



*[CT0540] Social Network Analysis*

*Year: 2023/2024*

# Il tema del nucleare su Twitter

Lab Project

---

*De Biasi Andrea 890488*  
*Silvello Lorenzo 890441*

## Index

<b>1</b>	<b>Introduzione</b>	<b>1</b>
<b>2</b>	<b>Organizzazione</b>	<b>2</b>
2.1	Tecnologie utilizzate . . . . .	2
2.2	Divisione del lavoro . . . . .	2
<b>3</b>	<b>Scelta della fonte dei dati</b>	<b>3</b>
<b>4</b>	<b>Raccolta e restrizione dei dati</b>	<b>3</b>
<b>5</b>	<b>Analisi dei dati</b>	<b>5</b>
5.1	Rete dei retweet . . . . .	5
5.1.1	Osservazioni finali . . . . .	11
5.2	User mentions and top hashtags . . . . .	11
5.2.1	Conclusioni . . . . .	17
5.3	Word Frequency . . . . .	18
5.3.1	Osservazioni finali . . . . .	21
5.4	Sentiment analysis . . . . .	21
5.5	Topic modeling . . . . .	23
5.5.1	Osservazioni finali . . . . .	25
<b>6</b>	<b>Biases e limiti della ricerca</b>	<b>26</b>
<b>7</b>	<b>Conclusioni</b>	<b>27</b>

---

## 1 Introduzione

In un'epoca in cui il dibattito sul nucleare emerge come un elemento cruciale nel panorama globale, è essenziale analizzare le conversazioni online per comprendere le diverse prospettive e opinioni in merito. Il focus della presente ricerca si concentra sull'analisi di un dataset di tweet raccolti durante l'arco di un intero mese, con l'obiettivo di esplorare come le persone discutano del nucleare, sia in relazione all'aspetto bellico che alla produzione di energia.

Il nucleare rappresenta una tematica complessa e multiforme, coinvolgendo questioni che spaziano dalla sicurezza mondiale alla sostenibilità energetica. E' fondamentale comprendere come le piattaforme di social media, in particolare Twitter, siano veicoli di discussione e dibattito su questo tema.

Attraverso l'analisi del dataset di tweet, ci proponiamo di rispondere a diverse domande chiave:

*Quali sono le parole chiave e i termini più frequentemente utilizzati quando si discute del nucleare?*

*Quali sono le lingue utilizzate che generano il maggior numero di tweet sul tema nucleare?*

*La politica ha un ruolo nell'opinione pubblica sul tema del nucleare ?*

*Quali sono i temi che vengono affrontati quando si parla di nucleare ?*

*Quali sono i sentimenti che emergono nei tweet ?*

## 2 Organizzazione

### 2.1 Tecnologie utilizzate

Per la nostra ricerca abbiamo utilizzato le seguenti tecnologie:

- ◊ R Studio;
- ◊ Excel.

Utilizzati per la sezione di Sentiment Analysis e Topic modeling nei tweet. Formato dei file utilizzati per queste 2 analisi è il *csv*, per le restanti analisi è *json*.

- ◊ Google Doc;
- ◊ Overleaf.

Utilizzati per la scrittura della relazione. Questa relazione è stata scritta in Overleaf utilizzando il linguaggio latex. Come template è stato utilizzato un modello creato da un studente Dottorando di Ca' Foscari, solitamente utilizzato per scrivere le tesi di ricerca o lauree citetemplate.

- ◊ Python;
- ◊ Pycharm;
- ◊ Github;
- ◊ Jupyter Notebook.

Utilizzati per tutta l'analisi e la restrizione del dataset. Jupyter notebook è possibile dividere il codice in sezioni ed eseguirle singolarmente.

### 2.2 Divisione del lavoro

Il carico di lavoro è stato suddiviso equamente tra i 2 membri del gruppo, differenziandosi però nel compimento di alcuni compiti definiti inizialmente. Alcune attività sono state svolte collaborando insieme.

Definiamo dunque le attività svolte da *890488*:

- ◊ Word Frequency
- ◊ User Mentions e top hashtags
- ◊ Rete dei retweet

Attività svolte da *890441*:

- ◊ Sentiment analysis
- ◊ Topic modeling

Attività svolte insieme:

- ◊ Entrambi i componenti del gruppo si sono dedicati al download e alla restrizione del dataset. Il dataset non filtrato è stato diviso in 2 parti per poi assemblarlo a fine della restrizione.
- ◊ Stesura della relazione: per la stesura dell'analisi e le conclusioni il lavoro è stato svolto assieme confrontandosi sui risultati ottenuti.

### 3 Scelta della fonte dei dati

L'utilizzo di Twitter come fonte di dati permetterà di accedere a una vasta gamma di opinioni e prospettive, dando così vita a un quadro completo delle dinamiche di discussione riguardanti il nucleare. Questa ricerca mira a contribuire alla comprensione approfondita di come il pubblico online percepisca e discuta questioni legate al nucleare, con potenziali implicazioni per la sensibilizzazione pubblica e la formulazione delle politiche.

Twitter è una piattaforma di micro-blogging che consente di comunicare attraverso brevi messaggi, foto e video. A differenza di altre piattaforme, come Facebook, Twitter è caratterizzato da una maggiore immediatezza, spontaneità e interattività.

Queste proprietà permettono lo scambio di idee in maniera più rapida ponendo così le persone a discutere e confrontarsi.

Twitter è quindi una piattaforma che favorisce anche il dibattito e la partecipazione politica, sia da parte dei cittadini che dei politici. Molti leader politici usano infatti Twitter per comunicare con i loro elettori, per diffondere i loro messaggi e per creare la loro immagine pubblica.[3] Questa peculiarità è molto importante in quanto il tema della domanda di ricerca è strettamente correlato al tema politico.

### 4 Raccolta e restrizione dei dati

La parte più dispendiosa in termini di tempo per questa ricerca è stata la restrizione del dataset. I dati utilizzati nella nostra ricerca sono stati presi da un archivio online e risalgono a novembre 2022 [5]. Il dataset è stato ripulito da questo sito web e non direttamente da Twitter, in quanto nei mesi successivi tutti i servizi ai programmatori forniti da Twitter sono stati annullati.

Il dataset è suddiviso nei vari giorni del mese e per ogni giorno sono presenti 1440 file json, uno per minuto.

Per le varie analisi i dataset dovevano essere caricati in memoria ram, ma non siamo in possesso di macchine abbastanza potenti per supportare un carico che va oltre i 32 GB di dati (peso di un dataset di un giorno).

Un file json conteneva tutti i dati dei vari tweet, il quale ognuno aveva a sua volta: il nome dell'utente, la sua immagine profilo, i vari hashtag e altre informazioni più o meno importanti per la nostra domanda di ricerca. E' stato dunque compiuto inizialmente l'eliminazione dei campi considerati inutili o inconsistenti. Come primo passaggio, quindi, per alleggerire il dataset abbiamo rimosso tutti questi dati superflui e unito i 1440 file di una giornata intera in un unico file così da ottenere un file per giorno, ottenendo file con un peso medio di 2GB.

Dopo aver fatto ciò, abbiamo filtrato i vari tweet selezionando solo i tweet che contenessero alcune parole chiavi da noi selezionate riguardo l'argomento di ricerca:

- ◇ radioattività
- ◇ fissione nucleare
- ◇ isotopi
- ◇ nucleare militare
- ◇ energia atomica
- ◇ rischi nucleari
- ◇ proliferazione nucleare



rappresentato. Alla lista delle lingue analizzate abbiamo poi aggiunto il francese, abbiamo scelto di aggiungere questa lingua perchè la Francia è una nazione molto attiva nel settore del nucleare ed energie rinnovabili.[2]

Per analizzare un utenza più vasta abbiamo aggiunto allo studio altre due lingue: il tedesco e l'italiano.

La scelta dell'aggiunta del tedesco è stata influenzata da alcune notizie di cronaca in cui si parla dello spegnimento delle ultime 3 centrali nucleari e volevamo analizzare se nel mese da noi scelto vi fossero tweet a riguardo e le emozioni che questi suscitavano.[1][4]

Abbiamo deciso poi di aggiungere la lingua italiana per analizzare come nella nostra nazione venisse affrontata la tematica del nucleare.

