', 'governor', 'period', 'like', 'account', 'committed', 'eligible', 'governance', '000', 'vote'],

TOPIC 8: ['lol', 'think', 'twitter', 'status', 'site', 'governance', 'get', 'something', 'reward', 'okay'],

TOPIC 9: ['one', 'contract', 'criterion', 'like', 'would', 'idea', 'smart', 'around', 'farm', 'miss'],

TOPIC 10: ['node', 'people', 'voting', 'know', 'vote', 'would', 'power', 'make', 'connected', 'algo'],

TOPIC 11: ['got', 'community', 'term', 'governance', 'mind', 'previous', 'imo', 'think', 'confused', 'point'],

TOPIC 12: ['defi', 'algo', 'people', 'apy', 'think', 'risk', 'would', 'vote', 'went', 'looking'],

TOPIC 13: ['http', 'com', 'discord', 'www', 'mean', 'channel', 'token', 'value', '491256308461207573', 'showing'],

TOPIC 14: ['pera', 'wallet', 'pool', 'governance', 'algorand', 'using', 'reward', 'foundation', 'folk', 'vault'],

TOPIC 15: ['yes', 'governance', 'thank', 'gov', 'period', 'vote', 'foundation', 'next', 'algorand', 'page'],

TOPIC 16: ['governance', 'reward', 'algo', 'period', 'wallet', 'commit', 'get', 'voting', 'algos', 'end'],

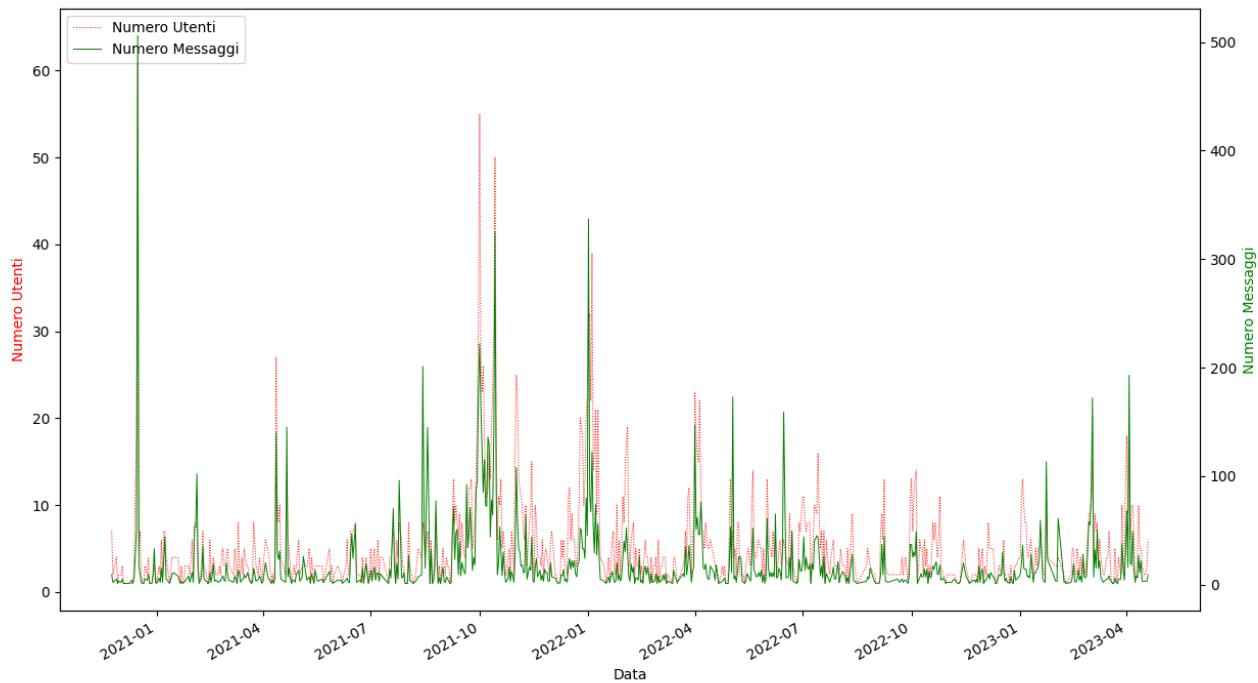
TOPIC 17: ['wallet', 'vote', 'need', 'option', 'would', 'dont', 'already', 'people', 'algo', 'committed'],

TOPIC 18: ['transaction', 'session', 'governance', 'voted', 'send', 'voting', 'sign', 'message', 'proposal', 'address'],

TOPIC 19: ['ledger', 'try', 'problem', 'arc', 'tried', 'support', 'make', 'vote', 'issue', 'nice']

Quarto quesito

In questo ultimo punto vogliamo porre l'attenzione su come l'andamento azionario ha inciso all'interno nella chat Discord che abbiamo analizzato. Un importante spunto di analisi è stato quello di monitorare come il numero dei messaggi e il numero di autori attivi è variato nel corso degli anni e se abbia seguito lo stesso andamento del mercato azionario e dell'andamento dei prezzi di ALGO.



