UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dipartimento di Scienze Fisiche, Informatiche e Matematiche
Corso di Laurea in Informatica

Tesi di Laurea Magistrale

Ricerca degli Utenti più Influenti di Twitter in Periodo Pandemico ed Analisi del Loro Contributo Utilizzando Tecniche di Text e Graph Analytics

RELATORE
Prof. Riccardo Martoglia
CORRELATORI
Prof. Marco Furini
Prof.ssa Manuela Montangero

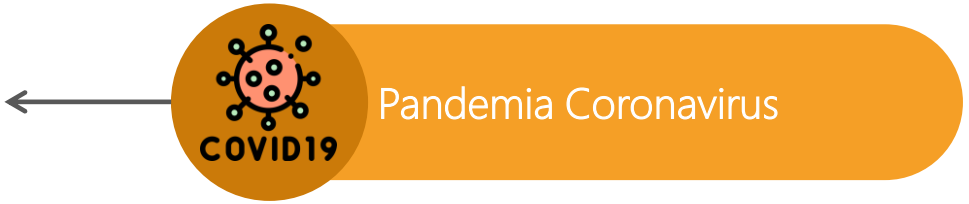
LAUREANDO
Luca Mariotti

Anno accademico 2020/2021

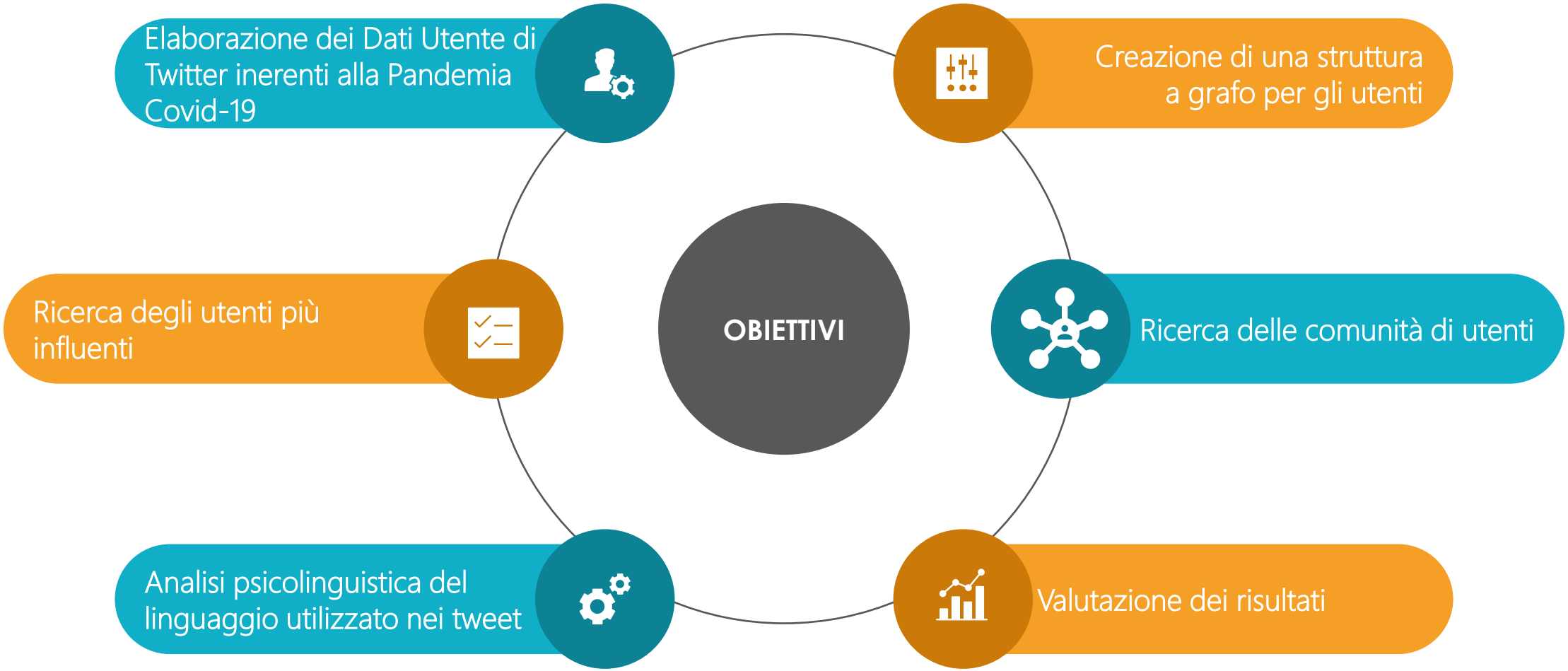
Ambito di Ricerca

Twitter è un **Social Network** che promuove **conversazioni** globali il cui obiettivo è permettere alle persone di **distribuire**, **creare** e **scoprire informazioni** affini ai loro **interessi**.

Il Coronavirus SARS-CoV-2 è un **virus respiratorio** che appartiene alla famiglia dei **coronavirus** e la sua comparsa ha **mutato** radicalmente lo stile di **vita** delle **persone**.