**Note:**

Bisogna specificare che i tweet in una specifica lingua non implica il fatto che gli utenti possiedano quella nazionalità. Vengono analizzate solo le lingue con le quali è stato scritto il testo, non la locazione degli utenti, ciò può causare un bias nell'analisi delle notizie.

## 5 Analisi dei dati

L'analisi dei dati è stata effettuata in Python utilizzando come IDE PyCharm e con l'utilizzo di jupyter notebook per suddividere più facilmente le sezioni. Come descritto in precedenza l'analisi è stata ristretta a 6 lingue e per ciascuna vengono fatte le seguenti analisi:

- ◇ Rete dei retweet
- ◇ Analisi dell'argomento
- ◇ Word frequency
- ◇ Rete delle menzioni di utenti
- ◇ Sentiment Analysis

### 5.1 Rete dei retweet

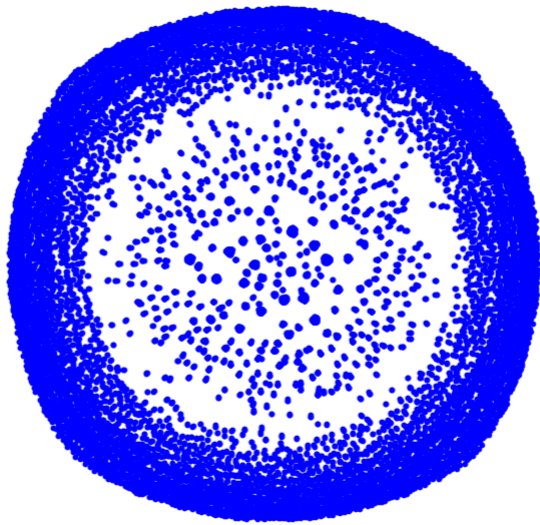
In questa sezione andiamo a creare una rete di retweet, una per ognuna delle 6 lingue selezionate. Abbiamo deciso nella nostra ricerca di analizzare i retweet in quanto sono una delle principali iterazioni che avvengono in questa piattaforma e si tratta di un post che viene creato in risposta ad un altro tweet

Creiamo con questa analisi dei grafici che rappresentano quali siano i post più retweettati. Nel grafico i nodi sono rappresentati dall'id del tweet gli archi che li collegano indicano se il tweet con uno specifico id è retweet di un altro.

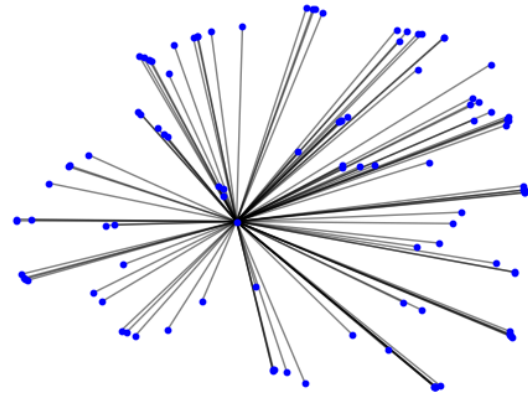
Per ciascuna lingua poi analizziamo il rapporto percentuale che c'è tra le componenti connesse e il numero totale di nodi per trovare se i post sono principalmente originali dell'utente o sono retweet di altri post. Grazie a questo riusciamo a capire il tipo di coinvolgimento che ha l'utenza sulla piattaforma.

**Note:**

Cliccando sul testo dei tweet si può accedere al loro link



(a) Rete dei retweet inglesi



(b) Componente connessa più grande

Figura 2: Rete dei retweet in Inglese, con la sua più grande componente connessa

La rete dei retweet inglese, in Figura 2.a è quella con più nodi rispetto a tutte le altre e presenta 14209 nodi e 7844 componenti connesse. Il rapporto percentuale per la lingua inglese tra componenti connesse e numero di nodi è del 55%.

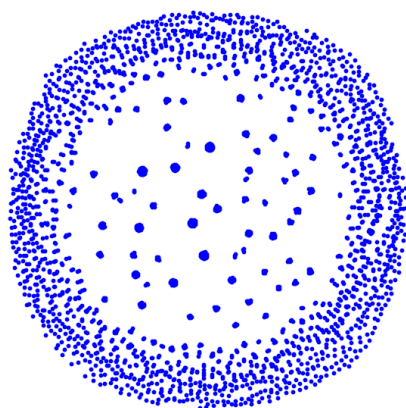
Possiamo notare che in questa rete come in quella spagnola abbiamo molte componenti connesse al centro e molti nodi singoli all'esterno del grafo; tale dato rappresenta che il tasso di pubblicazione dei tweet è più alto rispetto all'azione di retweet.

La Figura 2.b rappresenta la componente connessa più grande che comprende 107 nodi (corrispondente quindi a 106 retweet del post nel mese da noi selezionato) e riguarda il post dell'utente *mmpadellan* che scrive:

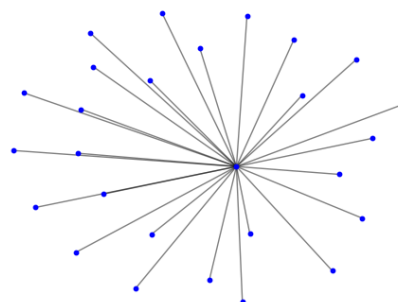
*"Hunter Biden is not employed by the government, did not steal classified nuclear documents or storm the US Capitol, but THAT'S Republicans' priority right now. Insanity."*

Ed è un post a difesa del presidente Biden, accusato di aver rubato documenti classificati sul tema del nucleare.





(a) Rete dei retweet spagnoli



(b) Componente connessa più grande

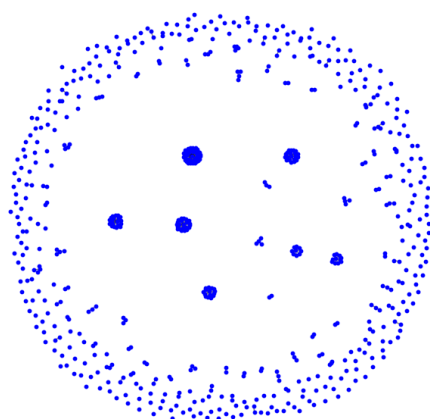
Figura 3: Rete dei retweet in Spagnolo, con la sua più grande componente connessa

La rete dei retweet spagnola nella Figura 3.a, seconda per numero di nodi, possiede 2482 nodi e 1306 componenti connesse. Il rapporto percentuale tra componenti connesse e numero di nodi è del 53%.

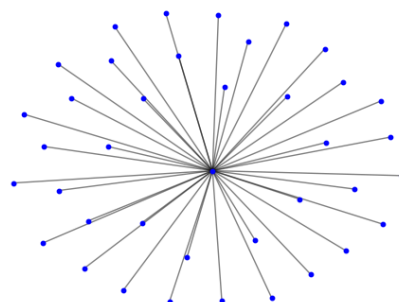
La Figura 3.b rappresenta la componente connessa più grande, comprende 27 nodi (26 retweet), e rappresenta la rete dei retweet il quale post originale è scritto da *David\_qva*:

*"NO LO DIRÁ CNN!! ...Ucrania "bombardeó el edificio especial No 2 de la Central Nuclear de Zaporizhzhya, donde se almacena combustible nuclear" informó GazetaRu... y aunque aseguran los niveles de radiación son estables, imágenes muestran jjescajes del agua de enfriamiento!!"*

Questo tweet riguarda un bombardamento da parte dell'ucraina ad un apparente deposito di combustibili nucleari.



