

Green Pass on Twitter: Social Interaction in Italy

Andrea Munarin,
Ca' Foscari University of Venice, Italy
Corresponding Author e-mail: 879607@unive.it

Abstract

In Italia, nel 2021 sono entrati in vigore i decreti che introducono Green Pass e Super Green Pass. La certificazione verde “COVID-19 – EU digital COVID certificate” nasce per agevolare la libera circolazione in sicurezza dei cittadini nell’Unione Europea durante la pandemia di COVID-19, ma nel corso dei mesi ha assunto una funzione più ampia e strutturata. Si sono così generate due correnti di pensiero nella popolazione italiana: chi manifesta in favore all’abolizione di tali regole e chi sostiene l’idea intravedendo una soluzione per contenere la diffusione del virus. Lo studio descritto in questo documento verterà sulle interazioni tra questi due schieramenti. Strumento chiave della ricerca è la cosiddetta “*sentiment analysis*”: un campo dell’elaborazione del linguaggio naturale che si occupa di costruire sistemi per l’identificazione ed estrazione di opinioni dal testo. Grazie a questa è stato possibile determinare la polarizzazione di ciascun utente coinvolto nello studio e infine valutare la presenza di *echo chambers*. Considerata la forte influenza dei social network nella definizione dell’opinione pubblica, la ricerca si focalizza sulle interazioni degli utenti di Twitter. L’intera analisi è basata sui Tweet del mese di gennaio 2022 a seguito di una serie di riforme del governo dove green pass e super green pass sono come elementi cardine.

1. Introduzione

L’introduzione della certificazione verde è fonte di dibattiti e controversie, non solo in Italia, ma anche in altri Paesi. Sono presenti diverse critiche alle legislazioni che prevedono l’utilizzo del Green Pass. Alcune di queste denunciano i difetti di questo provvedimento sanitario illustrando possibili perfezionamenti. Ad esempio, uno studio effettuato in Israele: “*The Israeli Experience with the “Green Pass” Policy Highlights Issues to Be Considered by Policymakers in Other Countries*” (by Ruth Waitzberg, ORCID, Noa Triki, Sharon Alroy-Preis, Tomer Lotan, Liat Shiran and Nachman Ash, ottobre 2021) ha analizzato i punti di vista della popolazione israelita sui temi vaccino e green pass, sottolineando, in conclusione, la presenza di diversi problemi. Secondo l’articolo i politici dovrebbero valutare strategie che integrino i suggerimenti proposti, in modo da curare gli aspetti maggiormente criticati. L’obiettivo finale è diminuire il grado di disapprovazione della popolazione. Alcune osservazioni proposte dall’analisi sopracitata sono: una maggiore attenzione alla protezione della privacy e la riduzione delle situazioni nelle quali il possesso della certificazione risulta obbligatorio, evitando di imporlo su attività di prima necessità. In questo documento verrà analizzata come l’introduzione delle normative sulla certificazione è stata percepita dagli italiani, in particolar modo facendo riferimento agli utenti di Twitter (rappresentativi di un sottoinsieme della popolazione) e ai tweet più discussi e con più interazioni del mese di gennaio 2022.

2. Data Collection

L’intera collezione di dati è stata raccolta tramite l’utilizzo delle Twitter API, disponibili per chiunque abbia un account Twitter. In particolare, il livello di accesso utilizzato è “elevated”, ma tutte le collezioni sono ottenibili anche con il piano base. I dati utilizzati per i profili dell’utente sono utilizzati solo per quelli di cui era garantito l’accesso evitando quindi di trattare informazioni di profili privati. I dati sono stati scaricati dai Tweet dell’ultimo mese (a causa delle limitazioni imposte dall’API), tuttavia il contenuto dei tweet e i nomi degli autori sono nascosti e mai presenti nei dataset utilizzati