Nel seguente grafico infatti, abbiamo creato una serie temporale per individuare possibili picchi e possiamo notare che il numero dei messaggi e il numero degli autori è abbastanza variabile, il 16 Dicembre 2020 raggiunge un livello massimo pari a 500 messaggi, ma lo spunto interessante è che a novembre 2021 e a gennaio 2022, raggiunge altri importanti picchi che si aggirano intorno ai 350 messaggi. Anche per quanto riguarda gli utenti nella chat, il picco più alto si è verificato il 16 Dicembre 2020 con 64 utenti attivi, ma possiamo notare due ulteriori picchi, il 1 Ottobre 2021 con 55 partecipanti e il 14 Ottobre con 50 utenti attivi. Osservando il file csv abbiamo riscontrato il motivo per il quale si è verificato un punto massimo il 16 Dicembre 2020, in tale data è stata effettuata una live chat nel canale discord per discutere sulla decentralizzazione della Governance di Algorand.



Dal sito coinmarketcap.com abbiamo estrapolato il resoconto dell'andamento del prezzo dal 15 Settembre 2021 al 31 Ottobre 2021 e possiamo notare che in corrispondenza del 1 Ottobre 2021 e 14 Ottobre, date in cui abbiamo registrato un incremento nel numero degli utenti, l'andamento del prezzo prima cresce (1 Ottobre), mentre nella seconda data raggiunge uno dei picchi più bassi. Da questo possiamo affermare che questi eventi hanno suscitato un interesse tra gli utenti spingendoli ad essere maggiormente attivi nella chat.

Conclusioni e limiti

In conclusione, con riferimento al primo punto si nota che gli utenti, all'interno del canale, tendono a creare dei gruppi e a citare nei loro messaggi sempre gli stessi soggetti. Le abitudini sociali sono rimaste coerenti dal periodo iniziale di analisi fino alla fine. Ci sono 4 soggetti, all'interno delle 4 comunità più grandi, che ricoprono un ruolo centrale rispetto agli altri, a monte di ciò non possiamo trarre conclusioni sull'effettiva centralità degli utenti con riferimento al funzionamento di Algorand a livello di performance sul mercato. Con riferimento al secondo e terzo quesito, è emerso che i topic affrontati dagli utenti sono molto simili e costanti nel tempo, questo è dovuto al fatto che il canale è monotematico, non ci aspettavamo infatti temi che potessero discostarsi dalla Governance. Oltre alla caratteristica del canale, non possiamo trascurare che la bontà dell'analisi dipende anche dal metodo da noi utilizzato per pre-processare i contenuti testuali, infatti abbiamo notato che alcune parole, non discriminanti nell'influenzare i topic e che potrebbero essere considerate stopwords, non sono state rimosse, perché non facenti parte effettivamente delle stopwords di nltk. In relazione al quarto quesito siamo arrivati alla conclusione che vi sono delle associazioni tra il numero di autori e messaggi postati nel canale nei periodi temporali presi in analisi e l'andamento di Algorand sul mercato, per quanto non sappiamo con certezza se l'attività su Discord possa causare delle variazioni dell'andamento sul mercato o più probabilmente viceversa sono gli eventi esterni relativi al mondo Algorand a rendere la chat più attiva.

In generale, il caso ha presentato dei limiti perché non avevamo accesso ad una profilazione più dettagliata sugli utenti e questo non ci ha consentito di basare l'analisi su parametri differenti riguardanti gli autori che andasse oltre le menzioni o i topic di discussione.

Bibliografia:

<https://algorand.com/>

<https://networkx.org/documentation/stable/index.html>

<https://memgraph.com/blog/community-detection-algorithms-with-python-networkx>

<https://www.algorand.foundation/general-faq#governance-faq-header>

<https://coinmarketcap.com/it/currencies/algorand/>