(a) Rete dei retweet portoghesi



(b) Componente connessa più grande

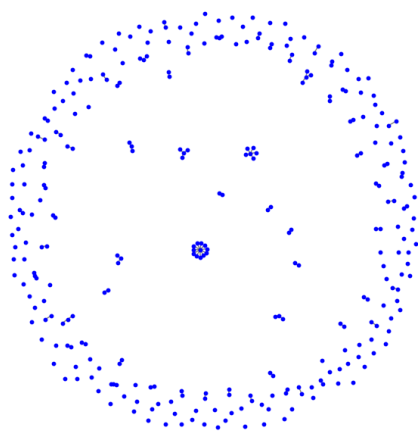
Figura 4: Rete dei retweet in Portoghese, con la sua più grande componente connessa

Nell'immagine 4.a viene rappresentata un'analisi della rete dei retweet portoghese, la quale possiede 624 nodi e 383 componenti connesse. Il rapporto tra il numero di componenti connesse e il numero di nodi in percentuale è del 61%.

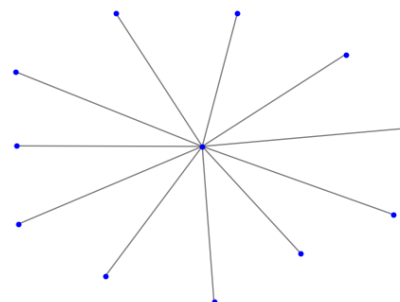
Analizzando la componente connessa più grande in Figura 4.b notiamo che comprende 39 nodi ed il nodo che connette la rete è rappresentato dal post scritto da *HistoriaNoPaint*:

*"Urgente: O cantor Manoel Gomes (o caneta azul) acaba de desenvolver armas termonucleares e diz que em caso de ataque da Coreia do Norte ao Brasil, irá revidar."*

Il tweet sopra citato rappresenta un post di satira.



(a) Rete dei retweet tedeschi



(b) Componente connessa più grande

Figura 5: Rete dei retweet in Tedesco, con la sua più grande componente connessa

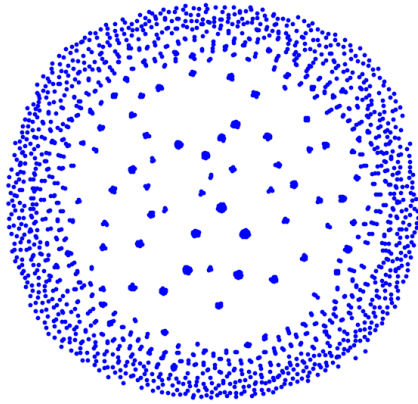
La rete tedesca, in Figura 5.a possiede invece 302 nodi e 214 componenti connesse. Il loro rapporto è del 71% il quale indica che per la maggior parte dei tweet non c'è stato retweet, analizzeremo meglio questo dettaglio comune tra i grafi nelle osservazioni finali.

Notiamo anche in questo caso molti nodi singoli senza archi ed al centro componenti connesse di dimensioni simili.

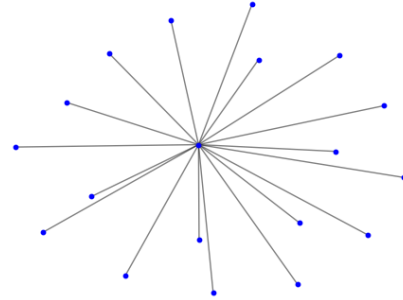
Prendiamo anche stavolta in considerazione la sottorete più grande, Figura 5.b, che comprende solamente 12 nodi ed è rappresentata dal tweet di *EL HOTZO*:

*"zu spät geboren um jemals ein Haus zu besitzen  
zu früh geboren um durch die postnukleare Wüste zu wandern  
exakt zur richtigen Zeit geboren, um mir von 3-5 Social Media-Apps das Gehirn so zerlöchern  
zu lassen bis ich keine zwei zusammenhängenden Gedanken mehr haben kann"*

Nel tweet viene evidenziata l'influenza negativa che hanno i social sui pensieri.



(a) Rete dei retweet francesi



(b) Componente connessa più grande

Figura 6: Rete dei retweet in Francese, con la sua più grande componente connessa

In questo paragrafo analizziamo la rete francese, Figura 6.a che possiede 1810 nodi e 1040 componenti connesse (notiamo come il nucleare in francia sia un argomento molto discusso sui social). Il loro rapporto è del 57%.

Il sottografo più grande, Figura 6.b, descritto dal tweet di *Florian Philippot* (politico francese ed ex esponente del Fronte Nazionale), possiede 19 nodi:

*"L'AIEA sait que ce sont les Ukrainiens qui bombardent la centrale nucléaire de Zaporijia !  
Voilà où finissent les armes qu'on donne à Zelensky : en potentiel nouveau Tchernoby!  
Arrêtons le carnage: plus une arme, plus un euro pour Zelensky !"*

Il tweet esprime dissenso per la guerra in Ucraina con il timore che possa accadere un disastro simile a Chernobyl.

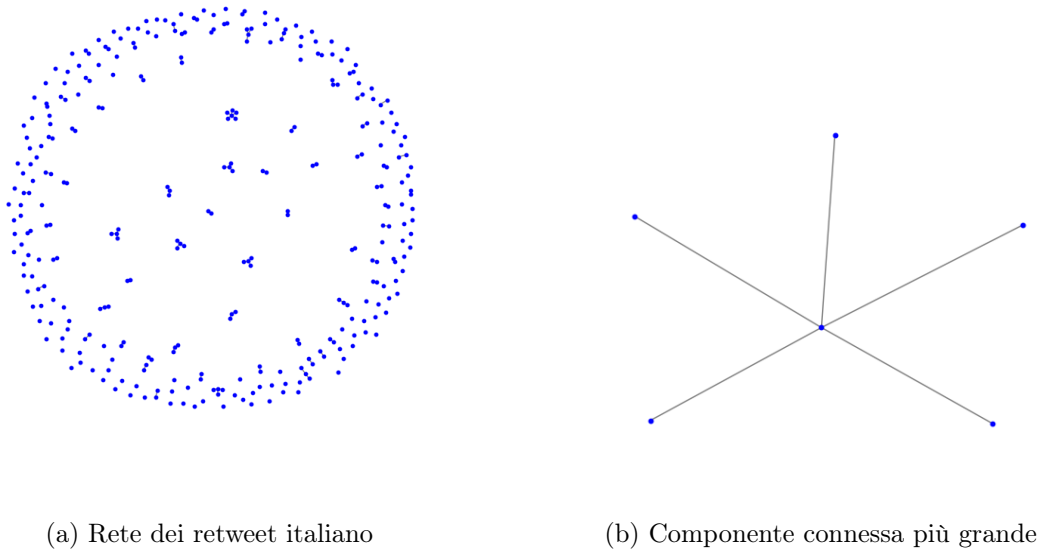


Figura 7: Rete dei retweet in Italiano, con la sua più grande componente connessa

Il grafo della rete italiana in Figura 7.a, possiede 321 nodi e 225 componenti connesse (notiamo come il nucleare in francia sia un argomento molto discusso sui social). I due dati hanno un rapporto del 70%.

Il sottografo più grande in Figura7.b, è rappresentato dalla rete del tweet di *Gianluca Martino* e possiede 6 nodi:

*"Chernobyl 26 aprile 1986. Mentre il mondo guarda attonito, dalle TV del tinello di casa, l'immane disastro che colpisce soprattutto l'Ucraina, Cuba organizza una gigantesca operazione di soccorso che sarà ricordata come la più lunga della storia (1990/2011) ..."*

L'argomento del tweet è la votazione per l'abolizione dell'embargo su Cuba.

### 5.1.1 Osservazioni finali

I temi trattati nei tweet più importanti sono vari ma comunque coerenti con il tema della ricerca. Un dettaglio importante da notare è la presenza dei politici tra i tweet più retweetati, questa informazione è in linea alle considerazioni iniziali in merito alla scelta della fonte di dati.

Dopo aver analizzato i grafi delle 6 lingue possiamo vedere come il rapporto tra le componenti connesse e il numero dei nodi sia sempre superiore al 50%. Questo dato indica che gli utenti quando parlano di argomenti legati alla tematica studiata preferiscano creare dei tweet piuttosto che creare un retweet. Questo crea con maggiore probabilità discussioni su più argomenti e quindi una vasta lista di topic.