nello studio. La collezione finale è, inoltre, un'aggregazione di dati, in questo modo risalire all'identità degli utenti discussi nello studio risulta ancora più complesso. Sono stati rispettati i termini, le condizioni e le politiche sulla privacy di Twitter. Sono stati identificati prima tutti i tweet del mese (basati sugli hashtag: #greenpass o #supergreenpass e solo in lingua italiana), successivamente dopo l'analisi del sentiment sono stati individuati i tweet ritenuti centrali nello studio: ordinando per le public metrics (ovvero considerano la somma di commenti, like e retweet) sono stati estratti i primi 5 tweet considerati positivi e i primi 5 tweet considerati negativi. Infine, basandosi sulle conversazioni di tali tweet (risposte tra gli utenti) sono stati determinati i profili dello studio. Da ciascun profilo sono stati scaricati e analizzati tramite *sentiment analysis* i tweet presenti nelle proprie timeline (in particolare dagli ultimi mille tweet di ciascuna timeline sono stati estratti quelli relativi al green pass e al vaccino). Nella tabella sottostante (Tabella 1) sono indicati il numero di dati finali ricavati per lo studio:

Dati	Numero
Tweet ultimo mese	7029
Tweet selezionati	10
Tweet delle "conversation" per le interazioni	1955
Utenti di cui si è scaricata la timeline	1242

Tabella 1 - tabella che contiene le informazioni relative ai dati scaricati per lo studio sul Green Pass

Infine, osserviamo il numero di Tweet "negativi" (6404) e "positivi" (625) dell'ultimo mese nella Figura 1.

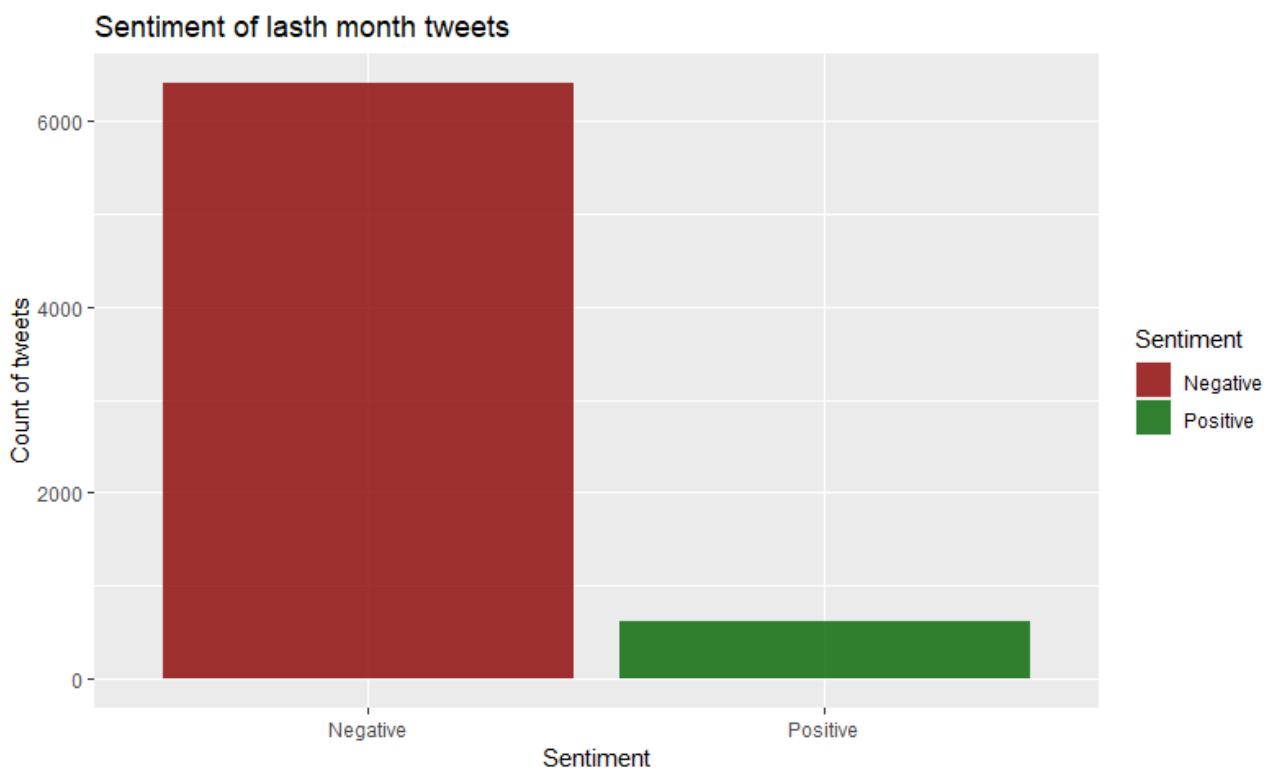


Figura 1 - Numero di Tweet dell'ultimo mese considerati "positivi" e "negativi" dall'analisi del sentiment