Obiettivi del progetto



Indice del progetto

Tecnologie

Linguaggio di programmazione

Strutture dati

Database a grafo

Libreria algoritmica

Libreria grafica

Live Editor

Indice del progetto

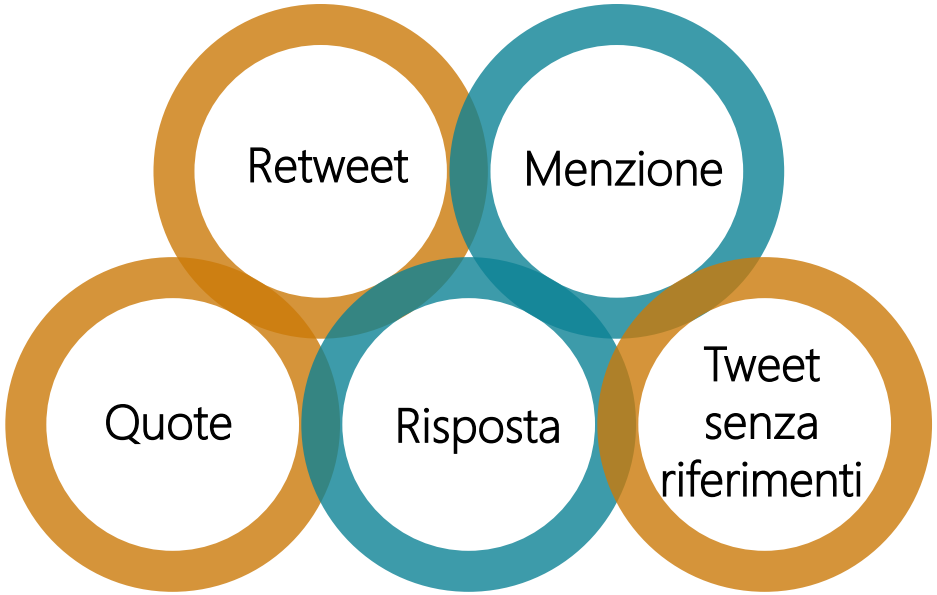
Struttura dei dati

I dati sono stati scaricati utilizzando le **API** di Twitter e fanno riferimento al periodo **Marzo 2020 – Dicembre 2021**.

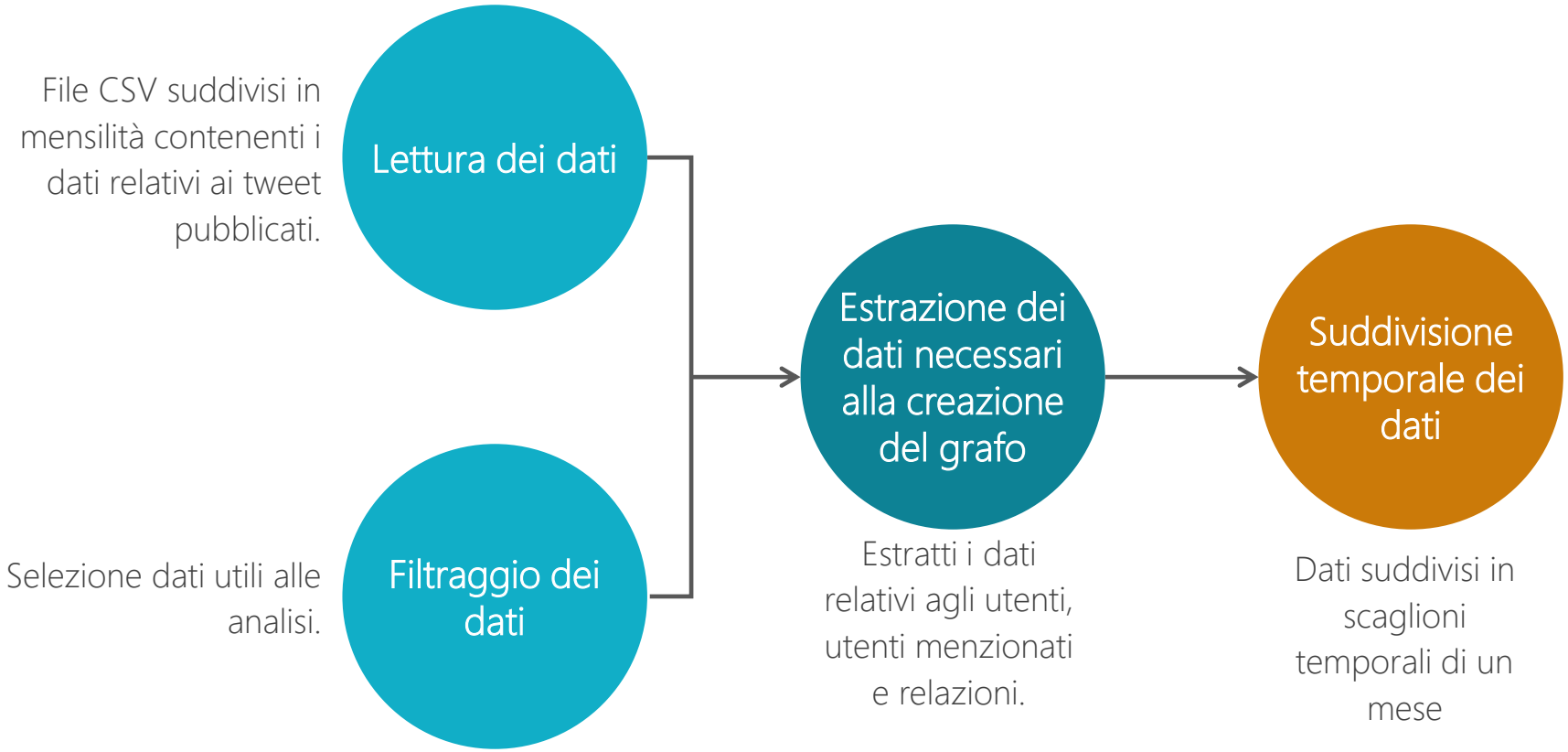
I dati hanno le seguenti caratteristiche:

- **Contengono** almeno un **hashtag** relativo alla **pandemia** di **Covid-19**.
- Ogni riga rappresenta un tweet con le **informazioni** relative al **tweet** e all'**autore** del tweet.
- In caso di **Retweet**, **Quote** e **Risposta** è presente il **riferimento** al **tweet originale**.
- In caso di **Menzione** sono presenti i **dati** relativi **all'utente menzionato**.
- Per ogni riga sono presenti **94 campi**.
- La dimensione complessiva dei dati è pari a **44 Gigabyte**.