## 5.2 User mentions and top hashtags

Su twitter oltre ai retweet vi è un'altra iterazione molto utilizzata, quella delle menzioni: una menzione è quando si tagga un utente o una pagina all'interno di un tweet per far sì che l'informazione raggiunga quel determinato utente.

Vengono poi utilizzati nei tweet gli hashtags: sono dei tag che identificano il contenuto dei tweet o il topic di cui parla, se si clicca su questi hashtags si viene reindirizzati a tutti i vari post che utilizzano quello stesso hashtag.

In questa sezione analizziamo per ogni lingua quali sono le persone o pagine più menzionate all'interno dei tweet per ogni lingua e andiamo ad analizzare gli hashtag più utilizzati per descrivere i tweet.

**Note:**

I nodi rappresentano gli utenti, sia quelli che scrivono i post che quelli menzionati.

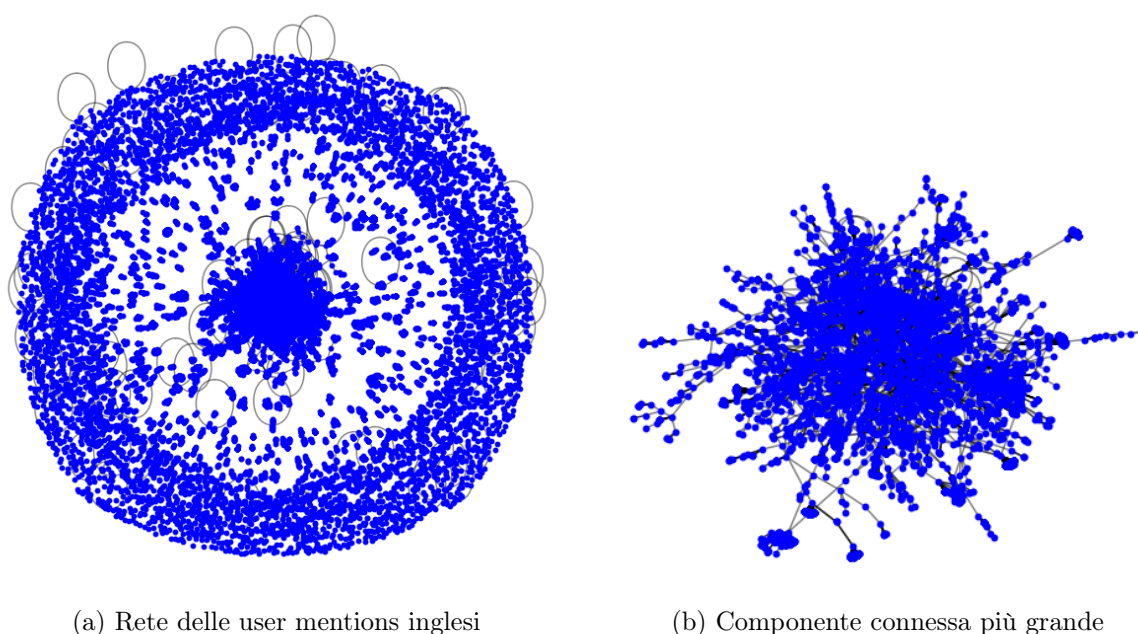


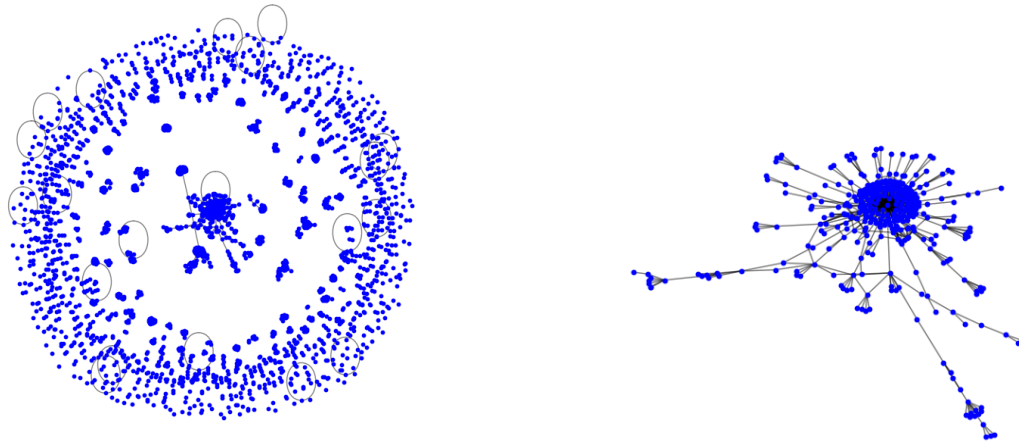
Figura 8: Rete delle menzioni in Inglese, con la sua più grande componente connessa

La Figura 8.a rappresenta la rete delle user mentions inglesi sono rappresentati 18548 nodi (utenti) e 4242 componenti connesse. Al contrario del grafico dei retweet stavolta vediamo che vi è una componente connessa principale con molti nodi (Figura 8.b).

L'utente più menzionato nella lingua inglese, cioè quello che crea la componente connessa centrale (menzionato 250 volte), è *Elon Musk* che sappiamo essere una figura molto influente nella comunità scientifica e tecnologica.

Le 5 hashtag più usate nella lingua inglese sono: "nuclear" usata 170 volte, "nftgiveaway" viene usato come hashtag 94 volte, è un dato che non è coerente con la nostra ricerca ma è rimasto all'interno del nostro dataset filterato. Questo probabilmente a causa di un bot che postato nel mese da noi selezionato dei tweet con le parole chiave da noi scelte per il filtraggio.

Vengono usati poi gli hashtag "russia" e "ukraine" (rispettivamente 89 e 77 volte; e come ultimo nella lista l'hashtag "breaking" (usato 53 volte) il quale viene utilizzato una notizia recente importante.



(a) Rete delle user mentions spagnole

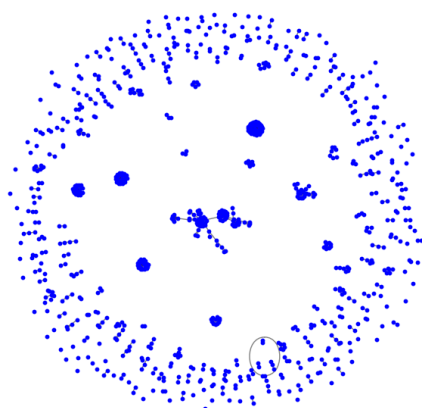
(b) Componente connessa più grande

Figura 9: Rete delle menzioni in Spagnolo, con la sua più grande componente connessa

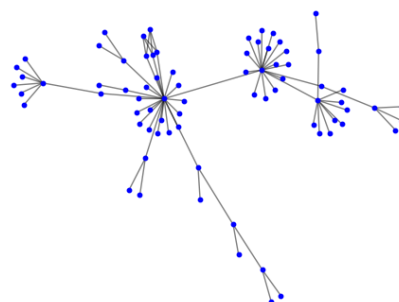
Nella rete delle menzioni spagnola in Figura 9.a, sono rappresentati 2775 nodi e 798 componenti connesse. Anche in questo caso notiamo che è presente un sottografo più denso al centro del grafo (Figura 9.b).

Questo sottografo è rappresentato dalla pagina "Operador Nuclear", taggata 219 volte. La pagina menzionata appartiene ad *Alfredo Garcia*, ingegnere con licenza di Operatore e Supervisore di reattori nucleari, ed è una pagina che tratta energia nucleare e divulgazione scientifica.

Nei post in lingua spagnola la lista dei 5 hashtag più utilizzati è la seguente: "rusia" (usata 19 volte), "ultimahora" ed "ahora" e come per l'hashtag "breaking" inglese vengono utilizzate per indicare notizie recenti che possono essere di carattere sia politico che sociale. Il quarto hashtag usato è "ucraina", usato 17 volte; come ultimo hashtag invece troviamo "apuntesoperador" (usato 13 volte) che viene utilizzato maggiormente da pagine di carattere scientifico le quali trattano di energia nucleare e informazione sul tema della radioattività.



