



Università degli Studi di Salerno
Dipartimento di Informatica

Corso di Reti Geografiche

Sentiment Analysis dell'Opinione Pubblica sul Conflitto a Gaza: Studio Comparativo su Reddit in Ambito Europeo

Gioacchino Tortorelli

Matricola: 0522501860

Contents

1	Introduzione	3
2	Architettura del Progetto	5
3	Dataset e Raccolta Dati	6
3.1	Introduzione alla Strategia di Raccolta	6
3.2	Selezione dei Subreddit Target	6
3.3	Architettura del Sistema di Raccolta	7
3.4	Processo di Pulizia e Filtraggio	7
3.5	Traduzione del Dataset	8
4	Fase 2 – Analisi Comparativa tra Nazioni	10
4.1	Preprocessing dei Dati	10
4.2	Engineering delle Feature	10
4.3	Metriche Comparative tra Nazioni	11
4.4	Risultati delle Analisi Comparative	11
4.5	Analisi dei Cluster Nazionali	12
5	Fase 3 – Analisi Tematica Transnazionale	15
5.1	Sistema di Categorizzazione Tematica	15
5.1.1	Architettura del Sistema di Classificazione	15
5.1.2	Definizione delle Categorie Tematiche	15
5.2	Distribuzione Globale dei Temi	16
5.3	Prevalenza dei Temi a Livello Globale	17
5.4	Profili Tematici per Nazione	18
6	Fase 4 - Sentiment Analysis	21
6.1	Modello e Metodologia	21
6.2	Risultati Globali	21
6.3	Analisi del Sentiment per Nazione	22
6.4	Analisi del Sentiment per Coppie Tema-Nazione	23
6.5	Clusterizzazione del Sentiment	24
6.6	Analisi della Polarizzazione e degli Estremi Emotivi	25
6.6.1	Analisi dei Commenti Estremi	26
6.7	Correlazioni Tematiche e Pattern Discorsivi	26
6.8	Analisi del Sentiment a Livello di Post	27
6.9	Analisi della Polarizzazione	27

6.9.1	Casi di Studio Dettagliati	28
7	Fase 5 - Analisi Temporale e Trend	31
7.1	Identificazione degli Eventi Significativi	32
7.2	Previsione dei Trend Futuri	32
7.3	Analisi degli Eventi Critici	34
7.4	Analisi dei Picchi di Volume per Nazione	35
7.5	Analisi Sincronia dei Picchi	37
7.6	Analisi Evoluzioni Tematiche per Tempo	43
8	Risultati e Conclusioni	45
8.1	Sintesi comparativa dei risultati	45
8.2	Interpretazioni socio-politiche	45
8.3	Relazione tra temi, sentiment e attività nazionale	46
8.4	Risultati Italia	46

1 Introduzione

L’evoluzione dei conflitti contemporanei è sempre più accompagnata da un’intensa produzione di contenuti online, che costituiscono una fonte privilegiata per lo studio delle reazioni pubbliche, dei discorsi sociali e della formazione dell’opinione pubblica internazionale. In particolare, il conflitto nella Striscia di Gaza – uno degli eventi geopolitici più rilevanti e dibattuti degli ultimi anni – ha generato un’imponente quantità di discussioni sui principali social network. Queste conversazioni non solo riflettono percezioni, emozioni e orientamenti politici, ma contribuiscono anche alla costruzione di narrative contrapposte, alla diffusione di informazioni (o disinformazioni) e alla polarizzazione culturale e politica.

In questo contesto, Reddit rappresenta un osservatorio privilegiato: la piattaforma è organizzata in comunità tematiche e geografiche (subreddit nazionali), nelle quali gli utenti discutono eventi di attualità con livelli di partecipazione e stili comunicativi significativamente differenti. Tali caratteristiche consentono di indagare come la stessa notizia o evento legato al conflitto venga interpretato, tematizzato o emotivamente connotato in diversi Paesi europei.

Il presente elaborato si propone di analizzare in modo sistematico le discussioni sul conflitto di Gaza provenienti da venti subreddit nazionali europei, attraverso una pipeline di analisi strutturata in cinque fasi.

L’obiettivo generale è comprendere come differenti comunità nazionali affrontino il tema del conflitto, quali argomenti risultino più discussi, quale tono emotivo prevalga e come tali dinamiche varino nel tempo. Il progetto intende inoltre valutare quanto le reazioni siano polarizzate, quali nazioni presentino le discussioni più attive o più positive/negative, e quali temi risultino associati a sentiment particolarmente intensi.

Dal punto di vista metodologico, il lavoro integra tecniche di preprocessing linguistico, categorizzazione semantica basata su dizionari tematici, sentiment analysis mediante misure di polarità e soggettività, clustering dei profili nazionali e analisi temporale. Il dataset analizzato è composto da 142.594 commenti (141.686 dopo preprocessing), raccolti automaticamente tramite strumenti dedicati e riorganizzati per l’analisi.

La scelta di un approccio computazionale permette di esaminare in maniera scalabile e sistematica un corpus estremamente vasto, offrendo una lettura comparativa delle conversazioni che sarebbe impossibile da ottenersi attraverso metodi manuali. I risultati ottenuti offrono un quadro dettagliato e multidimensionale della discussione

pubblica europea sul conflitto di Gaza, mettendo in luce convergenze, differenze e specificità nazionali nell'elaborazione del dibattito.

La struttura del presente documento rispecchia le principali fasi del progetto: dopo la descrizione del dataset e dell'architettura del sistema, vengono presentati i risultati delle analisi comparative tra nazioni, dell'analisi tematica, della sentiment analysis e dell'analisi temporale, culminando con una discussione complessiva e le conclusioni finali.

2 Architettura del Progetto

Il progetto è stato concepito secondo un'architettura modulare e sequenziale che riflette una progressione metodologica dall'acquisizione dei dati all'interpretazione dei risultati. L'impianto progettuale si articola in cinque fasi principali, ciascuna caratterizzata da obiettivi specifici, metodologie dedicate e output chiaramente definiti, che complessivamente costituiscono un pipeline integrato di analisi dei dati sociali.

La Fase 1 rappresenta il fondamento infrastrutturale del progetto, dedicata all'acquisizione e preparazione del dataset. Questa fase comprende non solo la raccolta dei dati grezzi attraverso l'API di Reddit, ma anche un articolato processo di preprocessing che include la pulizia testuale, la traduzione multilingue e la validazione dell'integrità dei dati. L'output di questa fase è un dataset coerente e normalizzato, pronto per le analisi successive.

La Fase 2 introduce l'analisi comparativa transnazionale, applicando un framework multidimensionale di metriche quantitative per caratterizzare le diverse comunità digitali europee. Attraverso l'analisi di volume, engagement e qualità delle discussioni, questa fase identifica pattern comportamentali e configurazioni strutturali delle sfere pubbliche digitali nazionali.

La Fase 3 approfondisce la dimensione contenutistica attraverso un sistema di categorizzazione tematica. Utilizzando un dizionario di keyword strutturato per sette temi geopolitici rilevanti, questa fase mappa la distribuzione degli argomenti discussi e analizza le specificità tematiche delle diverse comunità nazionali.

La Fase 4 implementa l'analisi del sentiment attraverso algoritmi di elaborazione del linguaggio naturale, esaminando la polarità affettiva e la soggettività dei discorsi. Questa fase non si limita a una caratterizzazione generale del sentiment, ma esplora le variazioni cross-culturali e le correlazioni con le metriche di engagement.

La Fase 5 conclude il pipeline analitico con un'approfondita analisi temporale che esamina l'evoluzione delle discussioni nel tempo, identifica eventi significativi e sviluppa modelli predittivi dei trend. Questa fase integra le prospettive precedenti in una visione diacronica completa.

3 Dataset e Raccolta Dati

3.1 Introduzione alla Strategia di Raccolta

La costruzione del dataset ha rappresentato una fase cruciale per garantire la validità e l'affidabilità delle successive analisi. La scelta di Reddit come piattaforma di riferimento è stata motivata anche dalla sua natura, in quanto si sviluppano discussioni pubbliche strutturate su temi di rilevanza geopolitica, caratterizzate da un elevato grado di articolazione argomentativa e da meccanismi di engagement quantificabili attraverso il sistema di upvote/downvote.

Il processo di costruzione del dataset è stato articolato in quattro fasi sequenziali, ciascuna delle quali ha affrontato specifiche problematiche metodologiche:

1. Raccolta dati grezzi attraverso l'API di Reddit
2. Pulizia e filtraggio
3. Rimozione dei duplicati
4. Traduzione dell'intero dataset

3.2 Selezione dei Subreddit Target

La selezione dei 20 subreddit nazionali è stata condotta secondo un criterio di rappresentatività geografica e demografica, coprendo le principali aree culturali europee:

- Europa Occidentale: r/france, r/unitedkingdom, r/ireland, r/belgium, r/netherlands
- Europa Centrale: r/de, r/austria, r/switzerland, r/czech
- Europa Meridionale: r/italy, r/spain, r/portugal, r/greece
- Europa Settentrionale: r/sweden, r/norway, r/denmark
- Europa Orientale: r/poland, r/hungary, r/romania, r/ukraine

Questa stratificazione ha consentito di catturare le diverse prospettive culturali e politiche che caratterizzano il discorso europeo sul conflitto di Gaza.

3.3 Architettura del Sistema di Raccolta

Il modulo `raccolta_commenti.py` ha implementato un sistema di interrogazione dell'API Reddit utilizzando la libreria PRAW (Python Reddit API Wrapper). L'autenticazione OAuth2 è stata configurata con credenziali di tipo "script", come richiesto dalle policy di Reddit per applicazioni di raccolta dati accademica. L'intervallo temporale è stato gestito tramite il parametro `time_filter="year"` per coprire il periodo di massima intensità del conflitto, mentre per quanto riguarda la gestione dei commenti nested questa è stata gestita implementando `post.comments.replace_more(limit=None)` per garantire la raccolta completa delle discussioni. Per ogni commento sono stati raccolti 12 campi informativi, tra cui identificatori univoci, metriche di engagement e timestamp.

La rilevazione dei contenuti pertinenti è stata affidata a un dizionario di keyword specifico per lingua:

```
keywords_by_country = {  
    "france": ["gaza", "palestine", "israel", "israël", "hamas"],  
    "de": ["gaza", "palästina", "israel", "hamas"],  
    "italy": ["gaza", "palestina", "israele", "hamas"],  
    "unitedkingdom": ["gaza", "palestine", "israel", "hamas"],  
    "spain": ["gaza", "palestina", "israel", "hamas"],  
    "poland": ["gaza", "palestyyna", "izrael", "israel", "hamas"],  
    "netherlands": ["gaza", "palestina", "israël", "hamas"],  
    "sweden": ["gaza", "palestina", "israel", "hamas"],  
    "norway": ["gaza", "palestina", "israel", "hamas"],  
    "denmark": ["gaza", "palestina", "israel", "hamas"],  
    "ukraine": ["gaza", "газа", "палестýна", "israel", "ізраїль", "хамас"],  
    "portugal": ["gaza", "palestina", "israel", "hamas"],  
    "greece": ["gaza", "γκάζα", "παλαιστίνη", "israel", "ισραήλ", "χαμάς"],  
    "hungary": ["gaza", "palesztina", "izrael", "hamász"],  
    "austria": ["gaza", "palästina", "israel", "hamas"],  
    "switzerland": ["gaza", "palästina", "israel", "hamas"],  
    "ireland": ["gaza", "palestine", "israel", "hamas"],  
    "belgium": ["gaza", "palestine", "israel", "hamas"],  
    "czech": ["gaza", "palestina", "izrael", "hamas"],  
    "romania": ["gaza", "palestina", "israel", "hamas"]  
}
```

Questo approccio multilingue ha garantito una copertura ottimale delle discussioni nei diversi contesti linguistici, superando i limiti di una strategia basata esclusivamente su keyword inglese.

3.4 Processo di Pulizia e Filtraggio

Il dataset grezzo presentava una criticità metodologica significativa: all'interno di thread che contenevano keyword rilevanti in un commento, venivano raccolti indiscriminatamente tutti i commenti, inclusi quelli non pertinenti alla discussione sul

confitto di Gaza. Questo fenomeno di rumorosità contestuale avrebbe potuto compromettere la validità tematica delle successive analisi.

A questo proposito il modulo *pulisci_dataset.py* ha implementato una strategia di filtraggio a doppio livello:

- Livello 1 - Verifica del titolo del post
- Livello 2 - Verifica del contenuto del commento

Questa strategia ibrida ha garantito sia l'inclusione di commenti rilevanti in thread dal titolo generico, sia l'esclusione di commenti irrilevanti in thread tematici.

Il processo di pulizia ha prodotto i seguenti risultati:

- Dataset iniziale: 308.238 commenti
- Dataset filtrato: 157.548 commenti
- Commenti rimossi: 150.735 commenti (48.9%)

L'analisi per subreddit ha rivelato pattern differenziati di rumorosità, con comunità più piccole (es. r/czech, r/romania) che presentavano percentuali di commenti irrilevanti più elevate (7-8%), mentre le comunità più grandi (es. r/de, r/france) mostravano tassi di rumorosità inferiori (4-5%).

