RdS
RICERCA DI SISTEMA

ENEA
AGENZIA NAZIONALE PER LE NUOVE TECNOLOGIE, L'ENERGIA E LO SVILUPPO ECONOMICO SOSTENIBILE

MINISTERO DELL'AMBIENTE E DELLA SICUREZZA ENERGETICA

Ricerca di Sistema elettrico

Servizi social per le CER: Studio delle comunità energetiche mediante il discovery e l'analisi di reti sociali (LA1.22)

Emiliano Casalicchio, Mattia Samory

SAPIENZA
UNIVERSITÀ DI ROMA

RdS_PTR 22-24_PR1.7_LA1.22_340

Servizi social per le CER: Applicazione di tecniche di Process mining a comunità energetiche reali, ricognizione dei pattern di interazione e comparazione tra le diverse tecniche

LA1.22 - Servizi social per le CER: Studio delle comunità energetiche mediante il discovery e l'analisi di reti sociali

E.Casalicchio, M.Samory (Dipartimento di Informatica, Università di Roma La Sapienza)

Dicembre 2024

Report Ricerca di Sistema Elettrico

Accordo di Programma Ministero dell'Ambiente e della Sicurezza Energetica - ENEA Piano Triennale di Realizzazione 2022-2024

Obiettivo: Decarbonizzazione

Progetto: Tema di ricerca 1.7 – Tecnologie per la penetrazione efficiente del vettore elettrico negli usi finali

Linea di attività: *LA1.22*

Responsabile del Progetto: Dott.ssa Claudia Meloni, ENEA

Responsabile del Work Package: Ing. Angelo Frascella, ENEA

Responsabile Linea di Attività: Prof. Emiliano Casalicchio, ENEA

Mese inizio previsto: 13

Mese inizio effettivo: 13

Mese fine previsto: 36

Mese fine effettivo: 36

Il presente documento descrive le attività di ricerca svolte all'interno dell'Accordo di collaborazione: "Servizi social per le CER: Studio delle comunità energetiche mediante il discovery e l'analisi di reti social". Si ringrazia per la collaborazione alle attività svolte il dott. Daniele Marretta

Indice

1 Risultati attesi 4
2 Risultati ottenuti 5
2.1 Misura del raggiungimento dei risultati 5
2.2 Benefici per il sistema elettrico nazionale ed i suoi utenti 5
3 Prodotti attesi 6
4 Prodotti sviluppati 7
4.1 Modalità di accesso ai prodotti software sviluppati 7
5 Analisi degli scostamenti su attività e risultati 8
6 Sintesi delle attività svolte 9
7 Dettaglio delle attività svolte 10
7.1 Selezione delle tecniche per l'analisi di dataset 10
7.2 Implementazione di uno strumento software per la raccolta di dati 10
7.3 Creazione dei dataset 11
7.4 Analisi delle relazioni tra utenti e contenuti 12
7.5 Analisi del contenuto dei messaggi 15
7.5.1 Attinenza alle chiavi di ricerca 15
7.5.2 Analisi dei contenuti dei messaggi 19
7.5.3 Analisi del sentimento 21
8 Contributo delle eventuali consulenze alle attività sopra descritte 24
9 Pubblicazioni scientifiche 25
10 Eventi di disseminazione 26

Indice delle figure

Figura 1 Esempio di rappresentazione JSON di un messaggio 11
Figura 2 Elenco delle chiavi di ricerca identificate 12
Figura 3 Relazioni utente-utente e utente-hastag per i mesaggi raccolti da X.....13
Figura 4 Particolare del grafo di Figura 3 13
Figura 5 Grafo delle relazioni utente-utente per i messaggi raccolti da Meta..... 14
Figura 6 Particolare del grafo in Figura 5..... 14
Figura 7 Grafo delle relazioni utenti-hashtag per i messaggi raccolti da Meta 15
Figura 8 Attinenza dei messaggi alle chiavi di ricerca per i messaggi di X 16
Figura 9 Attinenza dei messaggi alle chiavi di ricerca per i messaggi di Meta.....17
Figura 10 Rilevanza delle chiavi di ricerca per i messaggi di X.....18
Figura 11 Rilevanza delle chiavi di ricerca per i messaggi di Meta 18
Figura 12 Rappresentazione bidimensionale dei cluster di messaggi di X.....19
Figura 13 Rappresentazione bidimensionale dei cluster di messaggi di Meta 20
Figura 14 Distribuzione del sentimento degli utenti rispetto alle parole chiavi per la piattaforma X.....22
Figura 15 Distribuzione del sentimento degli utenti rispetto alle parole chiavi per la piattaforma Meta 23