(a) Rete delle user mentions portoghesi



(b) Componente connessa più grande

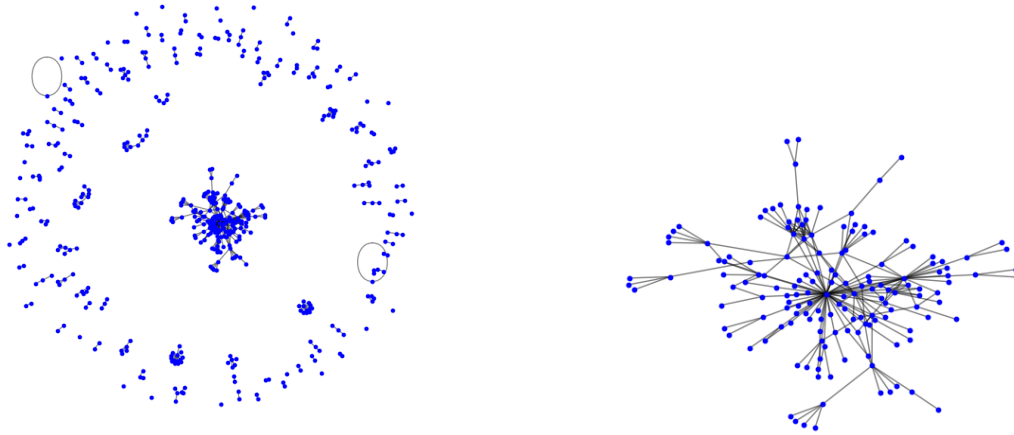
Figura 10: Rete delle menzioni in Portoghese, con la sua più grande componente connessa

Per la lingua portoghese troviamo una rete delle menzioni composta da 847 nodi e 312 componenti connessi di cui solo una con dimensione maggiore (Figura 10.a e 10.b).

L'utente rappresentante il nodo centrale è *História No Paint* ed ha ricevuto 38 menzioni. Quella citata è una pagina di satira di carattere scientifico, che è stata quella che ha creato la rete di retweet di grandezza maggiore per la lingua portoghese.

Non abbiamo svolto l'analisi dei 5 hashtag più utilizzati per la lingua portoghese in quanto nel dataset per il mese da noi selezionato questa funzione è stata poco utilizzata e l'analisi che avremmo avuto come risultato non sarebbe stata abbastanza accurata e coerente con la nostra ricerca.





(a) Rete delle user mentions tedesche

(b) Componente connessa più grande

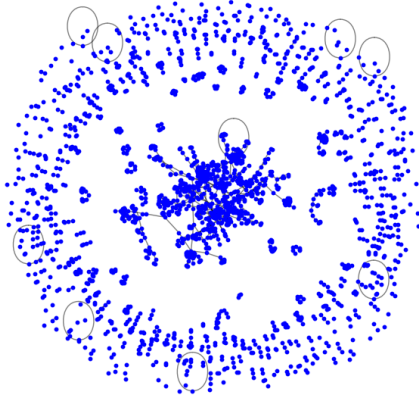
Figura 11: Rete delle menzioni in Tedesco, con la sua più grande componente connessa

Per la lingua tedesca è stato analizzato un grafo composto da 474 nodi e 118 componenti connesse (Figura 11.a). Il nodo che lega il sottografo più grande (Figura 11.b) è rappresentato dalla pagina twitter *Nuklearia e. V.*, un'associazione senza scopo di lucro che promuove l'uso civile dell'energia nucleare. La pagina viene menzionata nei tweet 63 volte, essa definisce l'energia nucleare come una forma di approvvigionamento energetico pulito, un questione molto discussa in Germania.

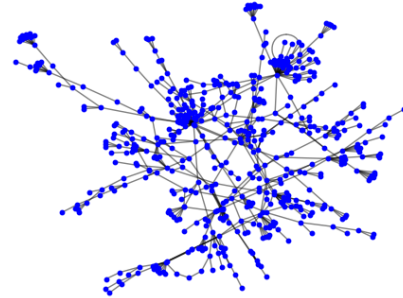
La lista dei 5 hashtag più utilizzati nei tweet in lingua tedesca ha come primo elemento *twitter-layoffs*, purtroppo non è un dato utile agli scopi della ricerca perchè utilizzato contro *Elon Musk* dopo l'acquisizione di Twitter.

La lista però continua con gli hashtag "ukraine" e "tschernobyl" i quali sono collegati con la ricerca e con la tematica della guerra in ucraina e la preoccupazione per le centrali nucleari presenti nella zona del conflitto.

Un altro hashtag utilizzato è "energiekrise" sta ad indicare la crisi energetica, argomento molto discusso e principalmente correlato con la produzione di energia nucleare. Quest'ultimo viene usato anche assieme all'hashtag "fridaysforfuture", movimento internazionale di protesta che hanno come fine la prevenzione del riscaldamento globale e il cambiamento climatico.



(a) Rete delle user mentions francesi



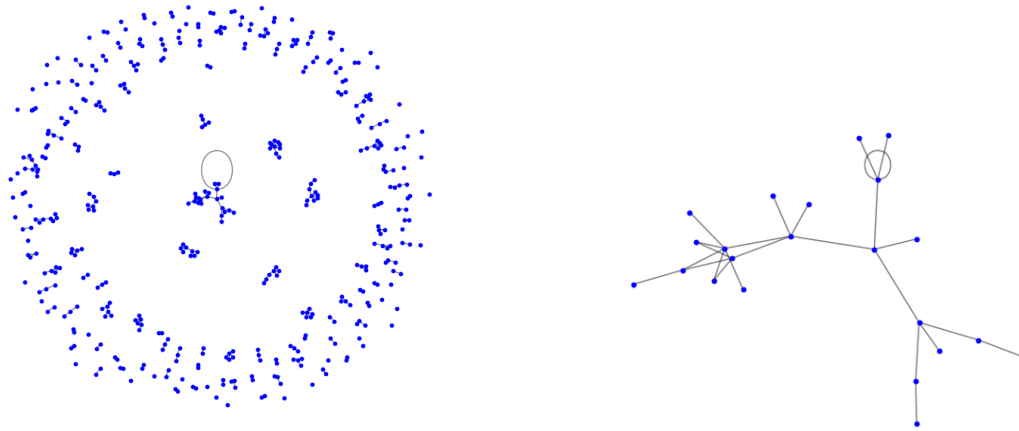
(b) Componente connessa più grande

Figura 12: Rete delle menzioni in Francese, con la sua più grande componente connessa

La rete delle menzioni per la lingua francese è composta da 2104 nodi e 513 sottografi (Figura 12.a). La componente connessa più grande (Figura 12.b) è rappresentata dalla menzione dell'utente *Florian Philippot* (politico francese, ex esponente del *Fronte Nazionale* e deputato europeo) che viene menzionato nei tweet 39 volte.

In lingua francese la lista dei 5 hashtag più utilizzati è composta come segue: al primo posto con 43 utilizzi vi è "nucléaire", al secondo e terzo posto vi sono "zaporijia" e "ukraine"; Zaporizžja è la città dove è situata una delle centrali che, citando i media locali, è stato un luogo preso di mira dall'esercito russo.

Come ulteriori hashtag vi sono "Pologne" che sta ad indicare la Polonia e viene usata in quanto in quel periodo iniziarono dei trattati per la costruzione della prima centrale nucleare polacca grazie all'appoggio USA. Come ultimo hashtag vi è "edf" ed è la sigla di "Électricité de France S.A.", la quale è la maggiore azienda produttrice e distributrice di energia in Francia.



(a) Rete delle user mentions italiane

(b) Componente connessa più grande

Figura 13: Rete delle menzioni in Italiano, con la sua più grande componente connessa

Come ultima, la rete italiana in Figura 13.a possiede 443 suddivisi che creano 176 sottografi . La componente connessa più grande (Figura 13.b) vede al centro l'utente con più menzioni in Italia è *Matteo Salvini*, politico italiano attivo nei movimenti sociali pro-nucleare.