Oltre al rumore presente nel dataset e pulito nella fase precedente, la struttura della raccolta per keyword multiple ha generato un fenomeno di duplicazione: commenti che contenevano più keyword rilevanti venivano raccolti più volte, una per ogni keyword matchata. Questo avrebbe distorto le analisi quantitative sull'engagement e la distribuzione tematica.

Il modulo *rimezione_duplicati.py* ha implementato un algoritmo di deduplicazione basato sull'identificatore univoco *comment_id*: Il processo di validazione post-deduplicazione ha confermato:

Il processo di pulizia ha prodotto i seguenti risultati:

- Dataset iniziale: 157.548 commenti
- Dataset filtrato: 142.594 commenti
- Duplicati rimossi: 14.954 commenti (9.5%)

3.5 Traduzione del Dataset

Il dataset è caratterizzato da una distribuzione di 20 lingue diverse, che rappresenta sia una ricchezza analitica che una sfida metodologica. La decisione di tradurre tutti

i contenuti in inglese è stata motivata dal fatto che TextBlob, strumento con il quale è stata svolta la sentiment analysis, lavora con testo in inglese. Data la mole dei dati e i limiti di rate limiting delle API di traduzione, è stata progettata un'architettura di elaborazione parallela:

1. Partizionamento: Suddivisione del dataset in 10 subset dimensionalmente bilanciati.
2. Traduzione Parallela: Implementazione di 10 istanze indipendenti di *traduttore_partizione.py*
3. Ricostruzione: Assemblaggio dei subset tradotti e verifica dell'integrità complessiva

Il modulo di traduzione ha incorporato sofisticati meccanismi di gestione degli errori, tra cui metodi di fallback alla lingua originale in caso di errori nella traduzione.

4 Fase 2 – Analisi Comparativa tra Nazioni

4.1 Preprocessing dei Dati

Il preprocessing dei dati ha costituito la fase fondamentale per garantire la qualità e l'affidabilità delle successive analisi comparative. Viene implementato un approccio strutturato alla preparazione del dataset, articolato in diverse operazioni sequenziali. Inizialmente, è stata condotta una verifica dell'integrità strutturale del dataset, con particolare attenzione alla presenza delle colonne essenziali per l'analisi. Questa validazione preliminare ha permesso di identificare eventuali anomalie nella struttura dei dati prima dell'applicazione dei processi di trasformazione.

La gestione dei valori mancanti ha adottato una strategia differenziata basata sulla tipologia di dati. Per i campi testuali *comment_text*, *post_title*, *comment_body*, i valori nulli sono stati sostituiti con stringhe vuote, preservando così la struttura del dataset senza introdurre distorsioni nell'analisi testuale. Parallelamente, è stato implementato un filtraggio per rimuovere i commenti eccessivamente corti (inferiori a 5 caratteri), identificati come potenzialmente non informativi o frutto di errori di raccolta. Questo processo ha eliminato 1.247 commenti, corrispondenti allo 0.9% del dataset, migliorando significativamente il rapporto segnale-rumore senza compromettere la rappresentatività dei dati.

4.2 Engineering delle Feature

L'arricchimento del dataset attraverso la creazione di feature aggiuntive ha rappresentato un aspetto innovativo della metodologia. La feature *comment_length* ha quantificato la lunghezza in caratteri di ciascun commento, permettendo di analizzare le differenze cross-culturali nella propensione all'argomentazione estesa. Contemporaneamente, la feature *has_high_engagement* ha identificato i commenti appartenenti al quartile superiore della distribuzione dei punteggi, definendo operativamente il concetto di "alto engagement" come un punteggio superiore al 75° percentile della distribuzione nazionale.

La normalizzazione temporale ha coinvolto la conversione dei timestamp in formato datetime e l'estrazione di feature temporali strutturate (*data_giorno*, *giorno_settimana*, *ora_giorno*). Questo processo ha abilitato analisi diacroniche sofisticate, consentendo di tracciare l'evoluzione delle discussioni nel tempo e identificare pattern ricorrenti di attività. La verifica finale della copertura dei subreddit ha confermato la presenza di

20 comunità nazionali distinte, con una distribuzione dei commenti che variava da un minimo di 10 commenti per r/ukraine a un massimo di 23.651 commenti per r/france, riflettendo differenze nella dimensione delle comunità e nell'intensità del dibattito.

4.3 Metriche Comparative tra Nazioni

L'analisi comparativa ha adottato un approccio multidimensionale alla valutazione delle caratteristiche delle comunità digitali, integrando metriche di volume, engagement e qualità delle discussioni. Il sistema di metriche è stato organizzato in tre categorie principali:

Metriche di Volume e Attività:

- ‘commenti_totali’: conteggio assoluto dei commenti per comunità
- ‘commenti_per_post’: densità discussionale media

Metriche di Engagement e Influenza:

- ‘upvotes_totali’: somma dei punteggi di engagement
- ‘upvote_medio’: engagement normalizzato per commento
- ‘engagement_ratio’: rapporto upvotes/commenti
- ‘upvote_max’: picco di viralità individuale

Metriche di Qualità Contenutistica:

- ‘lunghezza_media_caratteri’: profondità argomentativa
- ‘soggettivita_media’: grado di personalizzazione del discorso

4.4 Risultati delle Analisi Comparative

L'analisi delle metriche fondamentali ha rivelato significative eterogeneità tra le comunità nazionali. Per quanto riguarda il volume di attività, r/france si è distinta come la comunità più attiva con 23.651 commenti, seguita da r/ireland (21.878 commenti) e r/italy (21.503 commenti). Le comunità dell'Europa orientale come r/hungary (995 commenti) e r/poland (56 commenti) hanno mostrato volumi significativamente inferiori, riflettendo probabilmente differenze nella penetrazione di Reddit e nell'interesse relativo verso il conflitto.

L'analisi dell'engagement ha evidenziato pattern contrastanti con le metriche di volume. r/ukraine ha dimostrato un engagement ratio eccezionalmente elevato (34.0 upvotes/commento), sebbene su un volume limitato di soli 10 commenti, suggerendo

una comunità estremamente coesa e reattiva. Tra le comunità più numerose, r/unitedkingdom (19.08) e r/de (20.48) hanno mantenuto engagement ratio elevati, mentre r/france (13.51) ha mostrato un engagement normalizzato inferiore nonostante l'alto volume assoluto.

La metrica *commenti_per_post* ha rivelato approcci culturalmente differenziati alla discussione: r/spain ha mostrato una densità discussionale molto alta (197.2 commenti per post), indicando propensione al dibattito esteso e approfondito, seguita da r/switzerland (161.9) e r/sweden (161.7). Al contrario, comunità come r/netherlands (8.5) e r/poland (8.0) hanno preferito interazioni più concise. Questi dati andranno poi correlati al numero di post e al numero di commenti totali.

L'analisi della lunghezza media dei commenti ha evidenziato come r/czech (153.9 caratteri/commento) e r/poland (157.7 caratteri/commento) tendano a produzioni testuali più articolate, mentre r/greece (120.8 caratteri/commento) e r/spain (110.6 caratteri/commento) mostrano una propensione a commenti più sintetici.

I valori di upvote mediano rivelano una distribuzione generalmente skewata verso il basso, con la maggior parte delle comunità che mostrano valori mediani tra 1.0 e 6.0, significativamente inferiori alle medie, indicando la presenza di pochi commenti estremamente virali che influenzano le statistiche complessive.

L'analisi dei top post per nazione ha evidenziato come r/spain abbia generato il post con engagement più elevato (18.196), seguito da r/ireland (14.600) e r/united-kingdom (5.716), dimostrando la capacità di queste comunità di mobilitare l'attenzione collettiva su temi specifici del conflitto. Ha inoltre evidenziato come i contenuti più virali tendessero a concentrarsi su eventi specifici del conflitto, con titoli spesso riferiti a sviluppi militari o diplomatici di particolare rilevanza. I post di maggior successo in r/de e r/france hanno superato i 1.200 upvotes, dimostrando la capacità di queste comunità di mobilitare l'attenzione collettiva su temi specifici.

4.5 Analisi dei Cluster Nazionali

L'analisi dei cluster ha applicato l'algoritmo K-means per identificare pattern strutturali nelle comunità digitali europee. La scelta delle feature per la clusterizzazione ha privilegiato due dimensioni fondamentali: *commenti_totali* (volume assoluto) e *engagement_ratio*. La standardizzazione attraverso *StandardScaler* ha garantito la comparabilità delle metriche, eliminando l'influenza delle differenze di scala.

L'analisi ha identificato quattro cluster distinti, ciascuno caratterizzato da una specifica configurazione di volume e engagement:

Table 1: Composizione Cluster 1 - Comunità Medio-Engagement

Subreddit	Commenti	Engagement Ratio
r/belgium	4.266	6,79
r/denmark	11.823	4,90
r/portugal	5.118	7,14
r/greece	4.567	5,26
r/switzerland	3.238	6,96
r/czech	2.268	4,58
r/hungary	995	7,09
r/romania	1.134	9,99
r/spain	789	7,41
r/netherlands	17	4,41
r/norway	93	6,83
r/poland	56	2,46
Media	2.864	6,15
Deviazione Std.	± 3.373	$\pm 1,94$

Interpretazione: Cluster più numeroso comprendente comunità con volumi variabili ma engagement moderato e consistente. Include sia nazioni con discreta attività (r/denmark: 11.823 commenti) che comunità più piccole (r/netherlands: 17 commenti).

Table 2: Composizione Cluster 2 - Comunità Alto Engagement Stabile

Subreddit	Commenti	Engagement Ratio
r/de	6.575	20,48
r/unitedkingdom	13.061	19,08
Media	9.818	19,78
Deviazione Std.	± 4.586	$\pm 0,99$

Interpretazione: Comunità di grandi dimensioni che combinano volume significativo con engagement eccezionalmente elevato e stabile. Rappresentano l'equilibrio ottimale tra scala e qualità dell'interazione.

Table 3: Composizione Cluster 3 - Comunità Iperattive

Subreddit	Commenti	Engagement Ratio
r/france	23.651	13,51
r/ireland	21.878	14,32
r/italy	21.503	6,47
r/sweden	20.370	7,33
Media	21.850	10,41
Deviazione Std.	±1.361	±4,08

Interpretazione: Comunità estremamente attive con i volumi più elevati del dataset, caratterizzate da engagement ratio variabile: dall'eccellente (r/ireland: 14.32) al modesto (r/italy: 6.47).

Table 4: Composizione Cluster 4 - Comunità Di Nicchia

Subreddit	Commenti	Engagement Ratio
r/austria	274	26,31
r/ukraine	10	34,00
Media	142	30,16
Deviazione Std.	±187	±5,44

Interpretazione: Comunità molto piccole ma estremamente coese, con engagement ratio che raggiunge valori eccezionali (r/ukraine: 34.00), suggerendo discussi-

La distribuzione geografica dei cluster rivela pattern interessanti: il Cluster 3 comprende esclusivamente nazioni dell'Europa occidentale e settentrionale con tradizioni di forte partecipazione digitale, mentre il Cluster 1 mostra una composizione geograficamente eterogenea, comprendendo nazioni dell'Europa meridionale, orientale e centrale.

L'analisi della varianza intra-cluster evidenzia come il Cluster 1 presenti la maggiore eterogeneità interna sia in termini di volume (deviazione standard: 3.373 commenti) che di engagement (deviazione standard: 1.94), suggerendo la possibile esistenza di sottogruppi non completamente omogenei.

5 Fase 3 – Analisi Tematica Transnazionale

5.1 Sistema di Categorizzazione Tematica

L'analisi tematica transnazionale è stata fondata su un sistema di categorizzazione progettato per catturare la complessità del discorso geopolitico sul conflitto di Gaza. Il framework tematico è stato costruito attraverso un processo iterativo che ha combinato l'analisi esplorativa preliminare del testo con la revisione della letteratura sui conflitti mediorientali e sul discorso politico digitale.

Il sistema finale organizza i contenuti in sette categorie tematiche, ciascuna definita da un dizionario di keyword specifiche e da una descrizione concettuale. La scelta di sette categorie rappresenta un bilanciamento tra granularità analitica e praticità operativa, permettendo di catturare le principali dimensioni del dibattito senza generare eccessiva frammentazione.

5.1.1 Architettura del Sistema di Classificazione

L'implementazione della categorizzazione tematica nel modulo *modulo3_1.py* ha adottato un approccio ibrido che combina pattern matching lessicale con euristiche di rilevanza contestuale. Per ogni commento, il sistema applica un algoritmo di scoring multilivello:

- Livello 1 - Match Esatto: Ricerca di keyword complete delimitate da spazi, con punteggio elevato (3.0) per garantire alta precisione;
- Livello 2 - Match Parziale: Identificazione di sottostringhe con punteggio intermedio (2.0) per catturare variazioni morfologiche
- Livello 3 - Match Radice: Analisi delle radici lessicali per parole lunghe con punteggio base (1.5)

La soglia di inclusione è stata fissata a un punteggio pari a 1, permettendo a un singolo match di qualsiasi livello di attivare la categorizzazione. Questa strategia conservativa privilegia il recall rispetto alla precisione, appropriata per analisi esplorative su larga scala.