3. Analisi e risultati

La Polarizzazione è un fenomeno molto diffuso nei social network e verificarne la presenza può essere molto utile per determinare se un'opinione neutrale perde peso nel pubblico a favore di posizioni più estreme (nei limiti del contesto del social network su cui si basa l'analisi). Anche verificare la presenza di echo chambers ha molta rilevanza in uno studio dove si valuta l'interazione tra utenti con due correnti di pensiero differenti. In questo caso, osservare utenti contrari al green pass che interagiscono solo tra di loro, senza la presenza di contatti con tweet o utenti che invece sono favorevoli, determinerebbe la presenza di camere d'eco: una situazione in cui le informazioni, le idee o le credenze vengono amplificate o rafforzate dalla comunicazione e dalla ripetizione all'interno di un sistema definito. In questo documento l'analisi si focalizza proprio su questi due aspetti contestualizzati al tema del green pass e del super green pass in Italia.

Prima di iniziare l'analisi e visualizzare i risultati va discussa la possibile presenza di bias: distorsioni che compromettono l'integrità e l'oggettività dei dati. L'analisi è stata effettuata cercando il più possibile di limitare la presenza di queste distorsioni. Sono comunque presenti alcune operazioni effettuate nel processo di ottenimento dei dataset che potrebbero compromettere la correttezza dei dati, come:

- Bias relativi alla posizione geografica: gli utenti scelti non sono tutti quelli italiani ma solo un sottoinsieme (i tweet sono stati filtrati per lingua italiana, non abbiamo la certezza che gli utenti siano italiani).
- Bias relativi al tipo di dati: i dati rappresentano solo una parte della popolazione italiana, per cui vanno considerati solo nell'ambito degli utenti Twitter.
- Bias relativi alla creazione dei dataset: come vedremo in seguito se un profilo è segnalato privato o non esistente ma è un elemento presente nella costruzione dell'interazione sociale viene determinato per lui un valore di polarizzazione pari a 0 (polarizzazione nulla)
- Bias relativi alla tecnologia utilizzata: Sono presenti alcuni limiti dell'utilizzo dell'API di Twitter come il periodo relativi ai tweet ("Recent Tweet") e il numero dei tweet da scaricare.

Lo studio si avvale di diversi dataset, ognuno dei quali contiene informazioni utili allo studio:

- Tweet dell'ultimo mese: sui quali è stata effettuata un'analisi del sentiment ed è stata calcolata la somma delle metriche pubbliche.
- Tweet selezionati: i primi cinque tweet "*negativi*" e "*positivi*" basandosi sulla somma delle metriche pubbliche.
- Conversazioni: insieme dei collegamenti tra i vari utenti. Su ognuno dei dieci tweet sopracitati preso il "*conversation_id*" sono stati scaricati gli ultimi mille tweet relativi alla stessa conversazione (insieme delle risposte).
- Timeline: per ogni autore presente nella conversazione sono stati scaricati gli ultimi mille tweet pubblicati, filtrati poi per temi inerenti al green pass e infine valutati tramite sentiment analysis contando infine il numero di tweet "*positivi*" e il numero di tweet "*negativi*" pubblicati.

Osserviamo quindi nella Figura 2 tutti i tweet scaricati raggruppati per sentiment con la somma delle relative metriche.