Gli hashtag in Italia per i dati da noi selezionati sono stati poco utilizzati ma i 5 migliori rimangono comunque conformi con l'argomento di ricerca e sono: "nucleare"; "zaporizhzhia" che è la traduzione del nome della città ucraina citata nella ricerca per la lingua francese "Zaporizžja"; "italia", "ucraina" e, l'ultimo hashtag presente nella lista è "putin" usato sempre in tweet in cui si parla della paura di una guerra nucleare.

### 5.2.1 Conclusioni

Analizzando la rete delle user mentions si può osservare che alcune delle pagine o utenti più menzionate all'interno dei tweet sono le stesse che creavano i post centrali per la rete dei retweet nella lingua stessa, questo indica che sono pagine o utenti attivi riguardo queste tematiche che hanno una community più larga di altre pagine.

I risultati ottenuti da questa analisi sono conformi con le domande della nostra ricerca e le pagine menzionate approfondiscono l'ambito del nucleare principalmente nel tema di energia e guerra.

Gli hashtag più utilizzati sono "nucleare" e "ucraina" riguardo il conflitto presente ancora oggi tra Russia e Ucraina; mentre "apunesoperador" e "edf" sono stati gli hashtag più utilizzati riguardo il nucleare come fonte di energia.

In generale, prendendo in considerazione l'analisi condotta sugli hashtag, risulta essere più trattato il tema della guerra e la paura di un conflitto nucleare, ma approfondiremo la tematica delle emozioni nella sezione 'Sentiment Analysis'.





Figura 15: Word Frequency della lingua spagnola

Nei tweet di lingua spagnola, come si vede da Figura 15, la parola più utilizzata è "ucraina", seguite dalle seguenti: "rusia", "armas", "operadornuclear", "guerra", "energía", "central", "mundo", "eeuu", "corea", "chernobyl", "norte", "ataque", "jimin", "europa", "china".

Per la lingua spagnola vi sono tweet e parole che non sono correlati con la ricerca come "jimin" che è il nome di un cantante.

L'argomento più discusso rimane però la guerra e il timore possa accadere un disastro nucleare come quello accaduto a Chernobyl.



Figura 16: Word Frequency della lingua portoghese

Dalla Figura 16 vediamo che la parola più citata nei tweet in lingua portoghese è "chernobyl", viene utilizzata in quanto nei tweet, come si evince dalla sentiment analysis, vi è la paura che possa accadere un disastro nucleare causato dalla attuale guerra in Ucraina.

La lista completa è la seguente: "chernobyl", "grafite", "armas", "nucleares", "acaba", "desenvolver", "vai", "caneta", "azul", "termonucleares", "planaito", "caso", "bomba", "historiano-paint".

Anche per la lingua portoghese vengono utilizzati termini legati all'energia nucleare come ad esempio "azul" che tradotto vuol dire "blu" che è il colore utilizzato per indicare l'energia nucleare.



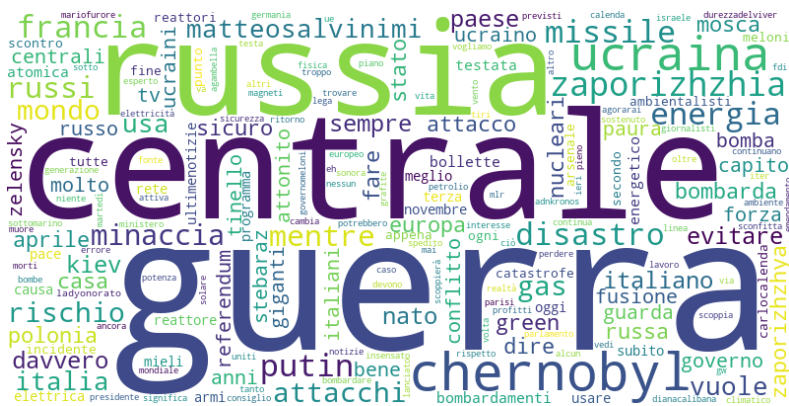


Figura 19: Word Frequency della lingua italiana

Guardando la Figura 19 possiamo notare che anche per la lingua italiana il tema principale dei tweet rimane la guerra.

Le parole più utilizzate sono rappresentate nella seguente lista: "centrale", "russia", "chernobyl", "ucraina", "zaporizhzhia", "energia", "putin", "missile", "francia", "disastro", "russi", "minaccia".

Oltre al tema della guerra è comunque citata la parola "eneriga" che è spesso affiancata a "francia". Infatti vi sono tweet in cui si discute contro la scelta di non utilizzare il nucleare in Italia come fonte di energia ma di supportarla comunque comprandola da altri paesi.

### 5.3.1 Osservazioni finali

Per rispondere alla domanda iniziale i termini più utilizzati sono legati principalmente alla guerra che è ancora attualmente in corso tra Ucraina e Russia in cui in territori legati al conflitto sono presenti centrali nucleari.

Come parole chiavi vi sono comunque termini che creano una discussione riguardo le energie rinnovabili e il nucleare come una di esse. In Germania ad esempio si è iniziato a parlare nel mese di novembre di petizioni a sostegno del mantenimento in funzione delle centrali nucleari.

## 5.4 Sentiment analysis

Ora vogliamo capire quali sono i sentimenti che emergono nei tweet postati dagli utenti e per far ciò abbiamo svolto la Sentiment Analysis.

La Sentiment Analysis è un campo dell'elaborazione del linguaggio naturale che si occupa di costruire sistemi per l'identificazione ed estrazione di opinioni dal testo. Abbiamo preso in considerazione 8 emozioni: *rabbia*, *anticipazione*, *disgusto*, *paura*, *gioia*, *tristezza*, *sorpresa* e *fiducia*. Per svolgere questa analisi è stata utilizzata la libreria *Syuzhet*, un pacchetto R che permette di estrarre il sentimento e la trama di un testo narrativo, usando diverse fonti di dati.[7] Di seguito sono riportate gli istogrammi delle analisi per le lingue prese in considerazione.



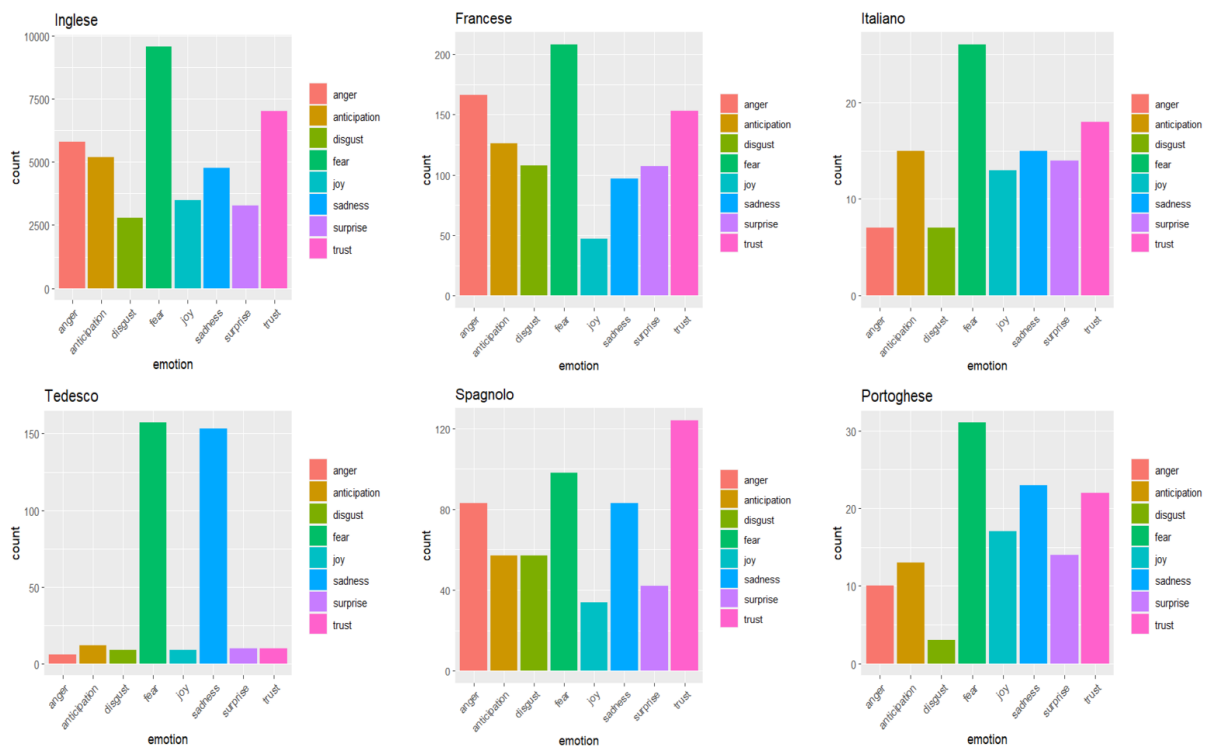


Figura 20: Sentiment Analysis delle lingue prese in considerazione



Possiamo notare in Figura 20 come nei vari istogrammi la distribuzione delle emozioni risulta non simile e questo ci porta a considerare come le emozioni che vengono riportate all'interno dei tweet cambia da un paese ( o insieme di paesi) a un altro.