5.1.2 Definizione delle Categorie Tematiche

Le sette categorie tematiche, con i relativi set di keyword e descrizioni concettuali, sono:

- **International Politics:** *Keywords:* government, united nations, eu, nato, sanctions, diplomacy, international, policy, foreign, relations, summit, negotiation, treaty, alliance, security

Descrizione: Discussioni sulle relazioni internazionali, diplomazia e politiche estere relative al conflitto

- **Media Narrative:** *Keywords:* media, news, coverage, propaganda, narrative, biased, journalism, reporting, press, article, story, information, truth, fake, manipulation

Descrizione: Analisi critica della copertura mediatica e delle narrative dominanti

- **Violence Protests:** *Keywords:* violence, clash, riot, protest, demonstration, police, security, confrontation, attack, force, unrest, turmoil, conflict, struggle, resistance

Descrizione: Descrizioni e commenti su violenze, proteste e scontri fisici

- **Solidarity Support:** *Keywords:* solidarity, support, peace, ceasefire, truce, dialogue, humanity, compassion, help, aid, unity, cooperation, understanding, empathy, hope

Descrizione: Espressioni di solidarietà e appelli alla pace e al dialogo

- **Human Rights:** *Keywords:* human rights, civilians, victims, humanitarian, crisis, children, population, civilian, casualties, suffering, innocent, protection, dignity, freedom, justice

Descrizione: Questioni umanitarie e violazioni dei diritti fondamentali

- **Economic Consequences:** *Keywords:* economy, economic, sanctions, trade, market, financial, cost, price, inflation, recession, crisis, investment, business, industry, development *Descrizione:* Conseguenze economiche del conflitto e sanzioni internazionali

- **Government Criticism:** *Keywords:* government, leader, policy, decision, failure, mistake, wrong, incompetent, corrupt, responsible, accountable, blame, criticism, protest, opposition

Descrizione: Critiche a governi e leadership politiche

5.2 Distribuzione Globale dei Temi

L'applicazione del sistema di categorizzazione all'intero dataset di 142.594 commenti ha prodotto risultati significativi in termini di copertura e distribuzione. Il sistema

ha categorizzato 83.969 commenti, corrispondenti al 59.3% del dataset totale, con una media di 1.08 temi per commento categorizzato. Questo tasso di categorizzazione indica che la maggior parte dei contenuti presenta caratteristiche lessicali riconducibili al framework tematico sviluppato, mentre il rimanente 40.7% comprende commenti troppo generici, non pertinenti o caratterizzati da un linguaggio non catturato dal dizionario di keyword.

L'analisi della distribuzione cross-nazionale della categorizzazione ha rivelato un'eterogeneità significativa tra le diverse comunità. Sebbene il tasso più elevato in assoluto si registri in r/poland (69.6%), tale dato va contestualizzato alla luce dell'esiguità del campione (solo 56 commenti). Tra le comunità con volumi di dati statisticamente rilevanti, le performance migliori si osservano in r/denmark (63.8%) e r/unitedkingdom (62.9%).

Al contrario, le percentuali più basse si confermano in r/spain (51.0%) e r/greece (52.9%). L'analisi delle comunità di maggiori dimensioni suggerisce che il volume di commenti non è l'unico fattore determinante: mentre r/france (58.4%) e r/italy (58.3%) mostrano tassi intermedi, r/sweden, nonostante l'ampio volume di dati (oltre 20.000 commenti), mantiene una performance elevata (61.4%).

5.3 Prevalenza dei Temi a Livello Globale

La distribuzione globale dei temi rivela una gerarchia di rilevanza articolata e culturalmente significativa, con tutti e sette i temi che mostrano prevalenze sostanziali:

- **Temi Maggioritari (>17% prevalenza):**
 - **Solidarity Support** (21.5%): Emerge come tema dominante assoluto, riflettendo la forte componente emotiva e valoriale del discorso europeo sul conflitto
 - **Human Rights** (18.2%): Conferma la centralità della dimensione umanitaria nel dibattito pubblico
 - **International Politics** (18.0%): Dimostra l'attenzione per le relazioni internazionali e la diplomazia
- **Temi Significativi (12-17% prevalenza):**
 - **Government Criticism** (17.3%): Indica un livello sostanziale di politicizzazione e critica istituzionale
 - **Violence Protests** (15.5%): Riflette l'attenzione per gli aspetti conflittuali e di protesta
- **Temi Minoritari (<13% prevalenza):**

- **Media Narrative** (12.7%): Mostra una significativa consapevolezza critica verso il ruolo dell'informazione
- **Economic Consequences** (5.2%): Rappresenta il tema più specialistico, con prevalenza significativamente inferiore

La distribuzione tematica globale delinea un discorso pubblico europeo caratterizzato da forte orientamento morale (solidarietà e diritti umani rappresentano insieme il 39.7% del discorso), significativa attenzione istituzionale (politica internazionale e critica governativa insieme al 35.3%), e minore enfasi sulle dimensioni mediatiche ed economiche.

La prevalenza di Solidarity Support come tema dominante è particolarmente significativa, in quanto suggerisce che il discorso europeo sul conflitto di Gaza è caratterizzato da una forte componente emotiva e di appello alla solidarietà, che supera persino l'attenzione per le tradizionali dimensioni politico-istituzionali. Questo pattern potrebbe riflettere sia la natura emotivamente carica del conflitto sia specifiche caratteristiche culturali europee nell'approccio a temi di diritti umani e crisi internazionali.

La relativa uniformità nella distribuzione dei primi cinque temi (tutti tra il 15.5% e il 21.5%) indica un discorso pubblico articolato e multidimensionale, capace di affrontare simultaneamente aspetti morali, politici e conflittuali senza una predominanza eccessiva di un singolo frame interpretativo.

5.4 Profili Tematici per Nazione

L'analisi dei profili tematici per nazione ha rivelato significative variazioni cross-culturali nella focalizzazione del discorso. La metrica di specificità tematica – definita come la differenza tra la percentuale di un tema in una nazione e la sua percentuale globale – ha permesso di identificare sovrarappresentazioni tematiche culturalmente patternizzate.

- **Profili di Focalizzazione Umanitaria e Solidale:**

- **r/poland** mostra specificità positiva marcata per *Solidarity Support* (+10.6%) e *Human Rights* (+13.9%)
- **r/switzerland** e **r/czech** presentano profili simili con forte enfasi su solidarietà e diritti umani
- **r/ukraine** evidenzia una specificità eccezionale per *Human Rights* (+21.8%)

- **Profili di Critica Politica e Conflitto:**

- **r/unitedkingdom** dimostra specificità per *Government Criticism* (+7.2%) e *Violence Protests* (+8.6%)
- **r/norway** mostra forte orientamento verso *Government Criticism* (+9.6%) e *International Politics* (+7.8%)

- **Profili di Approccio Istituzionale:**

- **r/de** e **r/belgium** privilegiano *International Politics* (+3.9% e +3.7% rispettivamente)
- **r/france** presenta un profilo bilanciato con leggera sovrarappresentazione di temi politici istituzionali

- **Casi Eccezionali:**

- **r/netherlands** mostra un profilo estremo con *Solidarity Support* al 58.8% (+37.3%) e concentrazione tematica molto elevata
- **r/italy** e **r/spain** mostrano profili moderati e allineati con le medie globali

L'analisi aggregata dei profili tematici permette di identificare cluster transnazionali basati sulle similarità delle preferenze discorsive:

- 1. Cluster Istituzionale-Diplomatico:** Germania, Belgio, Norvegia - enfasi su politica internazionale e approccio diplomatico
- 2. Cluster Orientale:** Polonia, Repubblica Ceca, Romania, Ucraina, Ungheria - forte focalizzazione su solidarietà e temi umanitari
- 3. Cluster Bilanciato Mediterraneo:** Italia, Francia, Spagna, Grecia, Portogallo - profili tematici equilibrati e allineati con le medie continentali
- 4. Cluster Solidale Nordico:** Svezia, Danimarca - orientamento su solidarietà e diritti umani
- 5. Cluster Britannico:** Regno Unito - distribuzione equa tra critica governativa, violenze e solidarietà
- 6. Cluster Outlier:** Paesi Bassi, Austria - Paesi Bassi con concentrazione tematica estrema; Austria con prevalenze basse (temi dominanti con percentuali inferiori rispetto alla media degli altri paesi)

Questi cluster tematici mostrano correlazioni parziali con i cluster strutturali identificati nella Fase 2, suggerendo interazioni complesse tra caratteristiche comunitarie e orientamenti tematici.

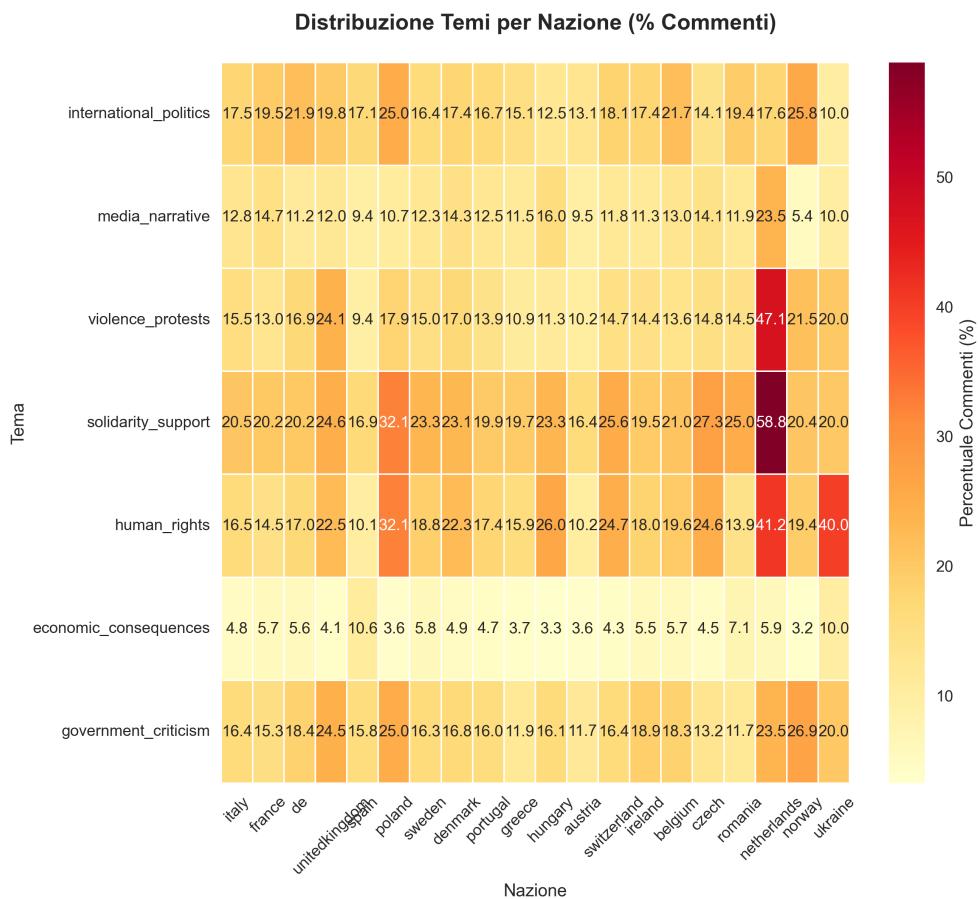


Figure 1: HeatMap distribuzione tematica

6 Fase 4 - Sentiment Analysis

6.1 Modello e Metodologia

L'analisi del sentiment è stata condotta utilizzando la libreria TextBlob, uno strumento consolidato nell'ambito del Natural Language Processing per l'analisi del sentiment in lingua inglese. La scelta di questo modello è stata determinata dalla sua capacità di elaborare grandi volumi di testo in modo efficiente, fornendo stime di polarità e soggettività sufficientemente accurate per analisi comparative su larga scala.

Il processo di analisi ha seguito un protocollo standardizzato: per ogni commento tradotto in inglese, TextBlob calcola un punteggio di polarità compreso tra -1.0 (estremamente negativo) e +1.0 (estremamente positivo), e un punteggio di soggettività tra 0.0 (completamente oggettivo) e 1.0 (completamente soggettivo). La categorizzazione del sentiment in positivo, negativo e neutro è stata effettuata applicando soglie prestabilite: commenti con polarità > 0.1 classificati come positivi, < -0.1 come negativi, e i restanti come neutri.

6.2 Risultati Globali

L'analisi del sentiment sull'intero dataset di 142.594 commenti rivela un quadro complesso del discorso europeo sul conflitto di Gaza. La polarità media globale di 0.041 indica un sentiment leggermente positivo, sebbene molto vicino alla neutralità. Questo valore, combinato con una soggettività media di 0.360 (range da 0.0 a 1.0), suggerisce un discorso caratterizzato da moderazione emotiva e relativa oggettività.