Indice delle tabelle

Tabella 1 Le cinque principali topic trattate sulla piattaforma X.....21
Tabella 2 Le cinque principali topic trattate sulla piattaforma Meta.....21

1 Risultati attesi

La linea di attività produrrà

- Strumenti software per l'analisi dei contenuti dei social media al fine di individuare le Energy Community
- Un dataset di messaggi/informazioni raccolti dai social media relativi alle Energy Community
- Analisi delle dinamiche delle Energy Community

Le librerie sviluppate e i db risultanti dalla raccolta saranno resi disponibili. A fine progetto sarà prodotto un rapporto tecnico.

2 Risultati ottenuti

I risultati ottenuti dal progetto sono i seguenti:

- **R1.** Definizione di una metodologia e implementazione di uno strumento software per la raccolta di dati sulle Energy Community a partire da piattaforme di social media (X e Meta nello specifico)
- **R2.** Creazione di un dataset di messaggi riguardanti il dominio delle Energy Community raccolti dalla piattaforma di social media X
- **R3.** Creazione di un dataset di messaggi riguardanti il dominio delle Energy Community raccolti dalla piattaforma di social media Meta
- **R4.** Analisi dei dataset ottenuti mediante tecniche di
    - Natural Language Processing
    - Analisi delle reti complesse
    - Analisi statistica

Le sezioni che seguono descrivono in che misura I risultati sono stati ottenuti ed il beneficio per il sistema elettrico nazionale e i suoi utenti.

2.1 Misura del raggiungimento dei risultati

Rispetto a quanto previsto dal capitolato i risultati sono stati pienamente raggiunti.

2.2 Benefici per il sistema elettrico nazionale ed i suoi utenti

I risultati del progetto contribuiscono al monitoraggio della percezione della società civile sulle comunità energetiche rinnovabili.

I risultati ottenuti contribuiscono agli obiettivi dal piano Ricerca di Sistema Elettrico Nazionale nella seguente misura:

- La metodologia di raccolta dei dati proposta ed il prototipo di strumento software per la raccolta permette di re-iterare il processo nel tempo monitorando l'evoluzione delle comunità energetiche, dei temi trattati, sia tecnologici che sociali e delle problematiche che emergono dagli utenti stessi. Ciò costituisce una base di conoscenza che istituzioni ed operatori possono utilizzare nel loro processo decisionale atto a realizzare la Decarbonizzazione e la Digitalizzazione ed evoluzione delle reti.
- I dataset creati sono stato uno strumento per identificare e valutare le tecniche di analisi dei dati che permettono appunto di determinare l'esistenza, l'entità e l'evoluzione delle comunità energetiche, gli attori coinvolti, i temi trattati, il sentimento degli utenti rispetto le varie tematiche. I dataset raccolti sono la fotografia dello stato delle comunità energetiche in un breve lasso di tempo, ma come evidenziato nel punto precedente, il processo di raccolta può essere re-iterato e l'analisi riprodotto portando ai benefici sopra elencati.

3 Prodotti attesi

I prodotti software attesi per la LA sono:

- Strumenti software per l'analisi dei contenuti dei social media al fine di individuare le Energy Community
- Un dataset di messaggi/informazioni raccolti dai social media relativi alle Energy Community

4 Prodotti sviluppati

I prodotto software sviluppati nella LA sono in linea con quelli attesi, ed in particolare

1. **SW1:** modulo software per la raccolta di dati sulle Energy Community a partire da piattaforme di social media (X e Meta nello specifico)
2. **SW2:** modulo software per la raccolta di documenti pubblici disponibili sul Web e riguardanti il tema delle Energy Community.
3. **SW3:** Dataset di messaggi riguardanti il dominio delle Energy Community raccolti dalla piattaforma di social media X (ex Twitter)
4. **SW4:** Dataset di messaggi riguardanti il dominio delle Energy Community raccolti dalla piattaforma di social media Meta
5. **SW5:** Moduli software per l'analisi dei dataset

4.1 Modalità di accesso ai prodotti software sviluppati

I prodotti software sviluppati ed i dataset sono stati resi disponibili ad ENEA sotto forma di codice sorgente e file di testo.