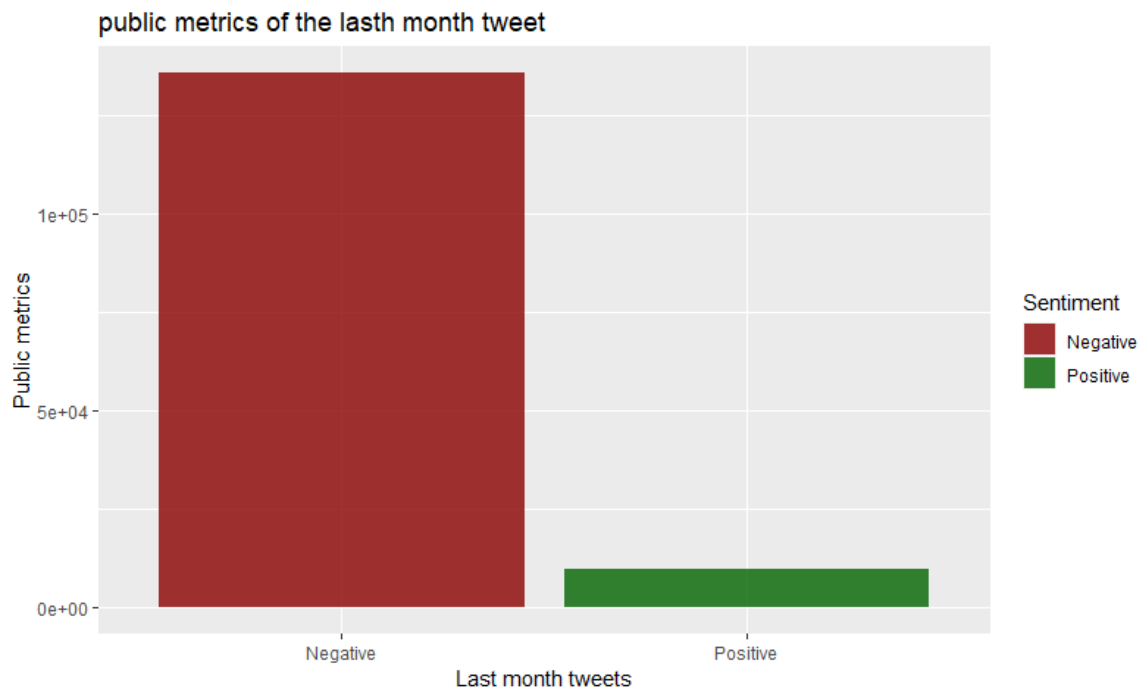


Figura 2 - Impatto dei tweet "positivi" e "negativi" dell'ultimo mese misurati secondo la somma di like, commenti e retweet

Possiamo subito notare come i tweet ritenuti “*negativi*” hanno molto più impatto rispetto ai tweet considerati “*positivi*”. Questo è anche dovuto al fatto che i tweet “*negativi*” sono molti di più rispetto a quelli “*positivi*” come da Figura 1 presente nel capitolo “Data Collection”.

Già da questo si deduce come su Twitter, in Italia, gli utenti siano per la maggior parte contrari all’utilizzo di questo provvedimento sanitario per limitare la diffusione della pandemia. Vediamo quindi anche la somma delle metriche relativamente ai dieci tweet selezionati nella Figura 3.