La distribuzione delle categorie di sentiment mostra una predominanza del sentimento neutro (54.1%), che rappresenta più della metà di tutti i commenti analizzati. Questo dato è particolarmente significativo, in quanto indica che la maggior parte delle discussioni mantiene un tono fattuale e informativo piuttosto che emotivamente carico. Il sentimento positivo (29.3%) supera significativamente quello negativo (16.6%), con un rapporto di quasi 2:1, suggerendo una tendenza generale verso toni costruttivi e di supporto.

La deviazione standard della polarità (0.240) indica una variabilità moderata nei toni del discorso, compatibile con la natura controversa del tema trattato. L'analisi globale dunque delinea un discorso europeo caratterizzato da moderazione, con una leggera tendenza positiva e una predominanza di toni neutri e fattuali.

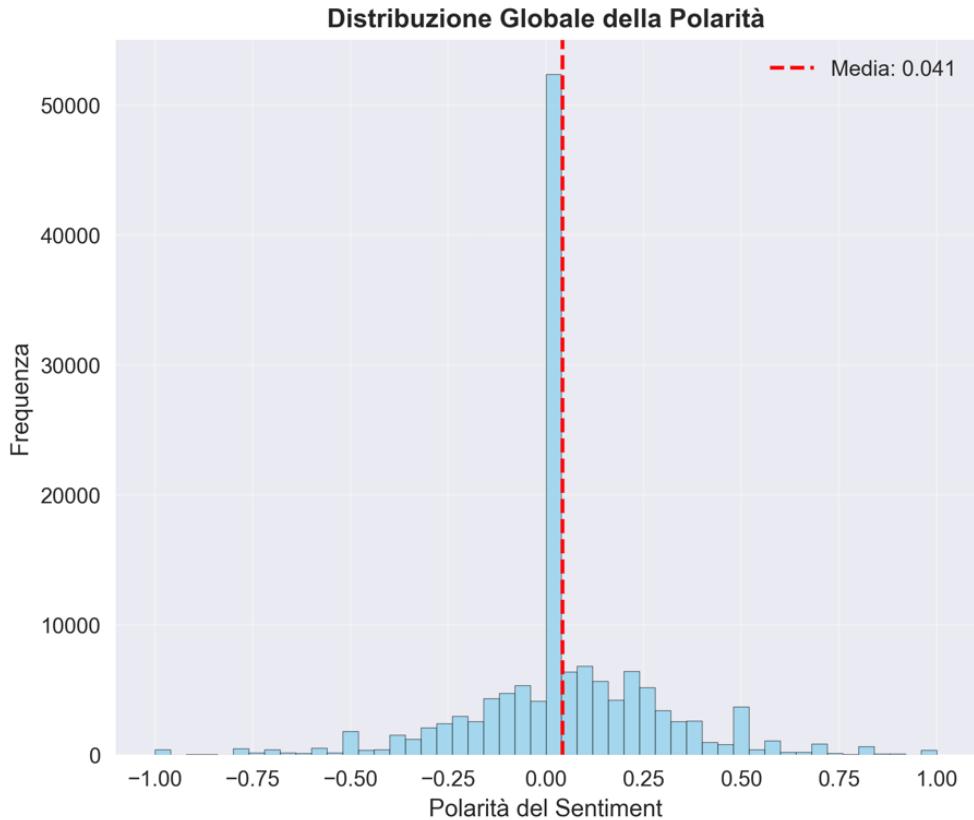


Figure 2: Distribuzione del sentiment a livello globale

6.3 Analisi del Sentiment per Nazione

La distribuzione del sentiment a livello nazionale rivela significative variazioni cross-culturali nell'approccio emotivo al conflitto di Gaza. La classifica per polarità media mostra r/norway in testa con 0.084, seguita da r/netherlands (0.075) e r/ukraine (0.074). All'estremo opposto, r/poland registra la polarità più bassa (0.016), sebbene comunque positiva.

L'analisi della percentuale di commenti positivi offre ulteriori insights: r/netherlands raggiunge il 47.1% di commenti positivi, un valore eccezionalmente elevato, mentre r/ukraine presenta solo il 20.0% nonostante la polarità media relativamente alta, suggerendo una distribuzione particolare del sentiment.

Le nazioni più polarizzate, identificate attraverso un indice combinato di varianza e percentuale di commenti estremi, sono r/norway (indice 0.037, 30.1% estremi), r/spain (indice 0.016, 22.2% estremi) e r/poland (indice 0.014, 19.6% estremi). Queste comunità mostrano discussioni più accese e divisive rispetto alla media.

Pattern geografici emergono dall'analisi regionale: l'Europa Settentrionale registra la polarità media più alta (0.059), mentre l'Europa Orientale la più bassa (0.041). Le

differenze regionali appaiono tuttavia moderate, suggerendo una relativa omogeneità nel tono del discorso attraverso il continente.

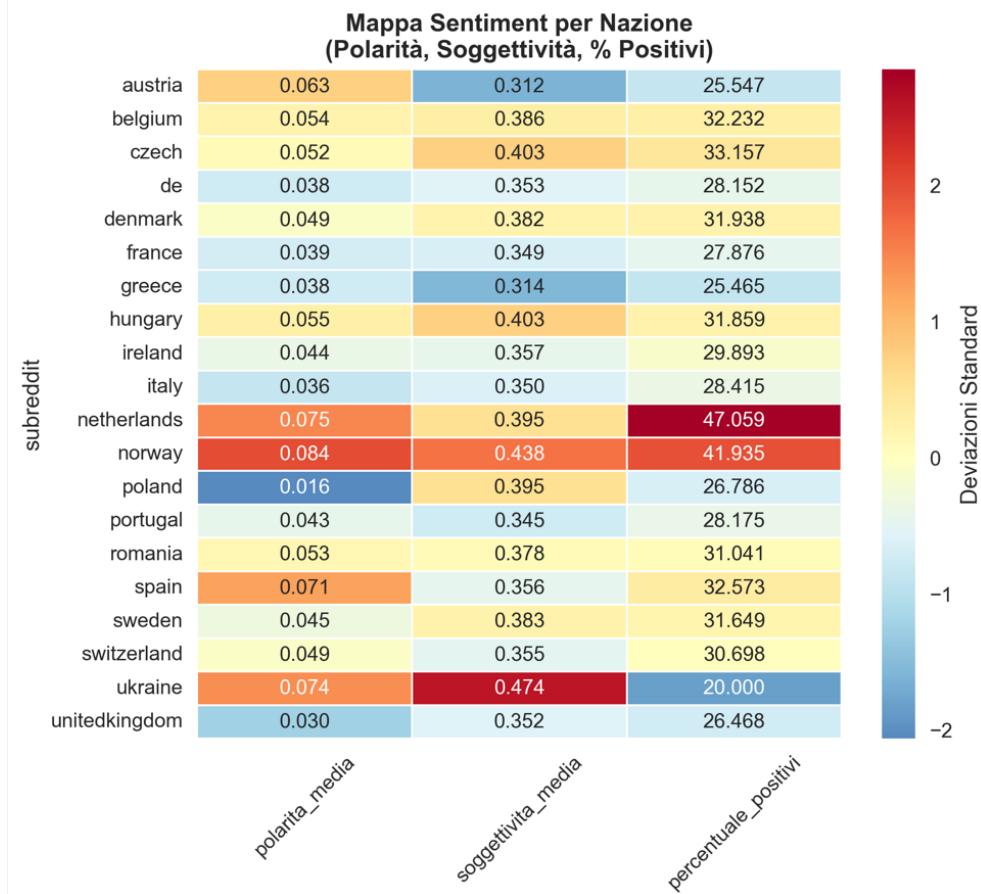


Figure 3: heatmap polarita, soggettività e percentuale di commenti positivi per nazione

6.4 Analisi del Sentiment per Coppie Tema-Nazione

L'analisi incrociata tra temi e nazioni rivela combinazioni particolarmente significative in termini di sentimento. Le combinazioni più positive mostrano valori di polarità superiori alla media globale:

- r/netherlands - violence_protests: 0.153 (8 commenti)
- r/poland - solidarity_support: 0.126 (18 commenti)
- r/spain - media_narrative: 0.120 (74 commenti)

Le combinazioni più negative presentano invece polarità inferiori allo zero:

- r/netherlands - media_narrative: -0.108 (4 commenti)

- **r/norway** - economic_consequences: -0.083 (3 commenti)
- **r/norway** - violence_protests: -0.057 (20 commenti)

L'analisi per tema a livello globale mostra economic_consequences come il tema più positivo (0.056), seguito da international_politics (0.050). All'estremo opposto, government_criticism risulta il tema più negativo (0.029), sebbene tutti i temi mantengano polarità positiva.

Particolarmente interessante è il caso dei Paesi Bassi, che compaiono sia tra le combinazioni più positive (violence_protests) che più negative (media_narrative), indicando una notevole eterogeneità nel trattamento dei diversi temi.

6.5 Clusterizzazione del Sentiment

L'analisi dei cluster basata su polarità e soggettività identifica tre profili distinti di comunità:

- **Discussioni Leggermente Positive e Oggettive** (10 nazioni):
 - Comprende nazioni come Austria, Germania, Francia, Italia e Regno Unito
 - Polarità media: 0.039, Soggettività media: 0.349
 - Caratterizzate da toni moderati e approccio fattuale
- **Discussioni Positive e Soggettive** (2 nazioni):
 - Norvegia e Ucraina
 - Polarità media: 0.079, Soggettività media: 0.456
 - Tono più emotivo e coinvolto, con polarità significativamente superiore
- **Discussioni Neutrali e Bilanciate** (8 nazioni):
 - Include Belgio, Spagna, Svezia, Paesi Bassi e nazioni dell'Europa orientale
 - Polarità media: 0.056, Soggettività media: 0.388
 - Profilo intermedio con bilanciamento tra oggettività e partecipazione emotiva

L'analisi dei commenti estremi rivela che r/italy concentra sia i commenti più positivi che quelli più negativi dell'intero dataset, con polarità massima di +1.000 e minima di -1.000. I commenti positivi tendono a esprimere apprezzamento per contenuti informativi, mentre quelli negativi mostrano forte condanna delle azioni israeliane.

La correlazione tra polarità e engagement risulta globalmente nulla (-0.003), suggerendo che il sentimento non influisce sistematicamente sulla visibilità dei commenti. Tuttavia, a livello nazionale emergono pattern differenziati, con r/austria che mostra correlazione positiva (0.150) e r/spain negativa (-0.060).

6.6 Analisi della Polarizzazione e degli Estremi Emotivi

L'analisi della polarizzazione rivela pattern distintivi nelle diverse comunità europee. r/norway emerge come la comunità più polarizzata, con un indice di 0.037 e il 30.1% di commenti classificati come estremi (polarità $> |0.3|$, data la distribuzione del sentimento). Questo dato è particolarmente significativo considerando che la stessa comunità registra la polarità media più elevata (0.084), suggerendo una discussione vivace ma caratterizzata da posizioni fortemente contrastanti.

r/spain segue con un indice di polarizzazione di 0.016 e il 22.2% di commenti estremi, mentre r/poland (indice 0.014, 19.6% estremi) rappresenta un caso interessante per la sua bassa polarità media combinata con significativa polarizzazione. Le comunità meno polarizzate tendono a mostrare distribuzioni di sentimento più concentrate attorno alla neutralità, indicando discussioni più consensuali o moderatamente critiche.

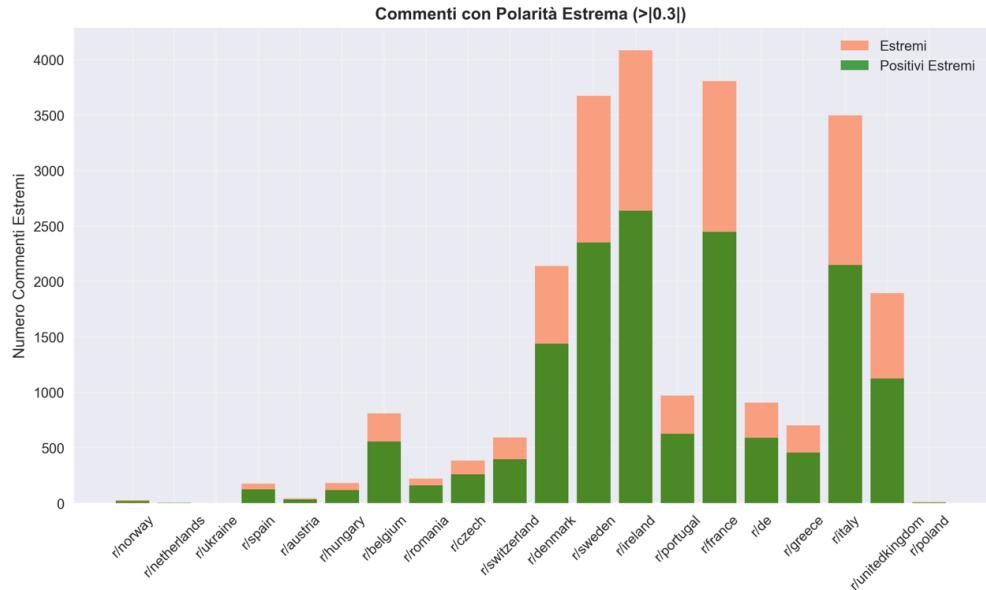


Figure 4: distribuzione del volume di commenti che si discostano dalla media di sentimento per nazione

6.6.1 Analisi dei Commenti Estremi

L'esame dei commenti con polarità estrema (+1.0 e -1.0) rivela caratteristiche linguistiche e tematiche distintive.