I prodotti sviluppati sono rivolti a ricercatori e non ad un'utenza generale.

I moduli software sono stati realizzati in linguaggio Python e sono eseguibili su qualsiasi piattaforma. Essi fanno uso di altre librerie software opensource. Il software, se ENEA lo ritiene opportuno può essere reso opensource e divulgato mediante piattaforme come GitHub.

I dataset, allo stato attuale non possono essere pubblicati in modalità open-access in quanto non anonimizzati. Essi non sono stati anonimizzati per non perdere contenuto informativo essenziale.

5 Analisi degli scostamenti su attività e risultati

Rispetto a quanto stabilito nel capitolato, non si rilevano scostamenti sui risultati ottenuti rispetto a quelli attesi.

6 Sintesi delle attività svolte

Identificazione delle metodologie e gli strumenti per: la determinazione della rilevanza di un testo rispetto ad una determinata chiave di ricerca, necessari per il filtraggio dei messaggi; il topic modeling; la sentiment analysis (risultato R1, §2).

È stato poi realizzato di un modulo software di scraping, basato sulla libreria Selenium, per la ricerca dei messaggi di interesse sulle piattaforme social e loro memorizzazione come oggetti JSON (risultato R1, §2).

Creazione dei dataset (risultati R2 e R3, §2). Sono prima state identificate keyword estratte sia mediante analisi manuale che automatica di pubblicazioni scientifiche. Successivamente, per ogni keyword sono state interrogate le piattaforme social Meta e X e sono stati ottenuti un totale di 200.000 messaggi per ogni piattaforma.

Analisi (risultato R4, §2) A partire dai dataset, sono state: analizzate le relazioni tra utenti, tra argomenti, e tra utenti e argomenti trattati; analizzati i contenuti dei messaggi (distribuzione delle chiavi di ricerca, topic-modeling, sentiment analysis).

7 Dettaglio delle attività svolte

Le attività svolte per raggiungere gli obiettivi del progetto sono:

- Selezione delle tecniche per l'analisi di dataset
- Definizione di una metodologia e implementazione di uno strumento software per la raccolta di dati sulle Energy Community a partire da piattaforme di social media X e Meta
- Creazione di dataset di messaggi
- Analisi dei dataset

7.1 Selezione delle tecniche per l'analisi di dataset

In questa attività sono stati identificati:

- Metodologia e strumenti per la determinazione della rilevanza di un testo rispetto ad una determinata chiave di ricerca, necessari per il filtraggio dei messaggi
- Metodologia e strumenti per l'analisi di similarità tra argomenti trattati nei messaggi ed per la modellazione dei temi trattati
- Metodologia e strumenti per l'analisi dei sentimenti espressi nei messaggi (sentiment analysis).

Per determinare il punteggio di attinenza di un messaggio rispetto alle parole chiave è stato utilizzato l'algoritmo di classificazione BM25[i] che permette di stimare la rilevanza di un documento rispetto a una query specifica. Essa fornisce un punteggio di rilevanza basato sulla frequenza delle parole nel documento, senza considerare l'ordine in cui esse appaiono.

Per analizzare i temi trattati nei messaggi (*topic modeling*) sono stati utilizzate le seguenti tecnologie. SBERT[ii], per rappresentare parole e frasi sotto forma di vettori numerici (detti *embedding*). Riduzione delle dimensioni degli *embedding* mediante UMAP (Uniform Manifold Approximation and Projection)[iii], un algoritmo non-supervisionato usato per la riduzione della dimensionalità dei dati, riducendo i vettori a due o tre dimensioni per facilitarne la visualizzazione. I vettori di dimensione ridotta vengono clusterizzati mediante l'algoritmo HDBSCAN[iv] e successivamente si determinano le parole più significative all'interno dei cluster utilizzando l'algoritmo c-TF-IDF. Per l'implementazione dei modelli e degli algoritmi sopra menzionati si è fatto uso della libreria software opensource BERTtopic[v],[vi].