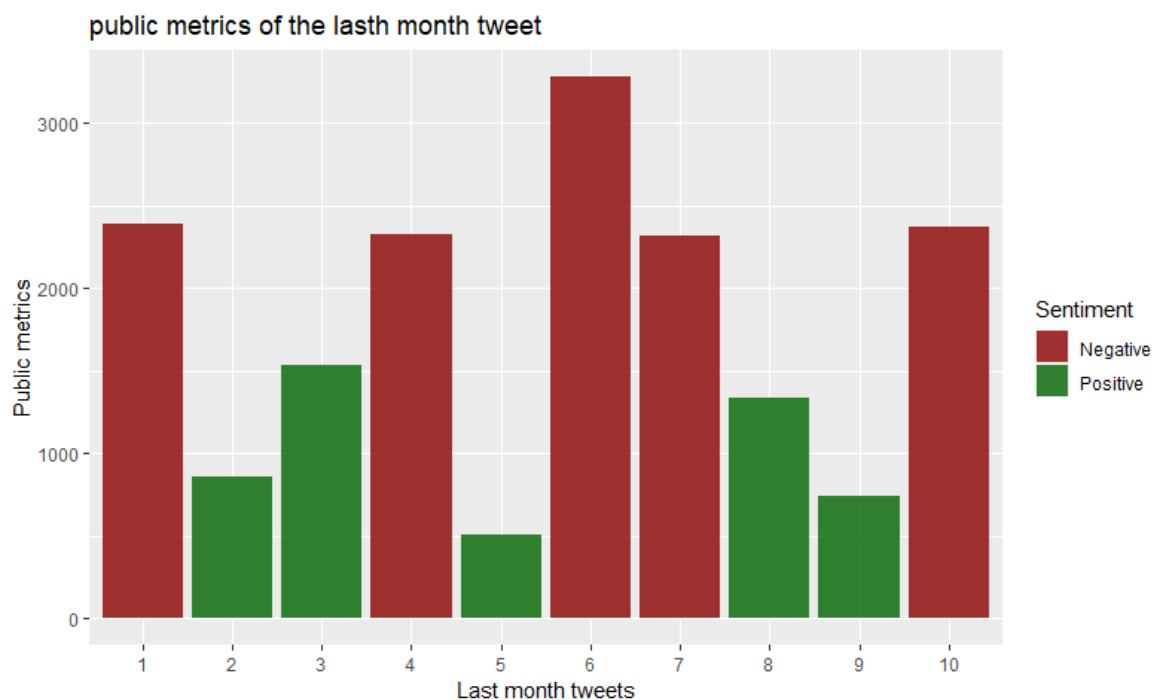


Figura 3 - Impatto dei dieci tweet selezionati da cui sono poi state scaricate le conversazioni

Come ci si poteva aspettare, anche dal grafico precedente i cinque tweet “*negativi*” hanno un impatto maggiore rispetto ai Tweet “*positivi*” per cui saranno anche quelli sui quali ci sarà maggior interazione e maggiori retweet.

Valutiamo ora la polarizzazione degli utenti. Per determinarla è stato utilizzato il dataset contenente le timeline: per ogni utente sono stati calcolati il numero di tweet ritenuti “*positivi*” e “*negativi*”. Sulla base di questi valori è stata determinata la polarizzazione come illustrato nell’equazione 1.

$$Polarizzazione = \frac{N_{Tweet_{positivi}} - N_{Tweet_{negativi}}}{N_{Tweet_{positivi}} + N_{Tweet_{negativi}}}$$

Equazione 1 - Formula per il calcolo della polarizzazione di un utente

Il valore di polarizzazione calcolato per ciascun utente risulterà compreso tra -1 e $+1$ dove:

- -1 : l’utente è completamente contrario all’utilizzo del green pass.
- 0 : l’utente viene considerato neutro (oppure non è stato possibile scaricare le timeline a causa di profili privati o non più esistenti) e non viene definito né a favore né contrario.
- $+1$: l’utente è completamente favorevole al concetto del green pass.

Sulla base di queste considerazioni, sono stati effettuati tutti i calcoli necessari e visualizzando i risultati otteniamo:

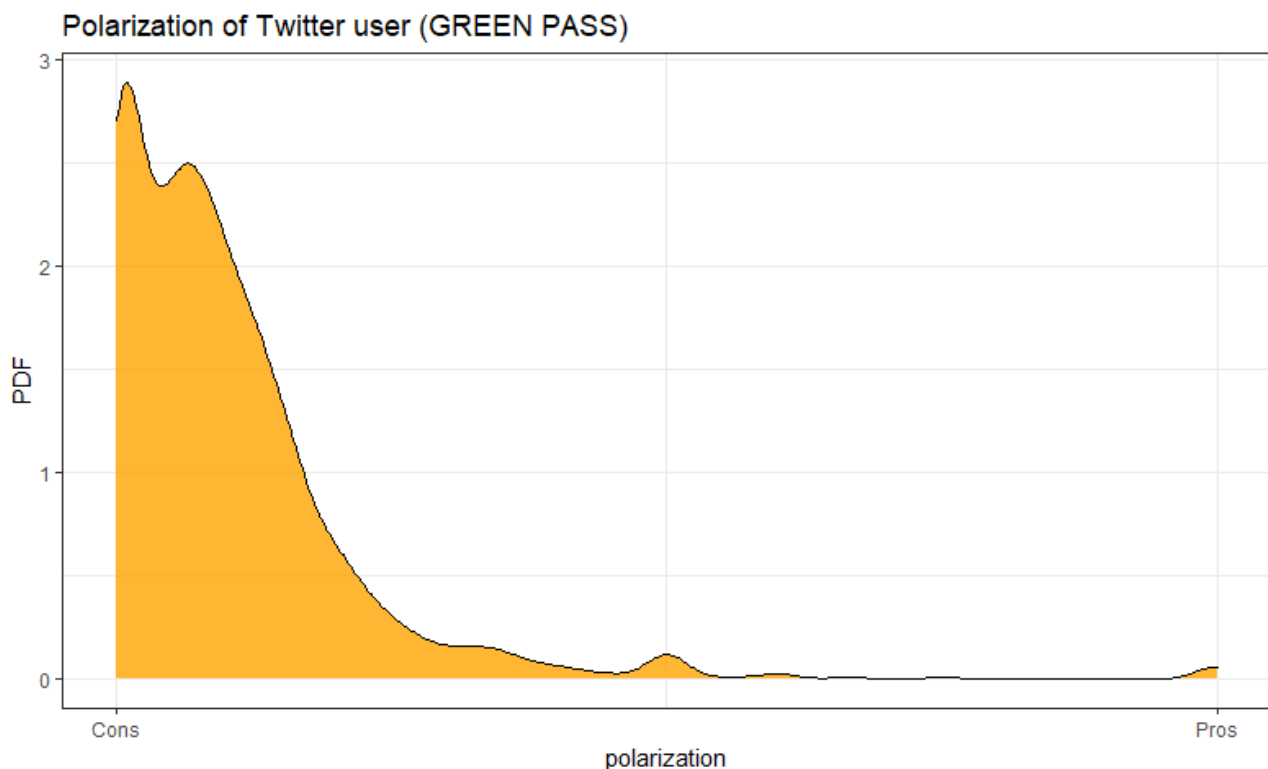


Figura 4 - Polarizzazione della comunità di utenti su Twitter sulla base dei tweet pubblicati e ricavati dalla propria timeline

Osserviamo quindi nella Figura 4 come gli utenti di Twitter sono fortemente polarizzati solo negativamente, con una bassissima percentuale di utenze favorevoli o neutre: la maggioranza di utenti sono contrari alle disposizioni degli ultimi decreti-legge.

Possiamo quindi constatare, considerando i grafici precedenti, che nei Tweet Italiani troviamo per lo più persone contrarie agli ultimi provvedimenti governativi. Questi utenti desiderano una soluzione diversa per contrastare la pandemia e isolare la diffusione dal Covid. È bene precisare come lo studio si concentri solamente sui pensieri relativi alla certificazione verde e non al vaccino o all'utilizzo dei tamponi per monitorare l'evolversi del virus.

Sulla base di questi dati ci aspettiamo che i tweet positivi non ricevano un grande consenso all'interno della piattaforma ma siano invece fonte di discussione e dibattito tra utenti che fanno parte della corrente di pensiero opposta. Ipotizziamo quindi sia presente una gigantesca echo chamber, dove gli utenti contrari possono anche entrare in contatto con quelli positivi o con i loro tweet ma l'opinione *"negativa"*, avendo un maggior consenso (come determinato precedentemente), non può fare altro che rafforzarsi. Quest'ipotesi implicherebbe una situazione nella quale qualunque tipo di tweet (*"positivo"* o *"negativo"*) sia una fonte di interazione per profili contrari alle normative, intravedendo poche eccezioni di profili neutri e ancora meno di profili favorevoli.

Andiamo quindi a verificare graficamente come avvengono le interazioni degli utenti e se le ipotesi precedenti si rivelano corrette. Costruiamo un grafo sulla base delle interazioni avvenute: tra due utenti è presente un collegamento se un utente ha risposto all'altro (dataset delle conversazioni). Per semplicità consideriamo un grafo non orientato. Ogni vertice del grafo verrà rappresentato da una forma e un colore diverso sulla base della polarizzazione calcolata precedentemente. Anche la dimensione dei vertici dipenderà sempre dalla polarizzazione. Inoltre, avremo due tipi diversi di *"link"* all'interno della nostra visualizzazione: i *link* verdi continui rappresentano le interazioni che sono avvenute a partire da un tweet *"positivo"* mentre quelli rossi e tratteggiati identificano una *"conversazione negativa"*. Andiamo quindi a visualizzare graficamente il risultato ottenendo:

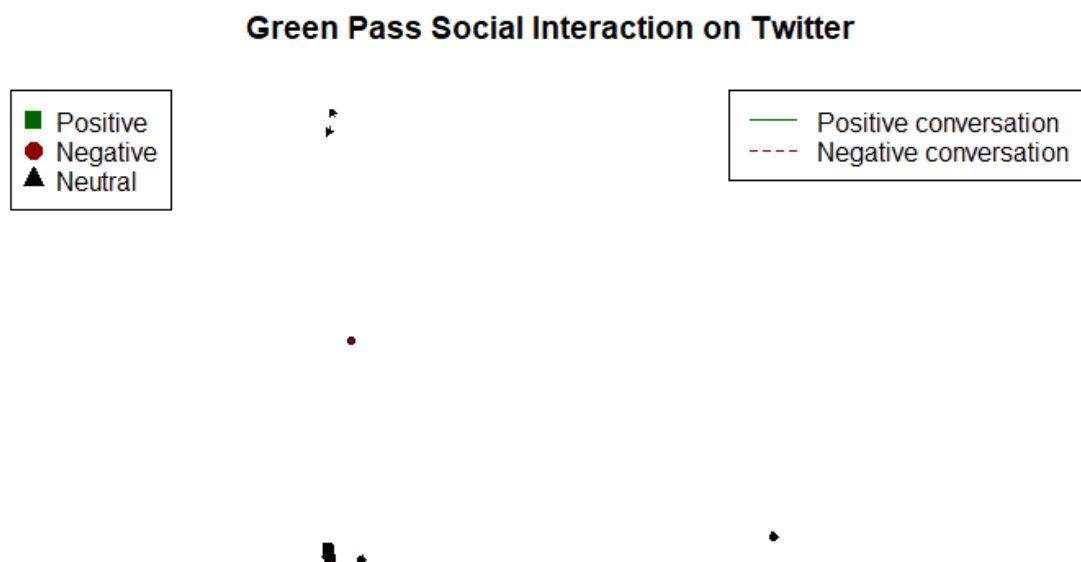


Figura 5 - Componenti connesse presenti nel grafo che rappresenta le interazioni tra le risposte degli utenti nei tweet selezionati

Come possiamo osservare dalla Figura 5 otteniamo un grafo non connesso, in particolare dalle dieci conversazioni risultano sei componenti connesse. Cerchiamo quindi di studiare la più rappresentativa delle interazioni, ovvero la componente connessa più grande tra tutte quelle nell'immagine sopra, e visualizziamola tramite l'algoritmo di Fruchterman Reingold:

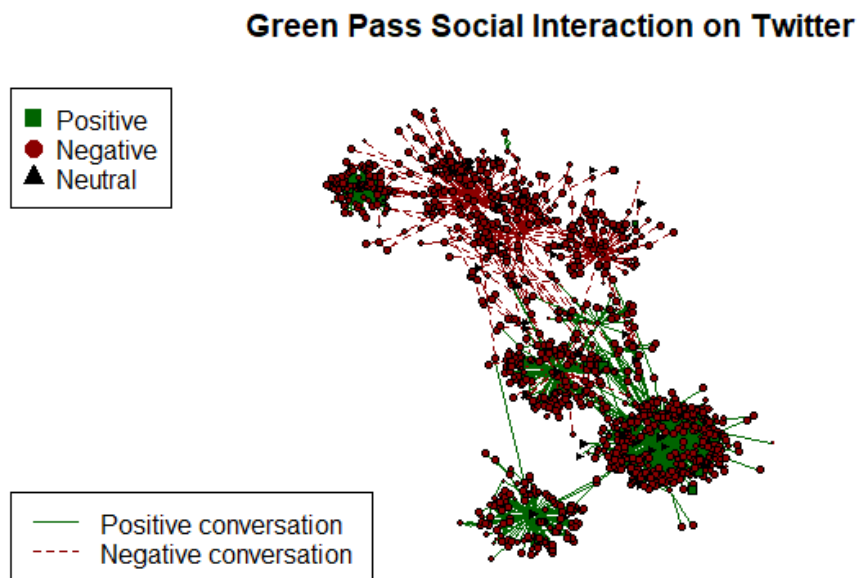


Figura 6 - Grafo della componente connessa più grande rappresentativo dell'interazione sociale in Twitter relativa al fenomeno del Green Pass

Possiamo notare sulla Figura 6 come siano presenti per lo più “*negative user*” e la maggior parte di questi si concentrano intorno agli utenti che hanno pubblicato i post “*positivi*” (alcuni dei quali sono segnalati tramite triangoli neri, quindi come se avessero polarizzazione nulla, probabilmente a causa delle distorsioni definite precedentemente).