Andando a stilare una classifica delle emozioni che prevalgono nei vari tweet possiamo notare che al primo posto abbiamo la paura, la seconda è la fiducia e terza l'anticipazione.

Se pensiamo alle emozioni che potremmo associare al nucleare possiamo dire che sono conformi alla classifica che abbiamo stilato precedentemente.

La paura è l'emozione più significativa ed è data dal fatto che generalmente il nucleare è visto come qualcosa di pericoloso, dal punto vista bellico ma anche dal punto di vista di uso per scopi di approvvigionamento energetico.

La fiducia tuttavia è l'emozione contrastante per certi versi alla paura però è importante sottolineare che è presente in modo prevalente in quasi tutti i paesi. Per quanto riguarda la lingua spagnola infatti è l'unica lingua in cui non primeggia la paura ma la fiducia.

Questa emozione possiamo includerla nella domanda di ricerca in quanto è un'emozione che viene provata da quelle persone che sono favorevoli all'approvvigionamento energetico tramite il nucleare .

Analizziamo un dato interessante, ovvero, la fiducia come emozione più diffusa nei tweet spagnoli. Notiamo che c'è una correlazione tra l'informazione precedentemente descritta e l'opinione pubblica spagnola in merito al tema della produzione di energia nucleare, sebbene sappiamo che i tweet di lingua spagnola non sono per forza stati scritti in Spagna.

Dunque nel caso specifico dei tweet in spagnolo, potrebbe esserci una maggiore fiducia nel nucleare per diversi motivi. Ad esempio, la Spagna ha una lunga esperienza nell'uso del nucleare per la produzione di elettricità, con otto reattori operativi dal 1981. Inoltre, la Spagna ha una forte dipendenza energetica dall'estero, il che potrebbe rendere il nucleare una fonte più attraente e affidabile.

Infine, la Spagna ha una cultura politica e sociale più aperta e tollerante verso il nucleare, rispetto ad altri paesi europei come la Germania o la Francia, dove ci sono state forti proteste e movimenti anti-nucleari.[9]

Valutiamo le emozioni meno presenti nei tweet ovvero: *gioia, sorpresa e disgusto*.

Le prime due emozioni difficilmente possiamo esprimerle nel parlare del nucleare, considerato che il tema della guerra può essere presente in molti tweet.

## 5.5 Topic modeling

In questa sezione andremo a descrivere un'altra analisi che mira a rispondere alla domanda:

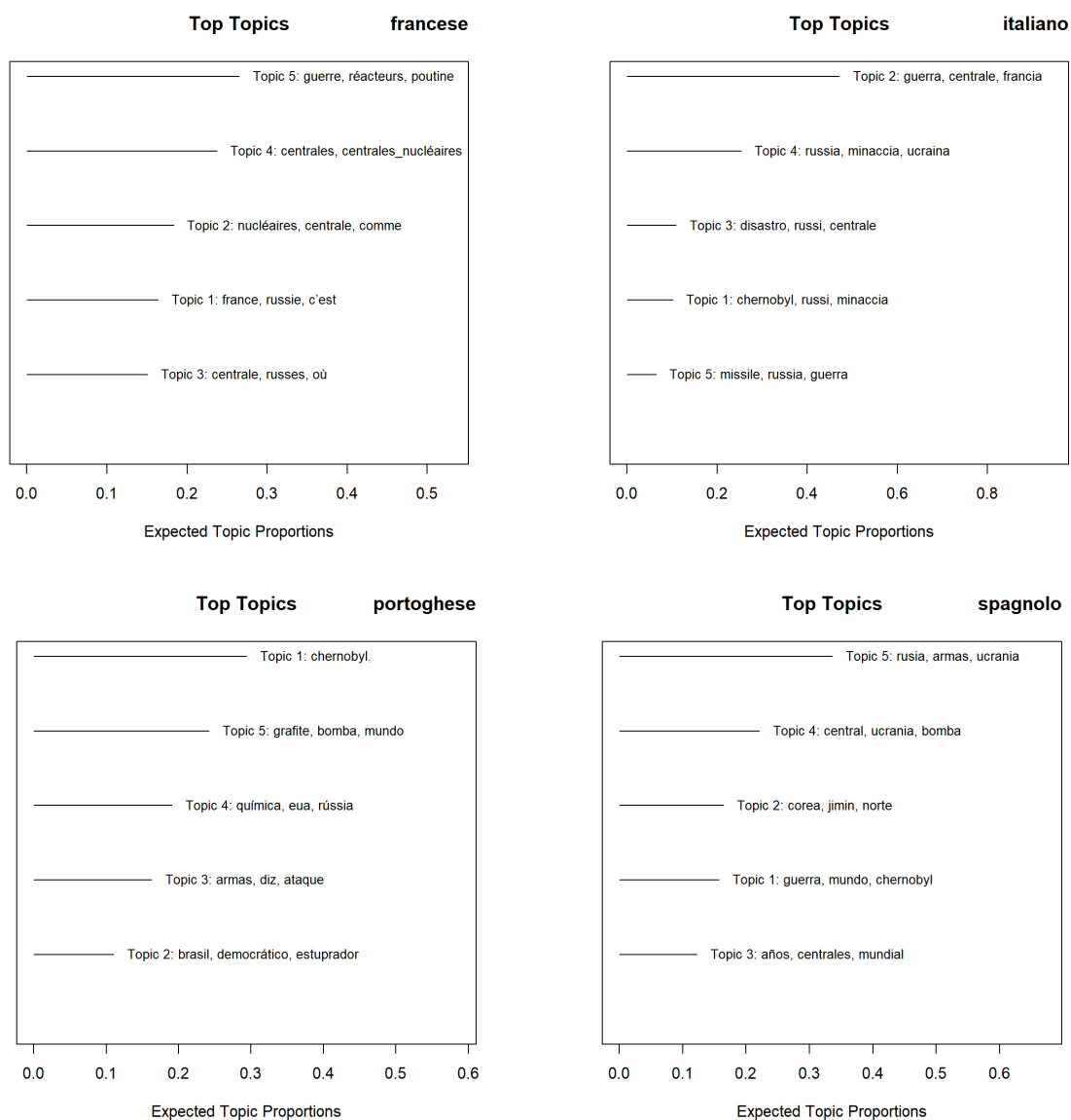
*Quali sono i temi che vengono affrontati quando si parla di nucleare?*

Per scoprire quali sono i temi principali, è stata usata una tecnica chiamata topic modeling, che raggruppa i documenti in base alle parole che contengono. Il topic modeling è un metodo che non richiede di conoscere i temi in anticipo, ma li trova direttamente dai dati. Ci sono diversi modelli di topic modeling, quella utilizzata in questa analisi è chiamata STM (Structural Topic Model).

Il modello dei topic strutturale ci permette di incorporare i metadati nel nostro modello e scoprire come diversi documenti possano parlare dello stesso tema sottostante usando parole diverse.

Per compiere questa analisi si è utilizzato la libreria di R *stm* [10], la quale, dato il testo filtrato dalle stop words, ha prodotto l'analisi desiderata.