7.2 Implementazione di uno strumento software per la raccolta di dati

In questa attività sono stati realizzati i prodotti SW1 e SW2 (cf. §4) ovvero moduli di scraping che fanno uso della libreria software opensource Selenium[vii]. Selenium permette di implementare azioni di navigazione di siti web e di automatizzarle. È stato quindi progettato ed implementato un agente scraper che mediante le funzionalità di ricerca messe a disposizione delle piattaforme social immette le query di interesse e raccoglie i risultati della ricerca. I risultati della ricerca vengono memorizzati in un oggetto JSON contenente le seguenti informazioni: *id*, identificativo univoco del messaggio assegnato dalla piattaforma social; *keyword*, la lista di keyword contenute all'interno del messaggio, username, il nome utente dell'autore del messaggio; *tag*, contiene informazioni su specifici destinatari del messaggio; *date*, la data di pubblicazione del messaggio; *hashtag*, gli hashtag associati al messaggio; *mentions*, altri utenti social a cui si fa riferimento nel messaggio; *sentiment*, esprime il sentimento associato al messaggio ed è un valore determinato dalla piattaforma social. Un

esempio di istanza di oggetto JSON rappresentate un messaggio memorizzato nel dataset è il seguente

7.3 Creazione dei dataset

In questa attività vengono realizzati i prodotti SW3 e SW4.

Per ricercare i messaggi sulle piattaforme social e quindi creare i dataset sono state prima identificate una serie di keyword, le quali sono state fornite come input ai moduli scraper


```
{
"id": "@COMANAGEproject2024-04-30T09:30:00.000Z",
"keyword": "citizen energy communities",
"username": "COMANAGE",
"tag": "@COMANAGEproject",
"date": "2024-04-30T09:30:00.000Z",
"text": "On April 25th, @sostAMB participated in the
      #EnergyCommunities Summit hosted by
      @UNEFotovoltaica. They shared insights on citizen
      advisory services from https:// and showcased the
      tools from the COMANAGE project for managing and
      maintaining energy communities. ",
"hashtags": [
      "#energycommunities"
      ],
"mentions": [
      "@sostAMB",
      "@UNEFotovoltaica"
      ],
"sentiment": "neutral"
}
```


Figura 1 Esempio di rappresentazione JSON di un messaggio

descritti in §7.2. La definizione delle keywords è iniziata con un'analisi pubblicazioni riguardanti le Energy Communities[viii],[ix],[x],[xi],[xii], da cui è stata stilata una lista preliminare, successivamente ampliata utilizzando il Large Language Model (LLM) *gpt-3.5-turbo*[xiii]. L'LLM è stato personalizzato fornendo un set di documenti sulle Energy Communities ed è stato interrogato con il prompt:

> *"I have to analyze the activity of Energy Communities on social media. Based on the document provided and your knowledge, extract a list of keywords for social media search."*

Infine, la lista è stata filtrata manualmente per eliminare eventuali parole poco rilevanti. Da questo processo sono state ottenute sessantaquattro chiavi di ricerca composte da una o più parole (cf. Figura 2).

anaerobic digestion	biogas
biomass	citizen energy communities
climate emergency	collective energy consumption management
collective self-consumption	community building
community energy projects	community engagement
community engagement in energy systems	community solar power
community wind farms	community-based renewable energy
community-owned energy	cooperative
decarbonization efforts	decentralized energy systems
democratic access to energy	distributed energy resources
eco-feedback for energy conservation	energy affordability and decentralization
energy autonomy	energy co-ops
energy communities	energy community regulations
energy cooperative	energy cooperatives
energy culture	energy efficiency
energy justice	energy marketplace
energy policy and consumer participation	energy resilience
energy sharing schemes	energy storage
energy sustainability and community benefits	energy transition
energy vulnerability and poverty	environmental impact
flexibility in household energy use	hydropower
interactive coordination systems	local energy autonomy
local energy initiatives	local energy projects
local grid connection	local stakeholders
microgrid communities	neighborhood energy systems
peer-to-peer platform	privacy concerns in energy data
pro-environmental behavior	renewable energy communities
renewable energy cooperatives	social acceptance of energy transitions
solar energy sharing	solar pv
sustainable energy communities	transparency in community energy usage

Figura 2 Elenco delle chiavi di ricerca identificate

Successivamente, da ogniuna delle piattaforme sono stati estratti 200.000 messaggia, circa 330 per ogni keyword in modo da avere un dataset bilanciato.