Tipologie di Tweet



Estrazione dei dati



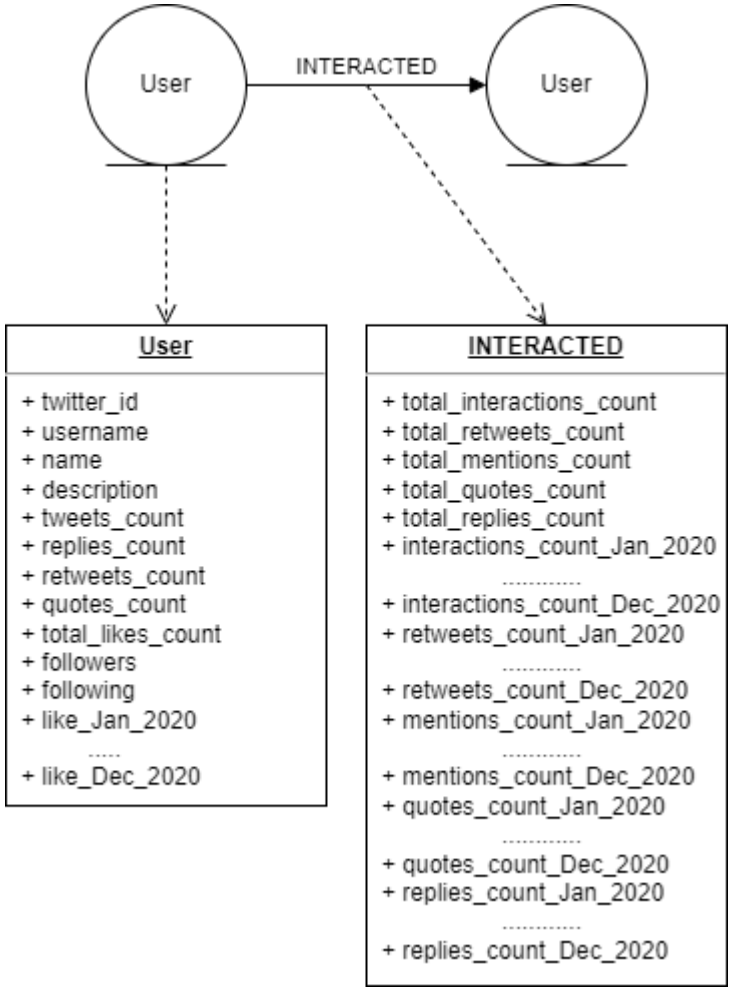
Creazione del Grafo

Obiettivo: memorizzare i dati in una struttura dati a grafo

I dati estratti in precedenza vengono letti attraverso tre **Query Cypher** per la creazione del **grafo Neo4j.**

Il **grafo** è costituito da un tipo di **nodo** ed un tipo di **relazione**:

- Il nodo **User** contiene le informazione degli utenti di Twitter che possono essere **autori** oppure **utenti menzionati**.
- La relazione **INTERACTED** rappresenta **l'aggregazione** di tutte le **relazioni**, suddivise per tipologia, avvenute tra due User.



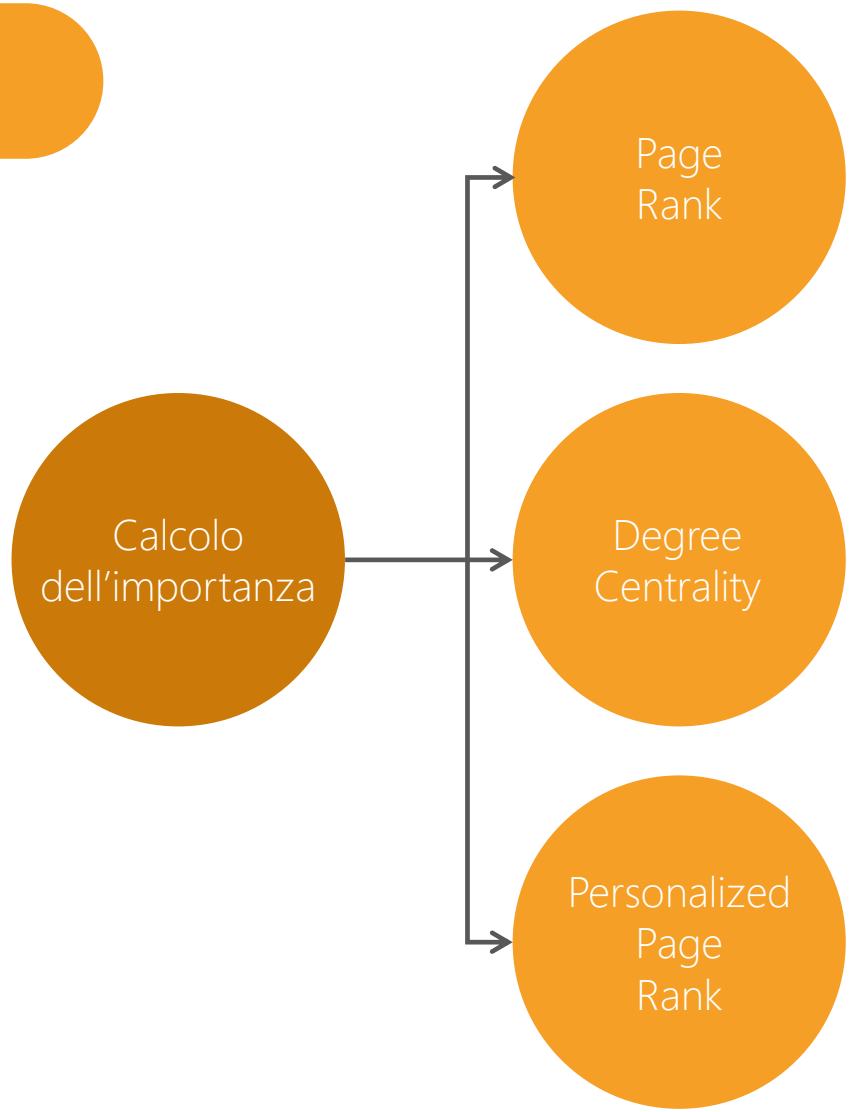
Algoritmi di Ranking

Obiettivo: ricerca degli utenti più influenti

Gli algoritmi sono eseguiti mese in mese, permettendo di analizzare **l'evoluzione temporale** ed il **cambiamento** delle figure **influenti.**

Elementi che definiscono **l'importanza** di un utente sono:

- **Like**: **apprezzamento** ricevuto
- **Interazioni: grado di attività**



- Misura **l'influenza** transitiva dei **nodi.**
- Numero di **interazioni** come **peso** degli **archi.**

- Due configurazioni:
1. Archi **entranti** -> Importanza **ricevuta.**
2. Archi **uscenti** -> Importanza **trasmessa.**
- **Peso** degli **archi** è combinazione di **like** ed **interazioni.**

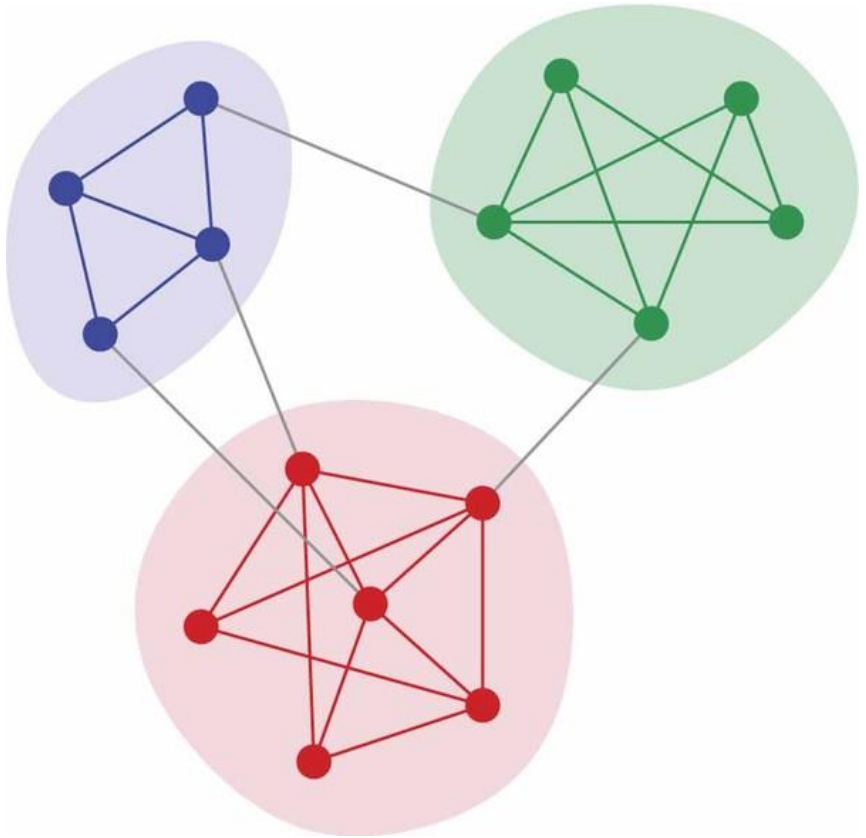
- Funzionamento **analogo** al **PageRank** con **l'aggiunta** dei **like.**
- Numero di **interazioni** come **peso** degli **archi.**

Community Detection

Obiettivo: ricerca delle comunità

Algoritmo di **Label Propagation** in due configurazioni differenti:

- **Retweet** come peso: ricerca di comunità con **idee affini.**
- **Interazioni** complessive come peso: ricerca di comunità con un alto **livello** di **interazione.**



Analisi Psicolinguistica

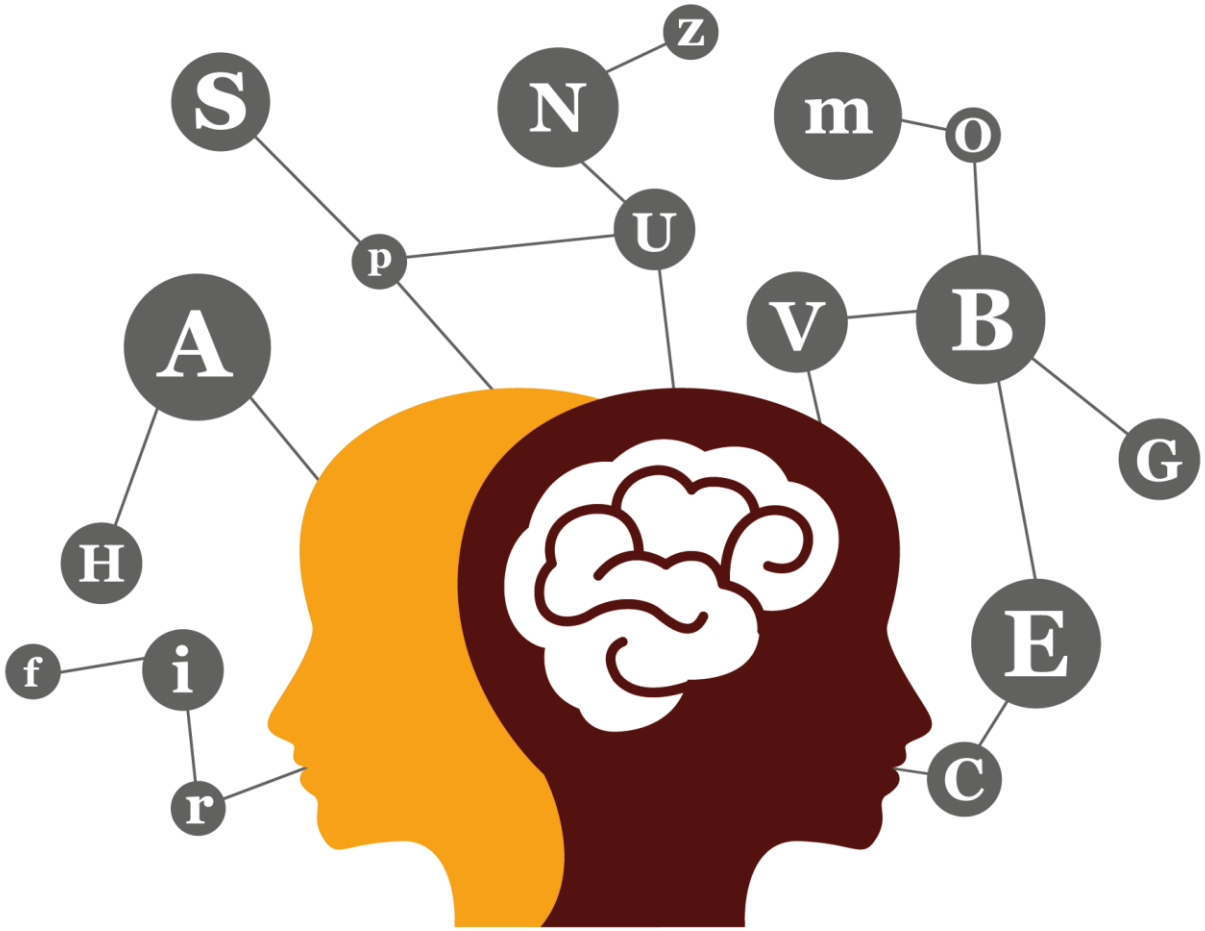
Obiettivo: analisi del contenuto dei Tweet