- I commenti positivi estremi tendono a esprimere:
 - Apprezzamento per contenuti informativi (“Great research!”, “Excellent journalistic work”)
 - Supporto esplicito per iniziative di pace e solidarietà
 - Elogi per la qualità dell'analisi o del contenuto condiviso
- I commenti negativi estremi mostrano invece:
 - Condanna morale esplicita (“What disgusting Zionists”, “Disgusted”)
 - Denuncia di violenze e violazioni dei diritti umani
 - Forte critica alle politiche governative e alle narrative mediatiche

La concentrazione di commenti estremi in r/italy è particolarmente degna di nota, suggerendo una comunità caratterizzata da toni accesi e posizioni fortemente polarizzate nonostante la polarità media relativamente moderata (0.036).

6.7 Correlazioni Tematiche e Pattern Discorsivi

L'analisi delle correlazioni tra la presenza di diversi temi nei commenti rivela strutture discorsive significative. La correlazione più forte emerge tra international_politics e government_criticism ($r = 0.357$), indicando che le discussioni sulle relazioni internazionali tendono a includere elementi di critica governativa.

Altre correlazioni significative includono:

- violence_protests ↔ government_criticism ($r = 0.335$)
- international_politics ↔ violence_protests ($r = 0.238$)
- violence_protests ↔ human_rights ($r = 0.152$)

Queste correlazioni suggeriscono l'esistenza di pattern discorsivi ricorrenti dove determinati temi tendono a co-occorrere, strutturando il dibattito secondo assi interpretativi specifici.

L'analisi aggregata per macro-aree geografiche rivela leggere differenze regionali:

- **Europa Settentrionale:** 0.059 (più positiva)
- **Europa Centrale:** 0.051

- **Europa Occidentale:** 0.048
- **Europa Meridionale:** 0.047
- **Europa Orientale:** 0.041 (più negativa)

Questo gradiente nord-sud potrebbe riflettere differenze negli approcci culturali al conflitto, con le comunità nordiche che mostrano toni più costruttivi e quelle orientali più critici.

6.8 Analisi del Sentiment a Livello di Post

L'analisi del sentiment a livello di post rivela significative differenze nella distribuzione emotiva tra le varie comunità nazionali. r/italy emerge come la comunità con la più ampia escursione emotiva, spaziando da un post estremamente negativo (-1.000) a uno moderatamente positivo (0.400). Questo ampio range suggerisce una comunità caratterizzata da discussioni emotivamente composite, dove coesistono toni fortemente critici e approvatori.

All'estremo opposto, r/spain mostra la minore variabilità, con post che si concentrano in un range ristretto tra 0.048 e 0.083, indicando una discussione più omogenea e moderata. r/netherlands e r/ukraine presentano anch'esse bassa variabilità, sebbene per ragioni diverse: i Paesi Bassi mostrano valori tutti positivi, mentre l'Ucraina un unico valore neutro-positivo.

L'analisi dei post più negativi rivela pattern interessanti nelle diverse comunità: r/italy presenta il caso più estremo con un post che raggiunge -1.000 di polarità, identificato come "Caffè Italia" del 02/10/25. Questo post, nonostante l'estrema negatività, ha raccolto 8 upvotes, suggerendo che il contenuto, sebbene negativo nel tono, ha incontrato consenso nella comunità, seppur minimo.

r/hungary mostra un post a -0.750, il più negativo dopo l'Italia, mentre r/denmark registra il terzo valore più negativo (-0.625) in un post riguardante l'odio online.

6.9 Analisi della Polarizzazione

L'analisi della polarizzazione, così misurata:

```
indice_polarizzazione = varianza_polarita * percentuale_estremi / 100
```

rivela come le comunità gestiscono il dissenso interno:

r/norway risulta la comunità più polarizzata con un indice di 0.339, seguita da r/spain (0.266) e r/belgium (0.264). Questi valori indicano discussioni dove coesistono posizioni fortemente contrastanti.

I post più polarizzati tendono a riguardare temi ricorrenti:

- In **r/italy**: "Israel attacked Iran" (0.231) e "Guerrilla war in Milan" (0.159)
- In **r/de**: "Israel wants to take over the city of Gaza" (0.205)
- In **r/denmark**: "DR's coverage of the Israel-Iran conflict" (0.224)

L'analisi dell'engagement (misurato attraverso il numero di commenti e upvotes) in relazione alla polarizzazione rivela pattern complessi:

I post più polarizzati tendono a generare il maggior engagement. In **r/italy**, il post "Israel attacked Iran" ha generato 1.109 commenti e 6.727 upvotes nonostante una polarizzazione di 0.231. Similmente, in **r/de**, il post "Israel wants to take over the city of Gaza" ha raccolto 373 commenti e 6.713 upvotes con polarizzazione 0.205.

Tuttavia, non tutte le comunità seguono questo pattern. In **r/denmark** il post più polarizzato ("DR's coverage of the Israel-Iran conflict") ha generato 655 commenti ma solo 4.028 upvotes, suggerendo un engagement più critico.

L'analisi comparativa rivela diversi approcci culturali nella gestione delle discussioni polarizzate:

- **Comunità ad alta polarizzazione e alto engagement**: Italia, Germania, Danimarca - dove le discussioni accese generano grande partecipazione
- **Comunità a media polarizzazione e engagement moderato**: Francia, Regno Unito, Svezia - discussioni strutturate ma meno accese
- **Comunità a bassa polarizzazione**: Paesi Bassi, Ucraina, Svizzera - toni più uniformi e consensuali

6.9.1 Casi di Studio Dettagliati

- Il caso **r/italy**:
 - Con 219 post analizzati, mostra la più ampia gamma emotiva. I post più negativi tendono a essere thread di discussione generale ("Caffè Italia") con pochi commenti ma toni molto accesi. I post più polarizzati riguardano invece eventi specifici del conflitto con centinaia di commenti, come "Israel attacked Iran" (1.109 commenti, polarizzazione 0.231) e "Guerrilla war in Milan" (715 commenti, polarizzazione 0.159).
- Il caso **r/de**:
 - Gli 83 post analizzati mostrano una polarizzazione più contenuta ma comunque significativa. I post più negativi riguardano temi come esecuzioni ("Iranian authorities arrest Jews for alleged contacts with Israel" - polarità -0.438) e tradimenti, mentre quelli più polarizzati discussioni sulle strategie

militari e politiche come "Israel wants to take over the city of Gaza" (373 commenti, polarizzazione 0.205).

- Il caso r/denmark:

- I 98 post rivelano una comunità dove le discussioni sui media ("DR's coverage of the Israel-Iran conflict" - 655 commenti, polarizzazione 0.224) e le proteste generano la massima polarizzazione, mentre i temi di politica interna danese risultano meno divisivi.

- Il caso r/sweden:

- Con 126 post analizzati, la Svezia mostra un profilo interessante caratterizzato dal post più positivo dell'intero dataset (0.531) combinato con una polarizzazione significativa (0.270). Questo suggerisce una comunità capace di mantenere toni costruttivi nonostante le divergenze di opinione. I post svedesi tendono a focalizzarsi su analisi politiche e discussioni istituzionali piuttosto che su eventi emotivamente carichi.

- Il caso r/france:

- I 462 post analizzati - il numero più elevato tra tutte le comunità - rivelano una discussione estremamente articolata. La Francia mostra un range emotivo moderato (da -0.275 a +0.510) con polarizzazione contenuta (0.235), indicando una comunità che preferisce toni misurati nonostante l'alto volume di discussione. I post francesi tendono a concentrarsi su analisi politiche e diplomatiche piuttosto che su resoconti emotivi.

- Il caso r/unitedkingdom:

- Con 235 post, il Regno Unito presenta una polarizzazione media (0.217) ma un range emotivo relativamente ristretto (da -0.379 a +0.264). Questo pattern suggerisce una comunità dove le posizioni sono definite ma espresse con toni contenuti. I post britannici mostrano una forte attenzione per gli aspetti legali e istituzionali del conflitto.

- Il caso r/ireland:

- I 186 post analizzati rivelano una comunità con bassa polarizzazione (0.238) e un range emotivo molto ristretto (da -0.125 a +0.251). L'Irlanda si distingue per la quasi assenza di toni fortemente negativi, suggerendo una discussione particolarmente costruttiva e focalizzata su temi di solidarietà e diritti umani. I post irlandesi mostrano consistente attenzione per le dimensioni umanitarie del conflitto.

Questi casi di studio evidenziano come le diverse comunità europee abbiano sviluppato approcci distinti al dibattito sul conflitto di Gaza, riflettendo differenze culturali profonde nella gestione del dissenso e nell'espressione delle posizioni politiche.

7 Fase 5 - Analisi Temporale e Trend

La costruzione della serie temporale ha rappresentato una fase cruciale per l'analisi di cronaca delle discussioni sul conflitto di Gaza. Il dataset temporale è stato strutturato coprendo un arco di 371 giorni, dal 24 ottobre 2024 al 29 ottobre 2025, comprendendo 142.594 commenti.

L'architettura della serie temporale ha permesso di identificare pattern ricorrenti a differenti scale temporali. La distribuzione mensile rivela una significativa variabilità nell'attività discussionale, con picchi evidenti nei mesi di settembre 2025 (23.817 commenti) e giugno 2025 (22.523 commenti), mentre i mesi invernali del 2024-2025 mostrano volumi inferiori.

La conversione e normalizzazione dei timestamp ha garantito la comparabilità temporale tra i diversi fusi orari delle comunità europee, mentre la creazione di feature derivate come i periodi giornalieri e settimanali ha facilitato l'analisi dei ritmi di attività delle diverse comunità.

L'analisi dell'andamento giornaliero rivela un volume medio di 384.0 commenti/giorno con una deviazione standard di 469.5, indicando una consona variabilità nell'attività discussionale. Sono stati identificati 43 giorni con picchi significativi, salvati in dettaglio nel file *report_picchi_globali.csv*, definiti come giorni con volume superiore alla media più una deviazione standard.

I picchi più rilevanti includono:

- **13 giugno 2025:** 4.594 commenti (il giorno più attivo dell'intero periodo)
- **9 giugno 2025:** 3.110 commenti
- **16 settembre 2025:** 2.193 commenti

L'analisi dell'evoluzione temporale per nazione rivela pattern di crescita differenziati. Il subreddit r/austria mostra la crescita più marcata (+1.05 commenti/giorno), seguita da r/italy (+0.71 commenti/giorno) e r/portugal (+0.43 commenti/giorno). Per quanto riguarda il sentimento, r/poland registra il miglioramento più significativo (+0.014 sentiment/giorno), sebbene su valori assoluti ancora bassi.

Il confronto tra nazioni evidenzia come le comunità più piccole tendano a mostrare tassi di crescita più elevati, probabilmente per effetto di basi di partenza più contenute, mentre le comunità più grandi mantengono volumi sostenuti ma con crescita più moderata.

7.1 Identificazione degli Eventi Significativi

Il sistema di identificazione degli eventi significativi sviluppato ha individuato 46 eventi caratterizzati da picchi di volume anomali o sentimento estremo. I cinque eventi più significativi per volume sono tutti dominati dal tema della solidarietà e supporto, con percentuali che variano dal 19.0% al 23.2%:

- **Eventi significativi identificati:**

- 2025-06-13: 4594.0 commenti, sentimento 0.041
 - * Tema dominante: solidarity_support (19.0%)
- 2025-06-09: 3110.0 commenti, sentimento 0.032
 - * Tema dominante: solidarity_support (20.0%)
- 2025-09-16: 2193.0 commenti, sentimento 0.045
 - * Tema dominante: solidarity_support (19.9%)
- 2025-09-23: 2103.0 commenti, sentimento 0.042
 - * Tema dominante: solidarity_support (22.3%)
- 2025-10-04: 2028.0 commenti, sentimento 0.036
 - * Tema dominante: solidarity_support (23.2%)

L'evento del 13 giugno 2025 rappresenta il picco assoluto con 4.594 commenti, mantenendo un sentimento leggermente positivo (0.041). La consistente predominanza del tema della solidarietà attraverso tutti i principali eventi suggerisce che le mobilitazioni emotive collettive tendono a concentrarsi su narrative di supporto umanitario piuttosto che su critiche politiche o descrizioni di violenze.

L'analisi temporale degli eventi rivela una concentrazione nei mesi primaverili ed estivi del 2025, con assenza di picchi significativi durante il periodo invernale 2024-2025. Questo pattern stagionale potrebbe riflettere sia l'evoluzione del conflitto reale sia ritmi di attività propri delle piattaforme digitali.