Nella figura successiva, avremmo dunque i risultati osservati:



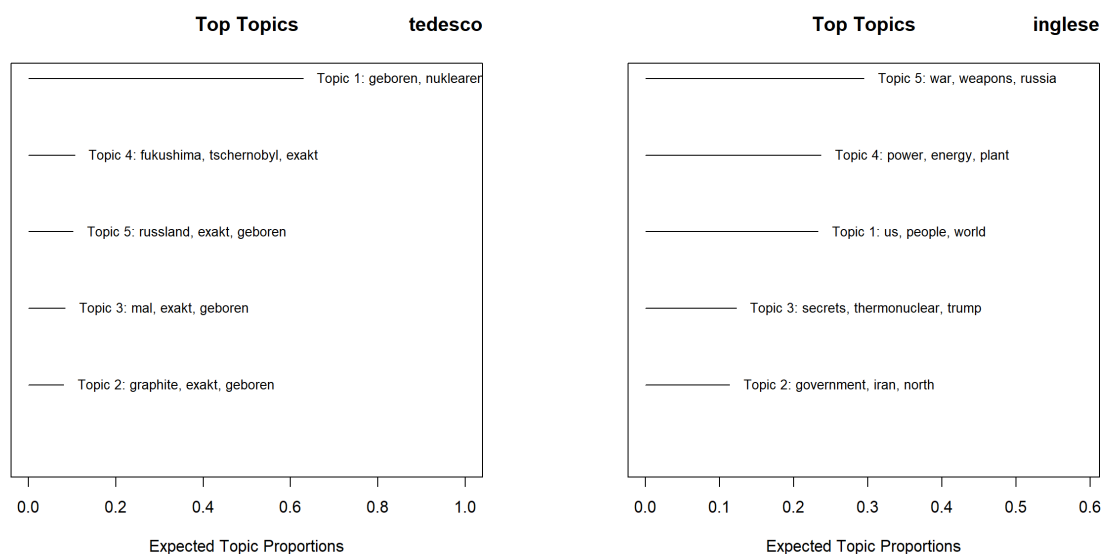


Figura 22: Topic modeling

### 5.5.1 Osservazioni finali

Nella Figura 22 sono comprese tutte le analisi sulle tematiche presenti nelle raccolta dei tweet, nelle varie lingue.

Il numero di topic chiesti in output è 5, questo numero è stato scelto dopo aver fatto alcune prove. In questi test abbiamo notato che con un numero di topic più alto il modello restituiva argomenti non strettamente collegati alla domanda di ricerca, mentre un numero troppo basso non descriveva sufficientemente il documento.

Si può notare, una volta aver tradotto le parole presenti nel grafico, che le tematiche più presenti sono principalmente:

- ◇ Guerra
- ◇ Centrale nucleare
- ◇ Russia
- ◇ Reattori
- ◇ Putin

Le tematiche risultanti sono quindi coerenti con le domande di ricerca.

## 6 Biases e limiti della ricerca

Alcuni errori potrebbero anche essere stati commessi durante il riconoscimento della lingua utilizzata nei tweet dove abbiamo utilizzato la libreria "langid" [8], questi errori possono essere dovuti alla lunghezza limitata del testo dei tweet.

Il riconoscimento della lingua dal testo crea un limite di ricerca per l'identificazione geografica del tweet in quanto lingue come inglese e spagnolo sono utilizzate globalmente e non solo dai paesi madrelingua.

La libreria utilizzata per la compiere la Sentiment analysis [7] potrebbe non essere stata precisa nel riconoscimento di certi emozioni, introducendo delle emozioni che non appartengono a un determinato tweet. Tuttavia il numero di errori dovrebbe essere trascurabile per le conclusioni che abbiamo tratto da questa analisi.

Un altro bias introdotto in fase di filtraggio del dataset potrebbe essere il non utilizzo di alcune parole chiave non identificate. In questo caso si potrebbe aver perso alcuni tweet che potevano risultare utili alla domanda di ricerca.

Il modello STM (Structural Topic Model), utilizzato, ha alcune limitazioni per il riconoscimento degli argomenti, come richiedere molte risorse e tempo computazionali, essere sensibile al numero di argomenti, produrre potenzialmente argomenti ridondanti e non catturare le sfumature semantiche e sintattiche del linguaggio naturale [6]. Questi problemi possono influire sulla qualità e sulla granularità degli argomenti, sulle prestazioni della classificazione e sul significato.

## 7 Conclusioni

L'analisi compiuta sul dataset di dati è risultata utile a capire come gli utenti di Twitter discutono sul tema del nucleare. Analizzando le lingue più utilizzate per la scrittura dei tweet ha fatto comprendere come il tema del nucleare sia largamente discusso in tutto il mondo.

La Rete di retweet e la User mentions ci ha aiutato a rispondere alla domanda iniziale:

*La politica ha un ruolo nell'opinione pubblica sul tema del nucleare?*

La presenza di politici come personaggi influenti su questa tematica è stata rilevata sia nella Rete dei retweet ma anche all'interno della User mentions. Possiamo dunque confermare che la politica e l'opinione pubblica sui social sono a stretto contatto.

Il tema dei tweet più con più retweet e gli hashtag utilizzati hanno principalmente il tema della guerra tra Russia e Ucraina mentre, il quale si evince anche nella analisi della Word Frequency. Il tema energetico è presente ma in maniera limitata. A confermare questo è stata l'analisi condotta 'Topic modeling' dove è stato riscontrato una presenza importante e ricorrente del tema della guerra per tutte le lingue.

In linea con ciò che è stato descritto precedentemente possiamo confermare che la paura è l'emozione più diffusa quando si parla di nucleare, data dai conflitti bellici tra Russia e Ucraina. Come analizzato nella sezione relativa alla Word Frequency si evince che vi è il timore di un'evoluzione della guerra che potrebbe portare ad un disastro come quello di Chernobyl.

## Riferimenti bibliografici

- [1] Redazione Bruxelles. “Berlino ascolta Greta: rinviato lo spegnimento delle centrali nucleari”. In: *Europa Today* (2022). URL: <https://europa.today.it/ambiente/berlino-ascolta-greta-rinviato-spegnimento-nucleare.html>.
- [2] Arianna Cerone. “La Francia punta sul nucleare: Otto nuovi reattori entro il 2050”. In: *Europa Today* (2024). URL: <https://europa.today.it/ambiente/energia-nuclare-francia-nuovi-reattori.html>.
- [3] Brian Dean. “Quante persone hanno usato Twitter nel 2022?” In: (2023). URL: <https://it.semrush.com/blog/statistiche-twitter/>.
- [4] Geopop. “Stop alle ultime 3 centrali nucleari in Germania dopo 60 anni, i motivi di questa decisione”. In: *Geopop* (2023). URL: <https://www.geopop.it/stop-alle-ultime-3-centrali-nucleari-in-germania-dopo-60-anni-i-motivi-di-questa-decisione/>.
- [5] hananc. *Archive Team Twitter Grabs 2022-11*. URL: <https://archive.org/details/archiveteam-twitter-stream-2022-11>.
- [6] Hongning Wang, Duo Zhang, ChengXiang Zhai. *Structural Topic Model for Latent Topical Structure Analysis*. URL: <https://aclanthology.org/P11-1153.pdf>.
- [7] Matthew Jockers. *Introduction to the Syuzhet Package*. URL: <https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html>.
- [8] Marco Lui. *langid.py Library*. URL: <https://github.com/saffsd/langid.py>.
- [9] Josep Espluga Trenc, Beatriz Medina, Albert Presas. “THE SOCIAL DIMENSIONS OF THE PERCEPTION OF NUCLEAR ENERGY. AN ANALYSIS OF THE SPANISH CASE (1960-2015)”. In: *Revista Internacional de Sociología* (2017). URL: <https://revintsociologia.revistas.csic.es/index.php/revintsociologia/article/download/688/874?inline=1>.
- [10] Brandon Stewart. *stm: Estimation of the Structural Topic Model*. URL: <https://cran.r-project.org/web/packages/stm/index.html>.