La **Psicolinguistica** è la fusione di **psicologia** e **linguistica**:

- La prima si dedica allo studio del **pensiero**, delle **emozioni** e del **comportamento**.
- La seconda studia le manifestazioni del **linguaggio.**

L'analisi assegna alle singole parole contenute nei tweet una specifica **sotto-categoria psicolinguistica:**

- **Macro-categoria**: racchiude molteplici sotto-categorie.
- **Sotto-categoria**: contiene i **termini specifici** che la identificano.



Categorie Psicolinguistiche

BIOLOGICA:
- Corpo
- Salute
- Morte
- Sintomi
- Malattie

EMOTIVA:
- Ansia
- Rabbia
- Emozioni Positive

MEDICA:
- Cure
- Covid
- Dispositivi Medici
- Vaccini
- Vaccini Anticovid

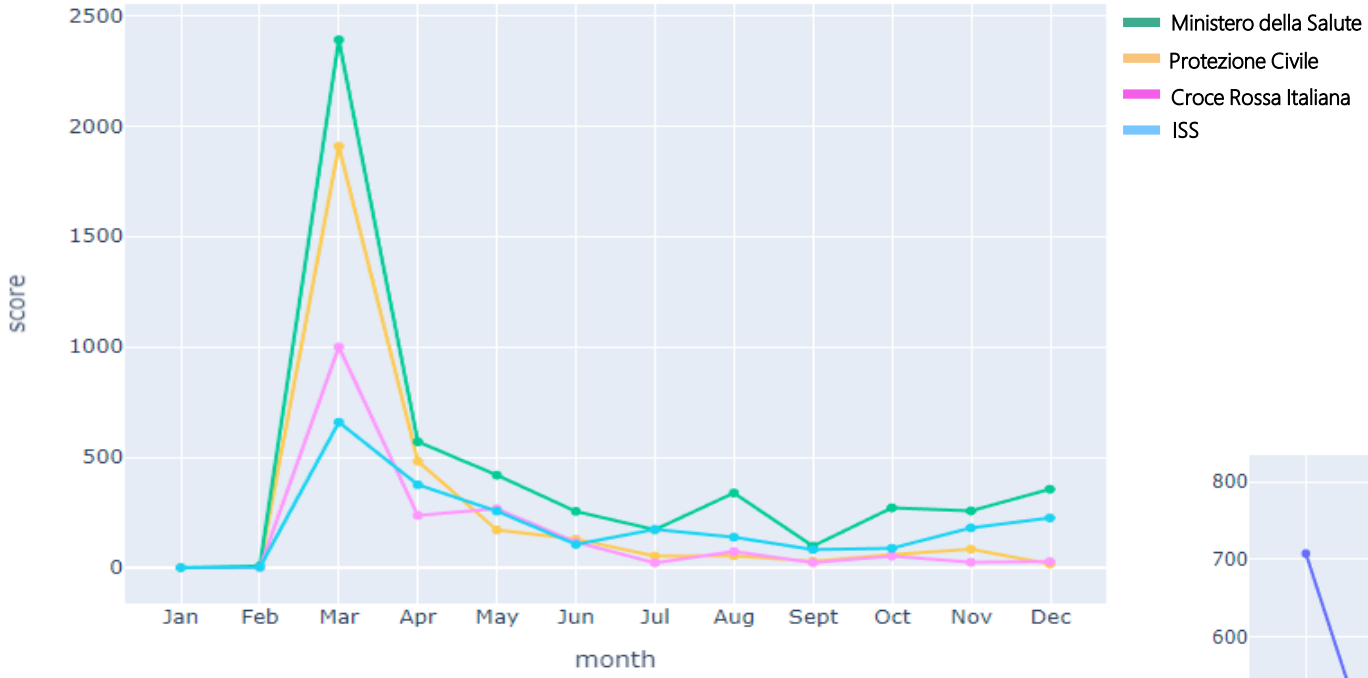
SOCIALE:
- Famiglia
- Lavoro
- Istituzioni
- Provvedimenti
- Campagna vaccinale

LINGUISTICA:
- Volgari
- Offensivi
- Linguaggio tecnico/scientifico
- Linguaggio ideologico