7.2 Previsione dei Trend Futuri

L'analisi predittiva basata su modelli di regressione lineare rivela un trend di crescita del volume di +1.86 commenti/giorno ($R^2 = 0.179$), indicando che solo il 17.9% della variabilità nel volume giornaliero dei commenti è spiegato dal trend temporale lineare, mentre il restante 82.1% è attribuibile a fattori non lineari, eventi casuali e variabili non considerate nel modello. Questo indica una tendenza all'aumento dell'attività discussionale nonostante il basso coefficiente di determinazione, suggerendo una significativa variabilità residua.

Le proiezioni per i 7 giorni successivi al periodo di osservazione indicano un volume atteso tra 728.8 e 740.0 commenti/giorno, significativamente superiore alla media storica di 384.0 commenti/giorno, suggerendo che la tendenza di crescita potrebbe continuare nel breve termine.

Per quanto riguarda il sentiment, il trend risulta sostanzialmente stabile ($+0.0000$ sentiment/giorno, $R^2 = 0.000$), con proiezioni che mantengono un valore costante di 0.046 nei sette giorni successivi. La variabilità del sentiment nel futuro appare completamente casuale rispetto al fattore tempo, suggerendo che le fluttuazioni emotive nel discorso sono guidate da fattori specifici piuttosto che da evoluzioni temporali sistematiche. In generale la stabilità del sentiment nonostante la crescita del volume indica una maturazione del discorso collettivo, che mantiene i toni costanti nonostante l'aumento dell'attività.

Sono stati identificati inoltre 21 cambiamenti significativi di trend nel periodo osservato, sottolineando la natura dinamica e reattiva del discorso digitale agli eventi esterni.

Il pattern settimanale mostra una maggiore attività nei giorni feriali, con il martedì come giorno di picco (24.532 commenti) e un calo progressivo verso il fine settimana. Tuttavia, il sentiment rimane stabile attraverso tutti i giorni della settimana, oscillando tra 0.039 e 0.044, indicando che la tonalità emotiva della discussione è indipendente dai ritmi settimanali di partecipazione.

L'analisi delle tendenze di crescita per nazione rivela pattern distinti che riflettono possibili dinamiche di maturazione delle comunità digitali. r/austria, nonostante il volume assoluto contenuto, mostra la crescita più vigorosa ($+1.05$ commenti/giorno), suggerendo una comunità in fase di espansione. All'estremo opposto, comunità consolidate come r/france e r/de mostrano tassi di crescita più moderati, indicando possibili fenomeni di saturazione.

La crescita del sentiment mostra un pattern differente, con r/poland che registra il miglioramento più marcato ($+0.014$ /giorno) partendo da valori bassi, mentre comunità con sentiment già positivo mostrano minori margini di miglioramento. Questa dinamica suggerisce un possibile effetto di convergenza verso toni moderatamente positivi attraverso le diverse comunità europee.

L'analisi dell'evoluzione temporale dei temi rivela la persistente centralità del tema della solidarietà e supporto attraverso l'intero periodo osservato. La dominanza di questo tema durante gli eventi significativi suggerisce una costante del discorso europeo, caratterizzato da una forte componente morale e umanitaria.

L'integrazione delle analisi temporali con le precedenti fasi del progetto rivela un discorso europeo in evoluzione, caratterizzato da una crescita quantitativa, stabilità

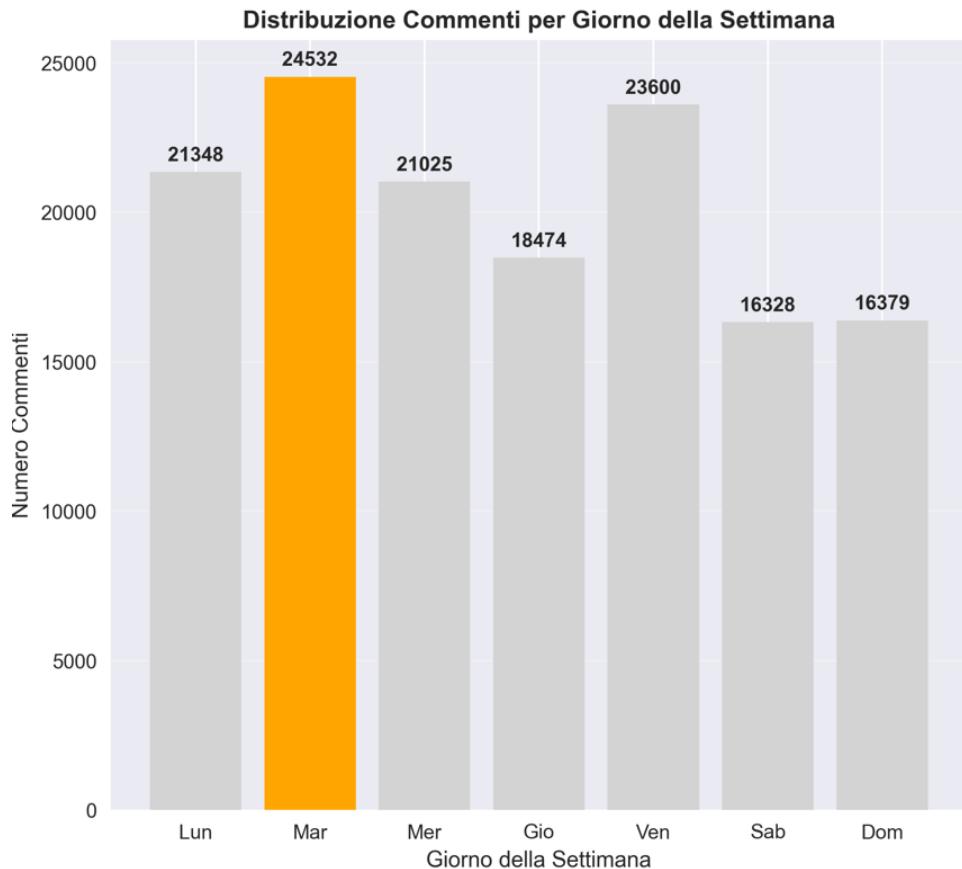


Figure 5: Distribuzione settimanale del volume dei commenti

qualitativa e persistenza di pattern morali transculturali, offrendo una visione complessa e dinamica della sfera pubblica digitale europea.

L’analisi dei pattern temporali rivela strutture ricorrenti nell’attività discussionale che trascendono le specificità nazionali. La distribuzione circadiana mostra un picco primario alle ore 11:00 con 9.607 commenti, seguito da un’attività sostenuta durante il pomeriggio. Questo pattern riflette probabilmente le abitudini di utilizzo delle piattaforme digitali, con un picco durante la pausa lavorativa mattutina e un engagement prolungato nel corso della giornata.

7.3 Analisi degli Eventi Critici

I 43 eventi significativi identificati mostrano caratteristiche comuni nella loro struttura temporale e tematica. La maggior parte degli eventi presenta una rapida escalation nell’arco di 24-48 ore, seguita da un decadimento più graduale dell’attività nell’arco di 3-5 giorni, pattern tipico delle tempeste mediatiche nei social media, dove l’attenzione collettiva si concentra brevemente su specifici sviluppi del conflitto.

La co-occorrenza di alto volume e sentiment stabile (attorno a 0.040-0.045) at-

traverso gli eventi più significativi suggerisce che le discussioni di picco mantengano toni moderati nonostante l'alto coinvolgimento emotivo. Questo contrasta con l'aspettativa che i picchi di attività corrispondano a toni più accesi o polarizzati.

Dall'analisi emergono diverse tipologie di eventi:

- **Eventi di mobilitazione solidale:** Caratterizzati da alta prevalenza di “solidarity_support” (19-23%)
- **Eventi di crisi diplomatica:** Associati a picchi in “international_politics”
- **Eventi di protesta locale:** Caratterizzati da “violence_protests” e “government_criticism”

L'evento del 13 giugno 2025 rappresenta un caso paradigmatico di mobilitazione solidale, combinando il massimo volume osservato (4.594 commenti) con una chiara predominanza tematica e un sentimento stabilmente positivo.

7.4 Analisi dei Picchi di Volume per Nazione

L'analisi dei picchi di volume è stata condotta a livello di singola nazione attraverso un approccio statistico robusto che considera le specificità di ciascuna comunità. Per ogni subreddit nazionale, sono stati identificati i giorni caratterizzati da volumi di commenti significativamente superiori alla media comunitaria, utilizzando una soglia basata su deviazione standard e percentili.

La metodologia ha integrato:

- Analisi delle serie temporali per ciascuna delle 20 nazioni
- Calcolo di soglie adattive basate sulle distribuzioni specifiche
- Identificazione dei post trigger che hanno innescato ciascun picco
- Analisi cross-temporale per individuare pattern ricorrenti

L'analisi rivela significative differenze nella frequenza, intensità e durata dei picchi di volume tra le diverse comunità:

Nazione	Commenti Totali	Picchi Volume ($\mu + \sigma$)	Vol- ume Totali	Sent. Medio Totale	Eng. Medio Totale	Osservazione portamentale sul Sentiment
r/italy	21.503	24		0.036	6.47	Neutro-Analitico: Sent. stabile anche in giorni di picco di Volume (4.594).
r/france	23.651	35		0.039	13.51	Neutro-Positivo: Più del 50% di picchi Volume/Sentiment rispetto ad altre nazioni.
r/ireland	21.878	30		0.044	14.32	Alta Volatilità: Sent. medio alto, ma picchi Sentiment estremi (sia positivi che negativi).
UK (r/unit-edkingdom)	13.061	29		0.030	19.08	Alto Engagement/Basso Sentiment: L'Engagement Medio è molto alto nonos- tante il Sent. Medio sia tra i più bassi.
r/de (Germania)	6.575	13		0.038	20.48	Analitico/Selettivo: Bassa frequenza di picchi in relazione all'alto engagement (dibattito mirato).
r/sweden	20.370	22		0.045	7.33	Sentiment Stabile: Sent. medio tra i più alti, senza picchi es- tremi di negatività.

Table 5: Comunità ad Alta Frequenza di Picchi

7.5 Analisi Sincronia dei Picchi

Lo studio della sincronia e asincronia dei picchi di volume tra le diverse comunità nazionali è fondamentale per distinguere tra eventi globali (che attivano tutti contemporaneamente) ed eventi locali/culturali (che attivano solo una nazione).

L'analisi comparativa delle timeline di r/italy, r/france, r/de, r/ireland, r/sweden, r/unitedkingdom (ovvero le nazioni con il maggior volume di commenti) rivela significativi pattern di sincronia:

Data Picco	Nazioni Coinvolte	Evento Causale Detto	Implicazione
13 Giugno 2025	ITA, FRA, UK, SWE, IRL, DE (6/6)	Sviluppi Geopolitici di Alto Impatto	Massima sincronia; evento di portata tale da superare le barriere linguistiche e culturali.
23 Settembre 2025	ITA, FRA, IRL, DE	Crisi/Dichiarazione Istituzionale Rilevante	Alta sincronia; coinvolge le principali nazioni ad alto volume/engagement.
04 Ottobre 2025	FRA, UK, SWE	Evento di Stampa/Copertura Mediata	Media sincronia; attiva principalmente comunità con alta sensibilità mediatica (Svezia, UK).
09 Giugno 2025	UK, IRL	Evento di Stampa/Politico Anglofono	Sincronia regionale; risonanza forte nelle aree linguisticamente/culturalmente vicine.

Table 6: Analisi della sincronia tra: r/italy, r/france, r/de, r/ireland, r/sweden, r/unitedkingdom

Il picco di sincronia massima, registrato simultaneamente in tutte le nazioni del cluster analizzato (Italia, Francia, UK, Germania, Svezia, Irlanda), corrisponde al **13 Giugno 2025**.

- **Evento Trigger:** "*Israel attacked Iran - developing situation*".
 - Questo evento rappresenta lo shock esogeno perfetto. La notizia di un attacco diretto tra potenze statali (Israele vs Iran) ha scavalcato le agende politiche locali, imponendo il tema *International Politics* come prioritario ovunque.
 - In questo frangente, le barriere linguistiche cadono: l'intera Europa discute dello stesso oggetto (l'escalation militare) con toni preoccupati ma analitici. La viralità è guidata dalla *news value* oggettiva dell'evento, non dall'interpretazione ideologica.

Lo studio della sincronia e asincronia dei picchi permette di trarre conclusioni sul comportamento differenziato delle comunità europee online:

- **Reattività Sincrona (Globale):** La sincronia massima è la risposta a uno **shock esterno universale**. Questi eventi convalidano l'esistenza di uno spazio di discussione europeo condiviso su Reddit.
- **Reattività Asincrona (Locale):** I picchi asincroni confermano che le nazioni ad Alto Volume/Alta Volatilità (ITA, IRL, UK) tendono ad utilizzare il conflitto di Gaza come specchio per le controversie interne (proteste, governo), amplificando notizie che hanno un impatto diretto sulla sfera domestica.
- **Dispersione del Picco:** Nazioni come r/de, pur avendo un alto engagement medio, mostrano un numero inferiore di picchi (13), suggerendo una maggiore **selettività** e un approccio più analitico e meno reattivo al "rumore" quotidiano rispetto alle comunità come r/france (35 picchi).