Da questo grafo possiamo dedurre che l'ipotesi sviluppata precedentemente era corretta: è presente, infatti, una gigantesca “camera d'eco” in cui tutti gli utenti considerati contrari alla certificazione si trovano d'accordo sia nelle risposte di un tweet “*positivo*” che su quelle di un tweet “*negativo*”. La scarsa presenza di profili favorevoli impedisce la creazione di un ambiente in cui anche la loro idea sia condivisa e rafforzata, ostacolando così la presenza di una seconda echo chamber con coloro che approvano la strategia governativa. Nel centro delle conversazioni notiamo come tutti i vertici siano fortemente polarizzati, mentre quelli con dimensioni minori e quindi meno polarizzati risultano essere vertici isolati, persone che non hanno partecipato attivamente alla conversazione. Gli utenti che uniscono le varie conversazioni, creando i ponti tra le varie componenti sono fortemente polarizzati negativamente.

È importante considerare come gli utenti di Twitter critichino attivamente le ultime decisioni governative sull'obbligo del green pass e sul suo valore.

4. Conclusioni

Il risultato della ricerca ha sottolineato la posizione della popolazione italiana su Twitter nei confronti della certificazione verde e del suo utilizzo. In particolare, ha mostrato come gli utenti siano fortemente contrari all'utilizzo di tale strategia sanitaria per limitare la diffusione del virus e desiderino una soluzione differente. Questo fenomeno è molto rilevante poiché se un utente dovesse decidere di iscriversi a Twitter, tra i tweet relativi alle ultime notizie sul green pass (che siano “*positivi*” o “*negativi*”) ritroverebbe, sia nei commenti che nelle risposte, utenti che non sono favorevoli, influenzando così il suo pensiero. Considerando quindi la potenza dell'influenza sociale attraverso i Social Network e la netta presenza di una camera d'eco è immediato pensare che di questo passo il malcontento comune non può far altro che aumentare portando a situazioni di estremismo. Risulta quindi importante valutare i risultati di quest'analisi per trovare una soluzione che garantisca una maggior approvazione da parte della popolazione. In questo modo aumenterà la presenza di utenti “favorevoli” e “positivi” generando un maggior numero di interazioni tra profili inversamente polarizzati, rompendo così la camera d'eco. Se questo avvenisse, anche posizioni favorevoli potranno trovare consenso creando un equilibrio che eviti la formazione di utenti con posizioni estremiste. Infatti, è altamente probabile che maggiore è la distanza emotiva tra lo stesso concetto in due echo chambers, maggiore è il loro effetto di polarizzazione sugli utenti.

Bibliografia

- Nozza, D., Bianchi, F., & Hovy, D. (2020). What the [MASK]? Making Sense of Language-Specific BERT Models. arXiv preprint arXiv:2003.02912.
- Bianchi, F., Nozza, D., & Hovy, D. (2021). FEEL-IT: Emotion and Sentiment Classification for the Italian Language. 11th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis.
- Sprugnoli, R. (2020). MultiEmotions-it: A new dataset for opinion polarity and emotion analysis for Italian. In 7th Italian Conference on Computational Linguistics, CLiC-it 2020 (pp. 402–408)
- The Israeli Experience with the “Green Pass” Policy Highlights Issues to Be Considered by Policymakers in Other Countries” by Ruth Waitzberg, ORCID, Noa Triki, Sharon Alroy-Preis, Tomer Lotan, Liat Shiran and Nachman Ash, 2021
- Fabiana Zollo, Michela Del Vicario, Riley Dunlap, Walter Quattrociocchi, Antonio Scala4, “The Climate Change debate on Facebook: A clash of contrasting narratives beyond rational arguments”