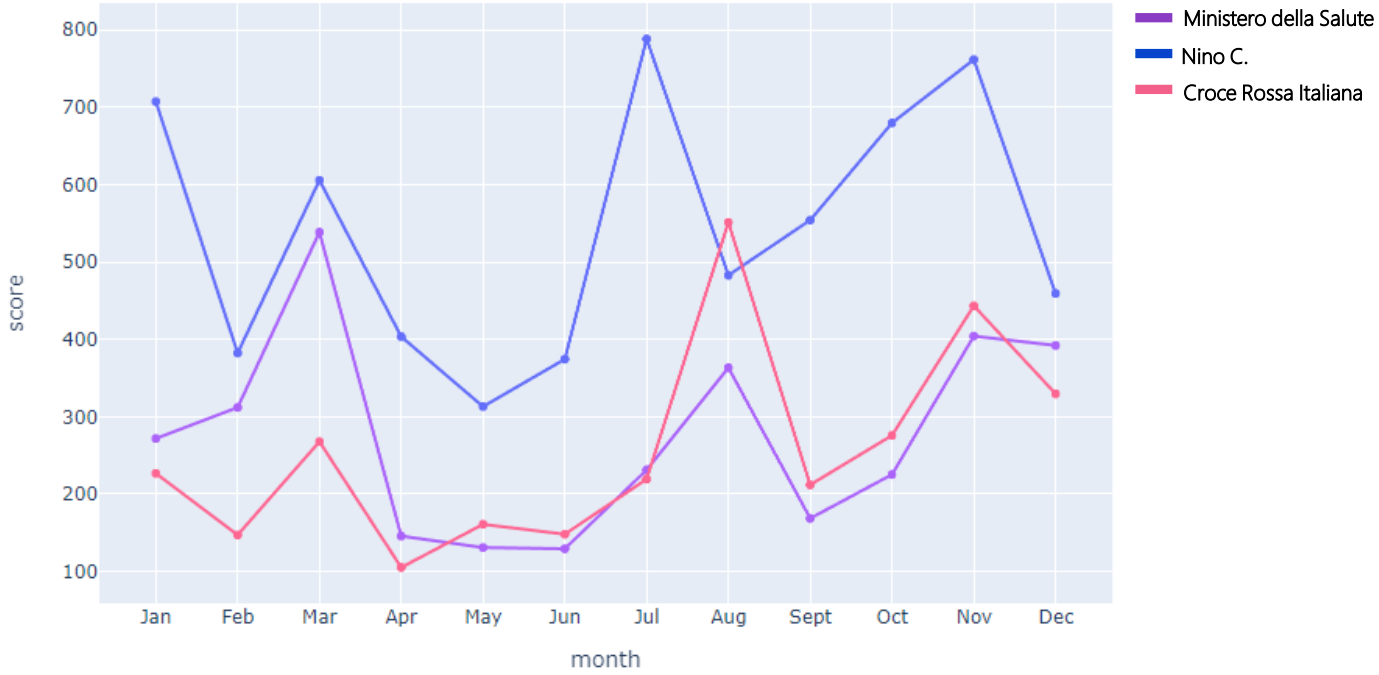
Indice del progetto

Risultati Ranking

PageRank Istituzioni 2020

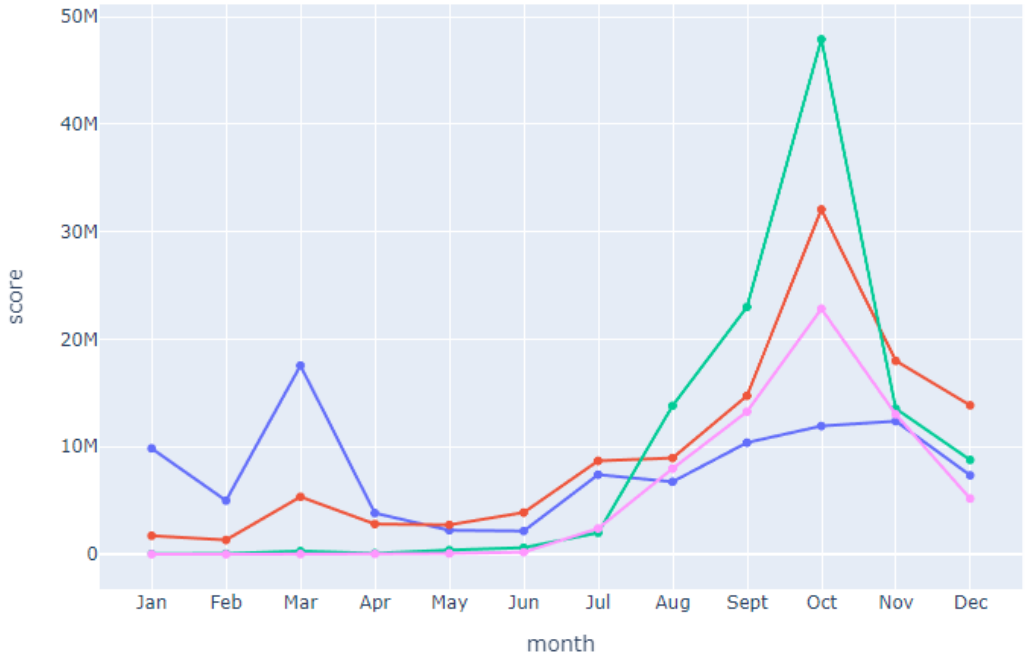


PageRank Medici vs Istituzioni 2021

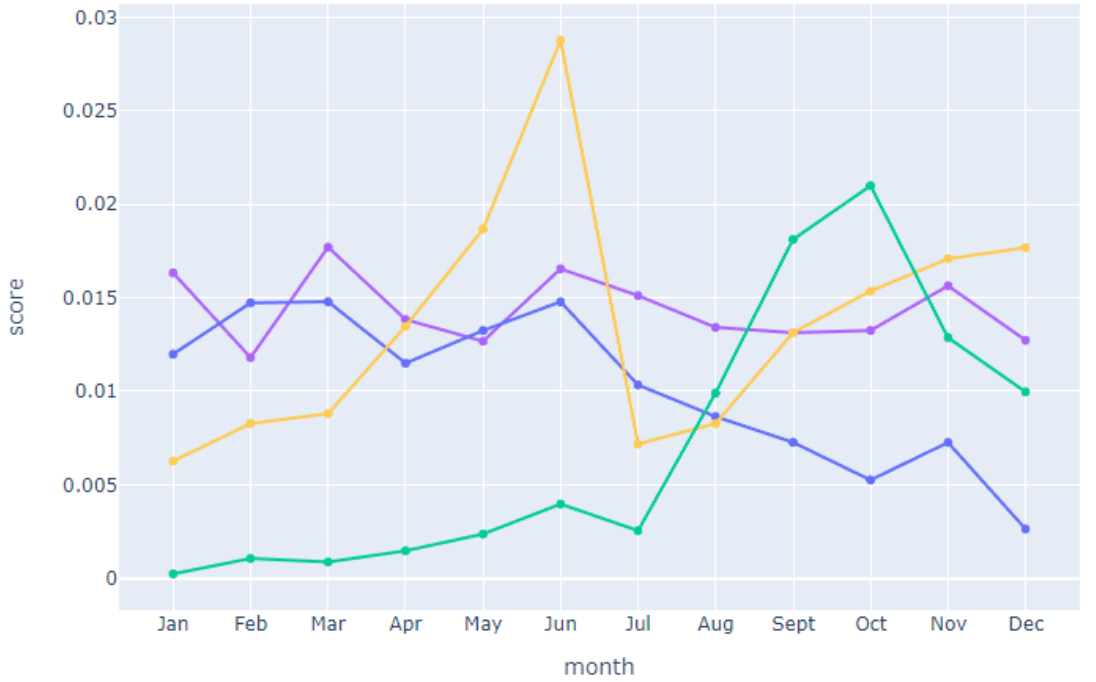


Risultati Ranking

Degree Centrality Medico vs NoVax 2021

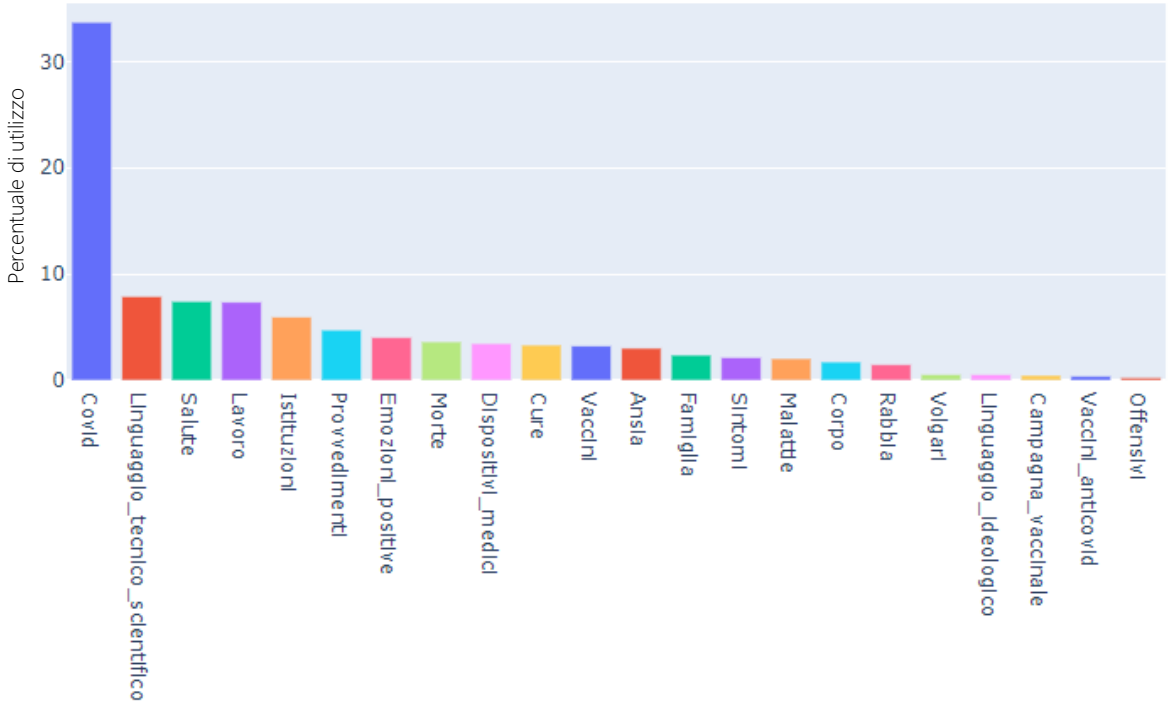


P.P.R. Medici vs NoVax 2021

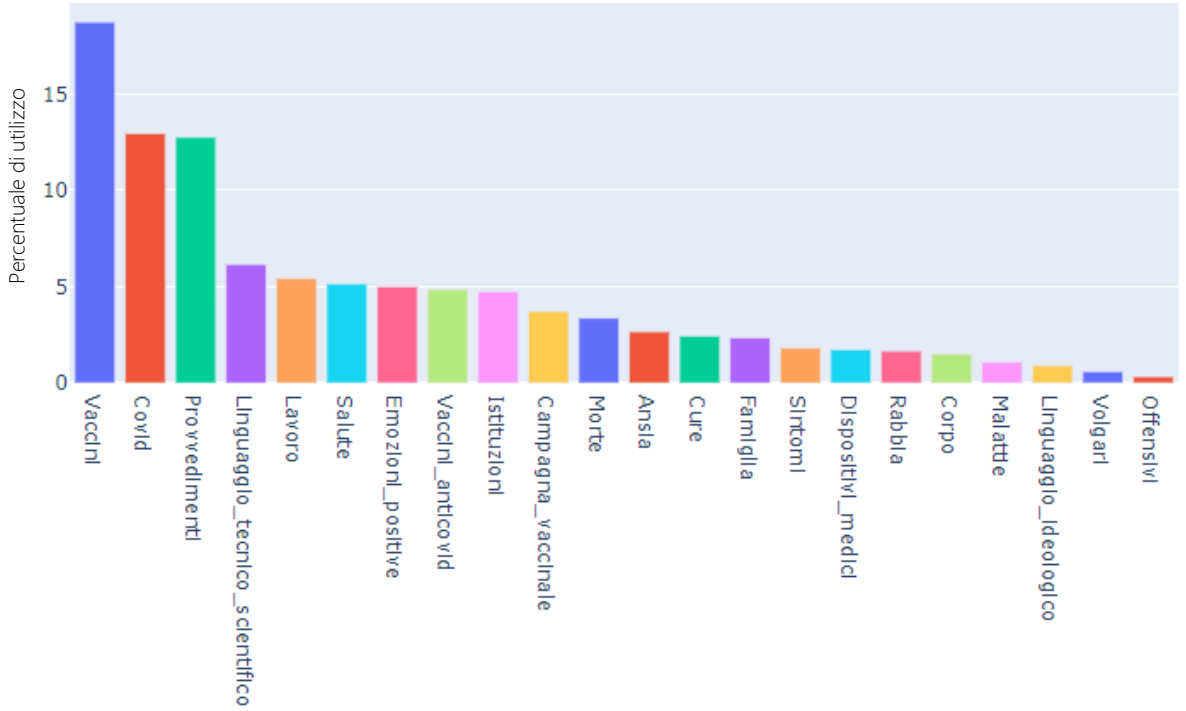


Risultati Psicolinguistici

Anno 2020



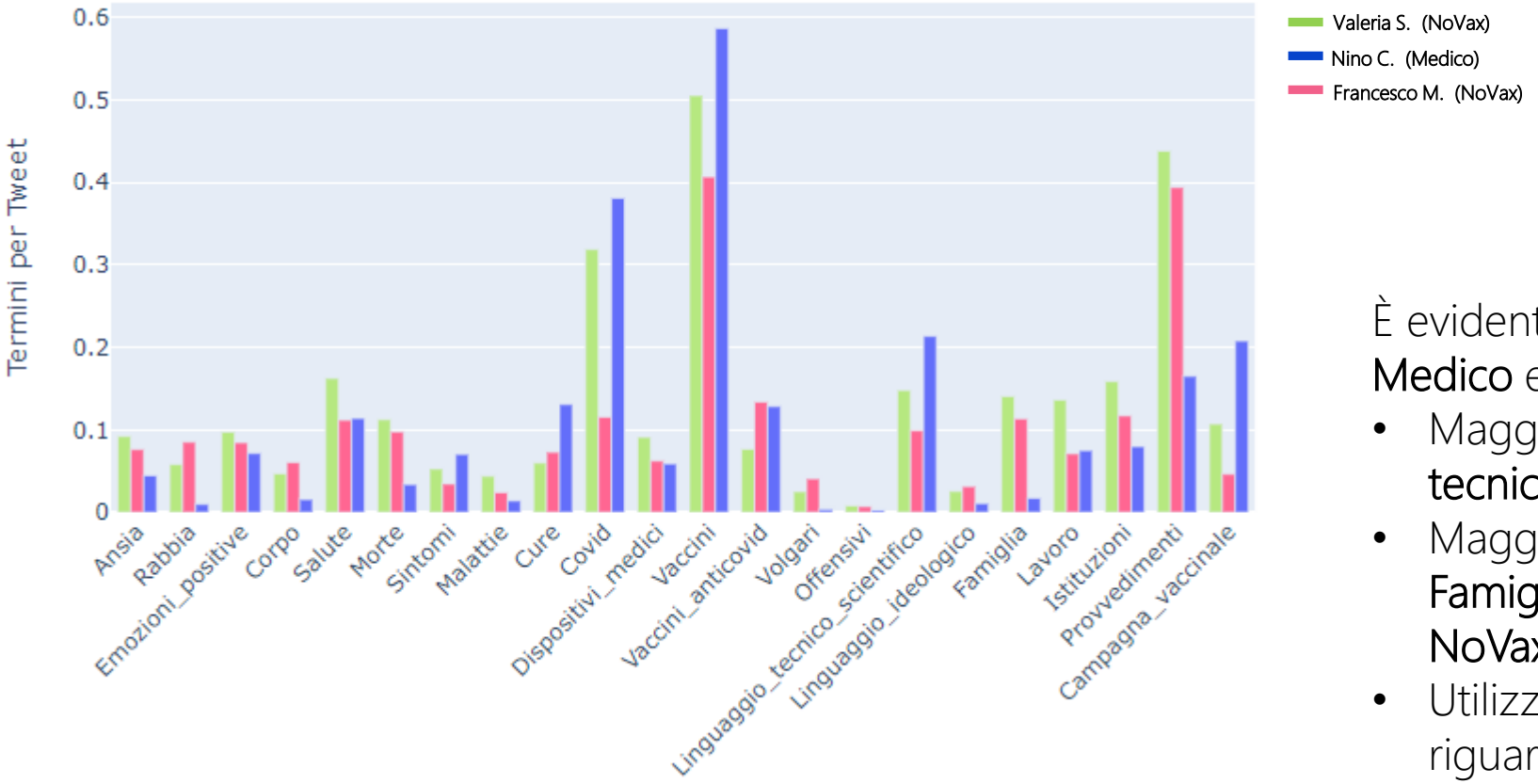
Anno 2021



I **risultati** evidenziano che:

- Nell'anno 2020 l'argomento principale di discussione è il **Covid.**