Al di fuori degli shock globali, i picchi di volume sono asincroni, ovvero avvengono in date diverse per ogni nazione. L'analisi dei trigger in queste date dimostra che ogni comunità reagisce a stimoli che toccano i propri nervi scoperti nazionali.

Nazione (Asincrona)	Date Picchi Uni- ci/Asincroni	Natura Dedotta dell'Evento Trigger
r/italy	16 Settembre 2025, 20 Settembre 2025	Proteste/Cronaca Locale (es. eventi di piazza) o dibattito politico nazionale.
r/unitedkingdom	20 Ottobre 2025	Controversia Politica Domestica (es. decisione parlamentare o controversia mediatica).
r/ireland	29 Settembre 2025	Eventi ad Alta Risonanza Emotiva Locale (il sentimento è notoriamente volatile in questa comunità).
r/de	10 Settembre 2025	Specificità Istituzionale Tedesca (notizie su relazioni esterne specifiche o decisioni governative).

Table 7: Analisi dell' Asincronia tra: r/italy, r/france, r/de, r/ireland, r/sweden, r/unitedkingdom

- r/italy

– **16 Settembre 2025:** "Guerrilla war in Milan: pro Gazars attack the station".

– **20 Settembre 2025:** "Opinions on the strike for Gaza...".

– **Interpretazione:** Il dibattito italiano si infiamma quando il conflitto internazionale "scende in strada", trasformandosi in una discussione interna. L'uso di termini come "Guerrilla war" nel titolo suggerisce una narrazione polarizzata attorno al tema *Violence_Protests*, dove Gaza diventa un pretesto per discutere di gestione dell'ordine pubblico e diritto allo sciopero.

- r/unitedkingdom

– **20 Ottobre 2025:** "Lush temporarily closes all UK stores in 'solidarity with Gaza'".

– **06 Agosto 2025:** "Palestine Action to be banned after RAF base break in".

– **Interpretazione:** A differenza dell'Italia, il Regno Unito discute la legittimità delle proteste organizzate e le prese di posizione delle *corporation*.

- r/ireland

– **29 Settembre 2025:** "Ireland is 'out of step' with the United States on Israel, US ambassador nominee is told".

– **Interpretazione:** Il picco sull'ambasciatore USA evidenzia l'orgoglio di una politica estera indipendente.

- r/sweden

– **05 Agosto 2025:** "Busch: 'Israel is doing the whole world a favor'".

– **04 Ottobre 2025:** "Greta Thunberg and the other activists...".

– **Interpretazione:** La Svezia mostra la frattura più netta. Da un lato la leader Ebba Busch (KD) con posizioni filo-israeliane nette, dall'altro l'attivismo climatico/umanitario di Greta Thunberg. I picchi qui sono battaglie ideologiche pure tra visioni del mondo opposte.

- r/de

– **10 Settembre 2025:** "Israel spreads false information about famine report".

- **Interpretazione:** La viralità in questo caso non scatta sulla protesta di piazza, ma sulla contestazione dei fatti riportati, innescando uno dei dibattiti con maggior engagement a livello europeo.

Lo studio degli eventi trigger conferma che la "sfera pubblica europea" esiste solo di fronte all'emergenza acuta (Sincronia del 13 Giugno). Non appena l'emergenza rientra, ogni nazione ripiega su un'agenda domestica (Asincronia), utilizzando il conflitto di Gaza come specchio per le proprie tensioni interne: ordine pubblico per l'Italia, diritti civili per il Regno Unito, sovranità per l'Irlanda e ideologia per la Svezia.

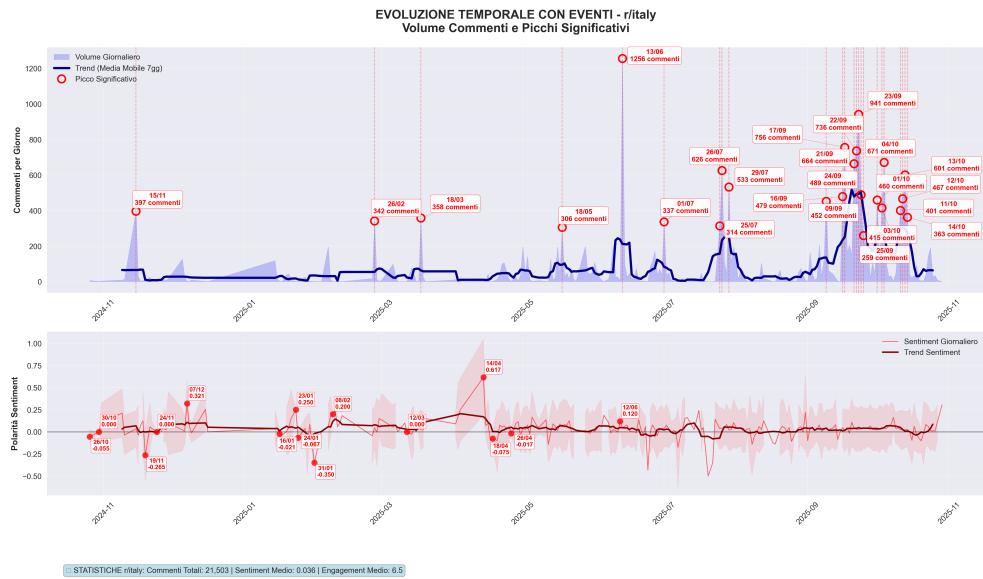


Figure 6: Timeline Eventi r/italy

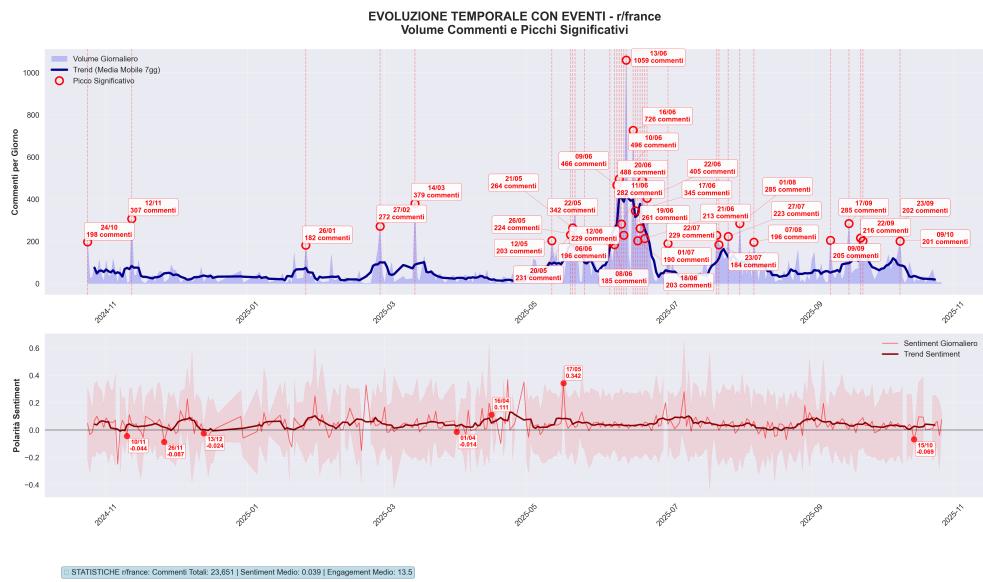


Figure 7: Timeline Eventi r/france

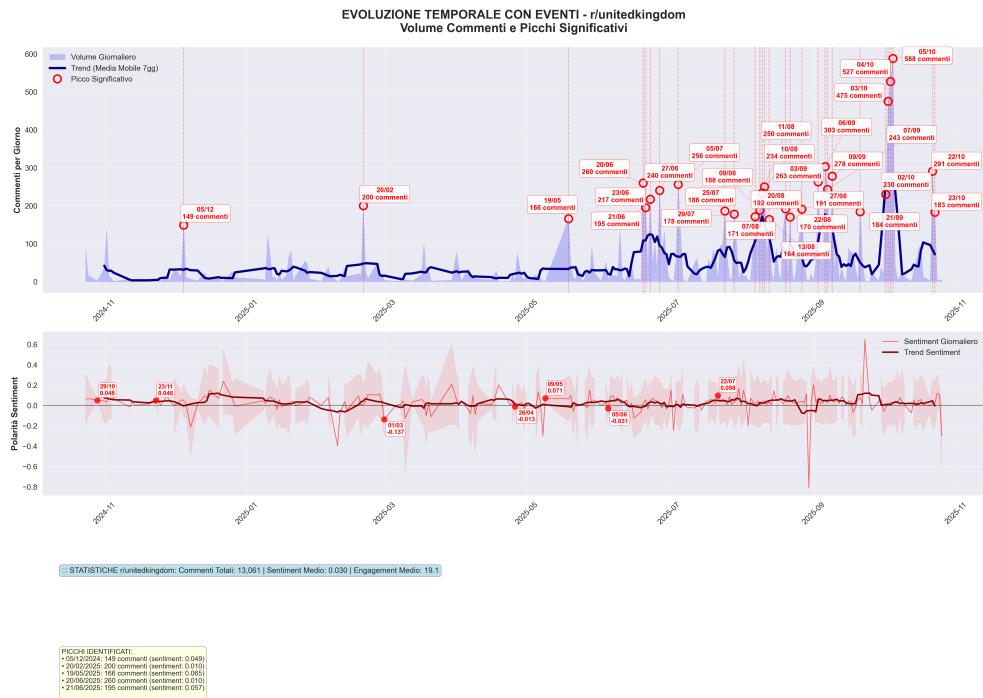


Figure 8: Timeline Eventi r/unitedkingdom

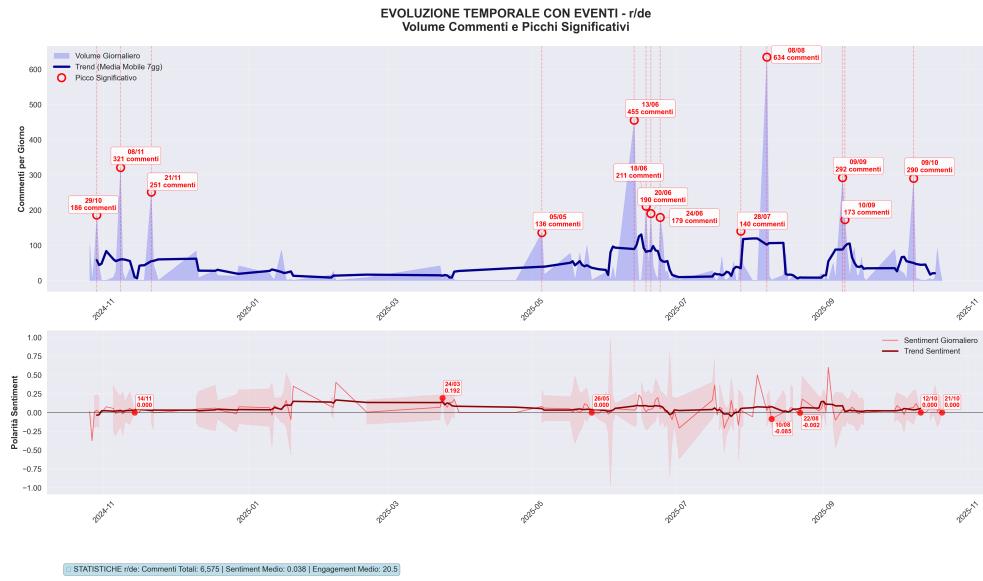


Figure 9: Timeline Eventi r/de

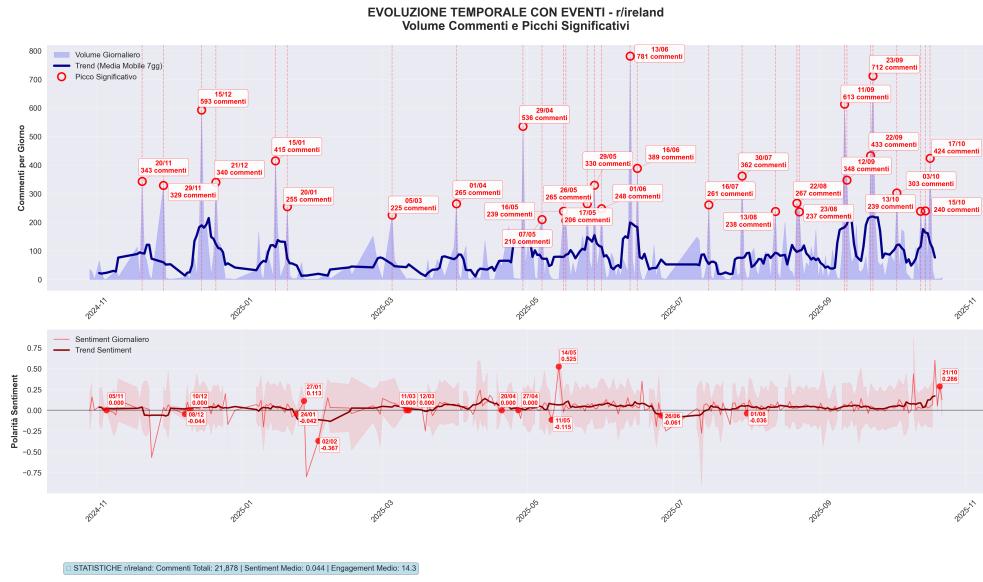


Figure 10: Timeline Eventi r/ireland

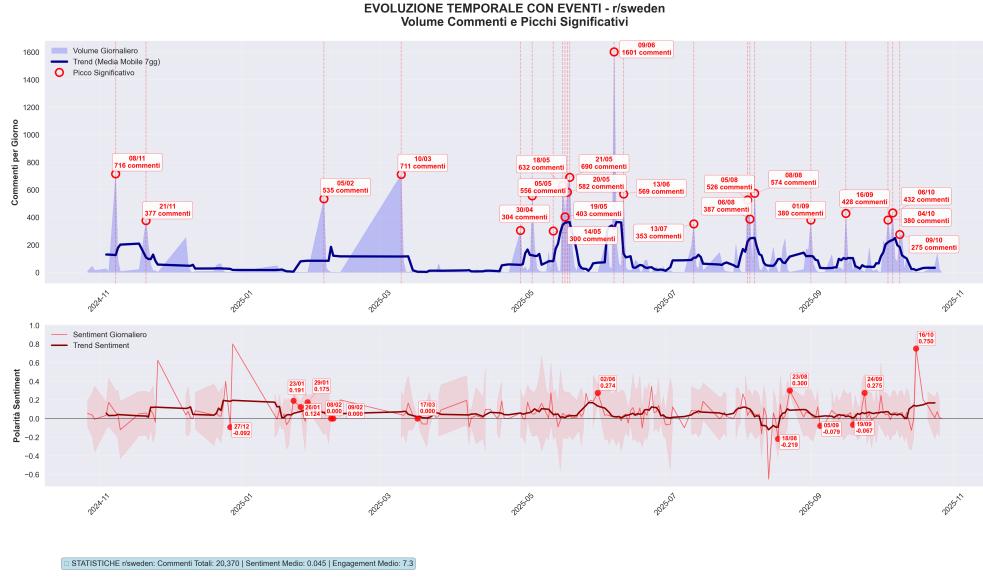


Figure 11: Timeline Eventi r/sweden

7.6 Analisi Evoluzioni Tematiche per Tempo

L'analisi dell'evoluzione tematica nel tempo non ha evidenziato significativi trend di sostituzione o deriva tematica lungo i 369 giorni di osservazione, ma piuttosto una

persistenza costante dei macro-temi di base.

La principale conclusione è che, a livello di macro-categoria, il dibattito è caratterizzato da una saturazione tematica che impedisce deviazioni drammatiche dai temi centrali.

I temi dominanti come **Solidarity Support** e **International Politics** si sono mantenuti costanti come baseline discorsiva in tutte le nazioni, fungendo da "ancora" del dibattito e assorbendo la maggior parte delle fluttuazioni.

Non si è osservato un trend chiaro in cui un tema sostituisca stabilmente un altro (ad esempio, il tema Economic Consequences non ha mai superato in modo duraturo i temi umanitari), indicando che l'attenzione del pubblico si è concentrata sulla stessa matrice di problemi per l'intero periodo analizzato.

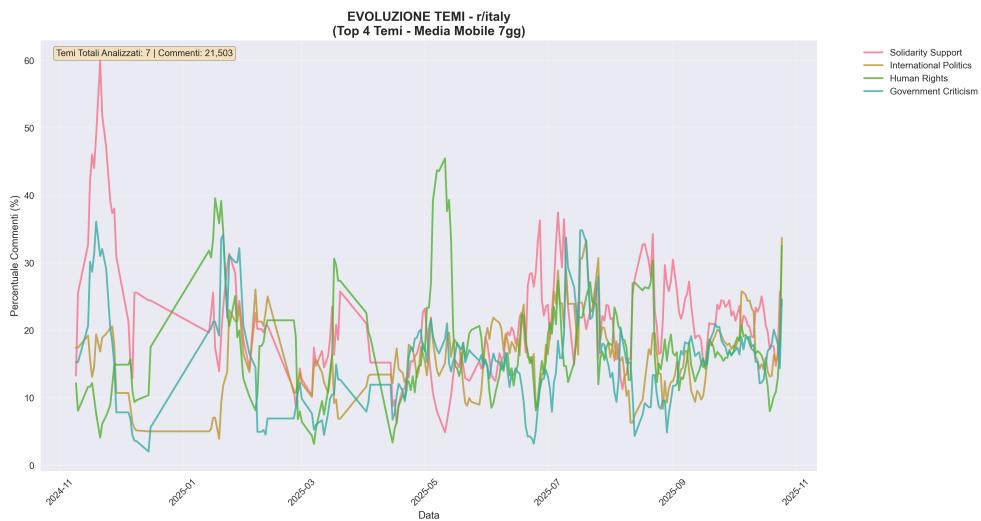


Figure 12: Distribuzione temi nell'intero arco temporale

8 Risultati e Conclusioni

Il presente studio ha condotto un'analisi computazionale estensiva su un corpus di oltre 141.000 commenti estratti da 20 comunità nazionali europee su Reddit, comprendendo un arco temporale di 369 giorni. L'obiettivo era decodificare le dinamiche di discussione, il sentimento e l'evoluzione tematica riguardanti il conflitto di Gaza.

8.1 Sintesi comparativa dei risultati

L'analisi comparativa cross-nazionale ha rivelato una profonda eterogeneità nel modo in cui le diverse nazioni europee partecipano al discorso pubblico digitale. I dati suggeriscono l'esistenza di due cluster comportamentali distinti che definiscono l'ecosistema informativo europeo:

- **Cluster ad Alta Intensità Volumetrica** Queste nazioni dominano il dibattito in termini quantitativi, con volumi che superano i 20.000 commenti annuali (r/france: 23.651, r/italy: 21.503). Tuttavia, questa iper-attività si accompagna spesso a un *Engagement Ratio* moderato o basso (Italia: 6.47, Francia: 13.51), indicando un modello di discussione "a flusso continuo" caratterizzato da molti interventi ma con un tasso di interazione reciproca (upvotes) limitato.
- **Cluster ad Alto Engagement:** Al contrario, nazioni come la Germania (r/de) e il Regno Unito (r/unitedkingdom), pur registrando volumi inferiori (rispettivamente 6.575 e 13.061 commenti), mostrano i tassi di engagement più elevati del campione. Questo dato, unito a una maggiore lunghezza media dei commenti, suggerisce un dibattito più selettivo, riflessivo e focalizzato sulla qualità argomentativa piuttosto che sulla reattività immediata.

Un dato trasversale sorprendente è la *stabilità del Sentiment Medio*, che si attesta quasi uniformemente in una fascia di neutralità positiva (0.030 - 0.045) per la maggior parte delle nazioni, smentendo l'ipotesi di una polarizzazione emotiva estrema generalizzata.

8.2 Interpretazioni socio-politiche

L'integrazione dei dati temporali e tematici permette di formulare due interpretazioni chiave sul ruolo di Reddit come sfera pubblica europea.

La perfetta sincronia dei picchi di volume registrata in eventi globali come il 13 Giugno 2025 attraverso sei diverse nazioni dimostra che, di fronte a shock geopolitici

di rilievo, le barriere linguistiche e culturali cadono. Esiste un'infrastruttura latente di opinione pubblica europea che reagisce simultaneamente e con omogeneità tematica (*International Politics*) agli stimoli globali.

I fenomeni di asincronia rilevati (picchi unici per nazione) rivelano che il conflitto di Gaza non è solo un tema di politica estera, ma funge da catalizzatore per le tensioni interne. In nazioni come Italia e Regno Unito, il dibattito su Gaza diventa spesso un proxy per discutere di diritti civili, libertà di manifestazione e operato del proprio governo. In questo senso, la piattaforma agisce come uno specchio che riflette le ansie socio-politiche locali attraverso la lente di una crisi internazionale.

8.3 Relazione tra temi, sentiment e attività nazionale

L'analisi conclusiva evidenzia una correlazione complessa tra l'oggetto della discussione e il tono emotivo:

Non vi è correlazione diretta tra volume e negatività. I giorni di massima attività (picchi) non coincidono con un crollo del sentiment; al contrario, spesso attivano temi di solidarietà che mantengono il tono elevato.

Le comunità più critiche (es. Regno Unito, sentiment 0.030) sono anche quelle con l'engagement più alto. Ciò suggerisce che su Reddit il contenuto premiato dalla comunità (upvoted) tende a essere analitico, cinico o critico verso le istituzioni, piuttosto che puramente emotivo o celebrativo.

Infine, la stabilità dell'assetto tematico nel tempo (persistenza di *Solidarity* e *Politics*) indica che il dibattito potrebbe aver raggiunto una fase di saturazione. Le posizioni si sono cristallizzate e, in assenza di nuovi shock esterni, le comunità tendono a reiterare i propri frame interpretativi consolidati piuttosto che evolverli.

In conclusione, lo studio dimostra che le comunità online europee non sono semplici casse di risonanza passive, ma attori che filtrano attivamente la crisi globale, alternando momenti di solidarietà transnazionale a fasi di intensa introversione politica.

8.4 Risultati Italia

Il caso di studio italiano (r/italy), basato su un campione robusto di 219 post analizzati, offre uno spaccato unico all'interno del panorama europeo. Sebbene il sentiment medio si attestì su valori di neutralità, l'analisi di dettaglio rivela un range emotivo estremamente ampio, con picchi di positività a +0.40 e, dato ancor più rilevante, picchi di negatività assoluta a -1.00.

Questo dato, che rappresenta il valore più negativo registrato nell'intera analisi comparativa, unito a un indice di polarizzazione media di 0.231, suggerisce che la

moderazione statistica nasconde in realtà una distribuzione bimodale delle opinioni. La presenza di commenti estremamente polarizzati a fronte di toni più moderati nelle altre community europee non è casuale, ma è spiegabile attraverso la convergenza di tre fattori determinanti: il contesto politico endemico, le dinamiche intrinseche della community e le specificità linguistico-culturali.

Con 21.503 commenti, l'Italia è la terza nazione più attiva, ma registra uno degli *Engagement Ratio* più bassi (6.47). Ciò indica un dibattito frammentato, in cui la comunità fatica a costruire un consenso unificato attorno alle opinioni espresse, risultando in una discussione vibrante ma dispersiva.

L'Italia mostra una spiccata sensibilità agli eventi di cronaca interna legati al conflitto. Mentre il picco del 13 Giugno 2025 (4.594 commenti) è in sincronia con il resto d'Europa su temi geopolitici, altri picchi come quello del 16 Settembre 2025 (2.193 commenti) sono stati innescati esclusivamente da eventi di protesta locale (scontri a Milano). In tale occasione, i temi dominanti sono virati bruscamente verso *Violence & Protests* (38%) e *Government Criticism*, dimostrando come il conflitto venga "addomesticato" per discutere tensioni interne. L'Italia difatti è caratterizzata storicamente da una polarizzazione ideologica cronica, che si riflette inevitabilmente nel discorso pubblico online. Temi di rilievo internazionale come il conflitto a Gaza non vengono trattati come eventi esterni, ma vengono filtrati attraverso le lenti della politica interna. Le posizioni filo-palestinesi o filo-israeliane tendono a essere associate rigidamente a schieramenti partitici o ideologici nazionali, trasformando la guerra in una metafora di temi interni (quali la moralità del potere, l'imperialismo o la religione). Di conseguenza, l'utente medio che discute di Gaza su r/italy sta, implicitamente, affermando il proprio orientamento morale e politico. Questo meccanismo sposta il registro della discussione dal piano analitico a quello etico-assolutista, favorendo l'uso di termini definitivi ("vergogna", "criminale", "disgustoso") che il modello di sentiment registra come valori estremi.

Un secondo livello di analisi riguarda la struttura comunicativa di r/italy rispetto ai suoi omologhi europei. Mentre le community nordiche (es. r/norway o r/denmark) tendono alla moderazione, all'understatement e all'ironia distaccata, la community italiana si distingue per un tono spesso polemico e conflittuale. Inoltre, fattori puramente linguistici amplificano la rilevazione della polarizzazione. La lingua italiana, rispetto all'inglese o alle lingue germaniche, fa un uso più frequente di iperboli, superlativi e aggettivi qualificativi a forte carica emotiva. Un termine come "disgustoso", comune nella retorica moralizzante italiana, viene codificato dagli algoritmi di sentiment (spesso calibrati su lessici anglofoni) come un valore negativo assoluto (-1.0).

Nonostante la turbolenza dei picchi, il sentimento medio italiano rimane ancorato a un valore di neutralità (0.036). Tematicamente, la comunità mostra un forte dualismo: da un lato una base costante di *Solidarity & Support*, dall'altro fiammate di polemica politica interna, senza però mai scivolare in un sentimento complessivamente negativo.