- Nell'anno 2021 la percentuale di utilizzo di terminologia **Covid** è dimezzata in favore di **Vaccini** e **Provvedimenti.**

Risultati Psicolinguistici



È evidente la differenza di **linguaggio** tra **Medico** e **NoVax**:

- Maggiore utilizzo di termini **tecnico/scientifici** da parte del **Medico.**
- Maggiore utilizzo di termini riguardanti **Famiglia** e **Provvedimenti** da parte dei **NoVax.**
- Utilizzo da parte del **Medico** di termini riguardanti **Cure** e **Campagna Vaccinale** doppio rispetto ai NoVax.

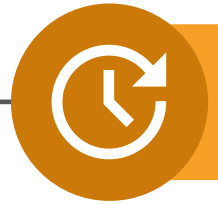
Indice del progetto

Conclusioni

Risultati Raggiunti

- **Medici** ed **Istituzioni centrali** nell'anno **2020** con relativo utilizzo di un linguaggio **Tecnico/Scientifico**.
- Da **Giugno 2021** forte **crescita** e centralità per utenti **NoVax** e **NoGreenPass** con relativo utilizzo termini riguardanti **Provvedimenti**, **Famiglia** ed **Idealistici**.
- Nell'anno **2021** sono presenti **due** grandi **comunità** con **ideologie opposte**.

Sviluppi Futuri



- Analisi della **variazione** dei membri delle **comunità** al passare dei mesi.
- Analisi di **veridicità** di contenuti e fonti dei **Tweet** degli utenti.

Grazie per l’attenzione