7.4 Analisi delle relazioni tra utenti e contenuti

Al fine di individuare eventuali Energy Communities online e analizzare le relazioni tra i diversi utenti, sono stati creati dei grafi. Per la rappresentazione è stata usata la libreria *ipysigma*[xiv] che permette la visualizzazione interattiva di grafi.

I dati sono stati modellati con tre diversi tipi di grafi che differiscono per le relazioni tracciate, ovvero: le relazioni utente-utente e utenti-hashtag; le relazioni utente-utente; le relazioni utente-hashtag. Per utente si intende l'autore del messaggio.

Nella prima tipologia, sono considerati nodi sia gli utenti sia gli hashtag. Viene creato un arco tra un utente e un hashtag ogni volta che l'hashtag è utilizzato in un messaggio dall'utente, e tra due utenti ogni volta che uno menziona l'altro. Eventuali nodi isolati sono stati eliminati. Un

esempio del grafo ottenuto per i messaggi di X è mostrato in Figura 3 e Figura 4. Questo tipo di grafo, esplorabile dinamicamente mediante un normale web browser permette di analizzare le relazioni tra i diversi utenti e tra i contenuti dei loro messaggi.

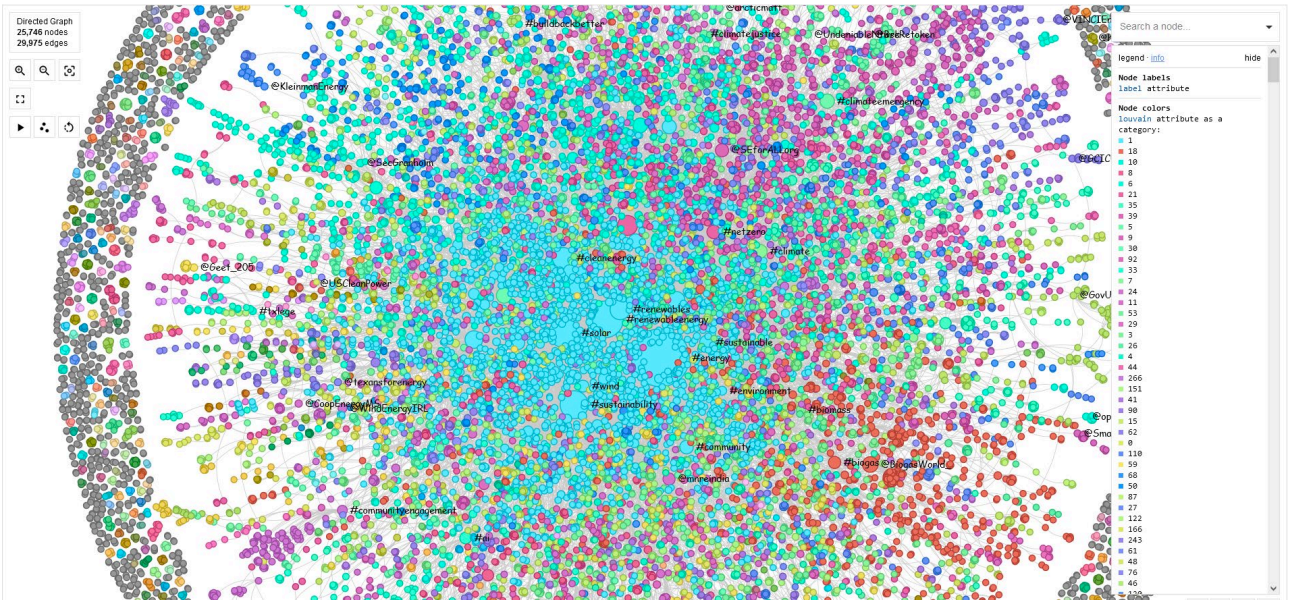

Figura 3 Relazioni utente-utente e utente-hastag per i mesaggi raccolti da X

Figura 4 Particolare del grafo di Figura 3

Nella seconda tipologia di grafo ogni nodo rappresenta un utente ed un arco tra due utenti viene creato nel caso in cui uno menzioni l'altro. Eventuali nodi isolati sono stati eliminati. Questa analisi, i cui risultati sono mostrati in Figura 5 e Figura 6, consente di evidenziare più facilmente la presenza di Energy Communities sui social media e di analizzare le relazioni tra i diversi utenti e le varie community.

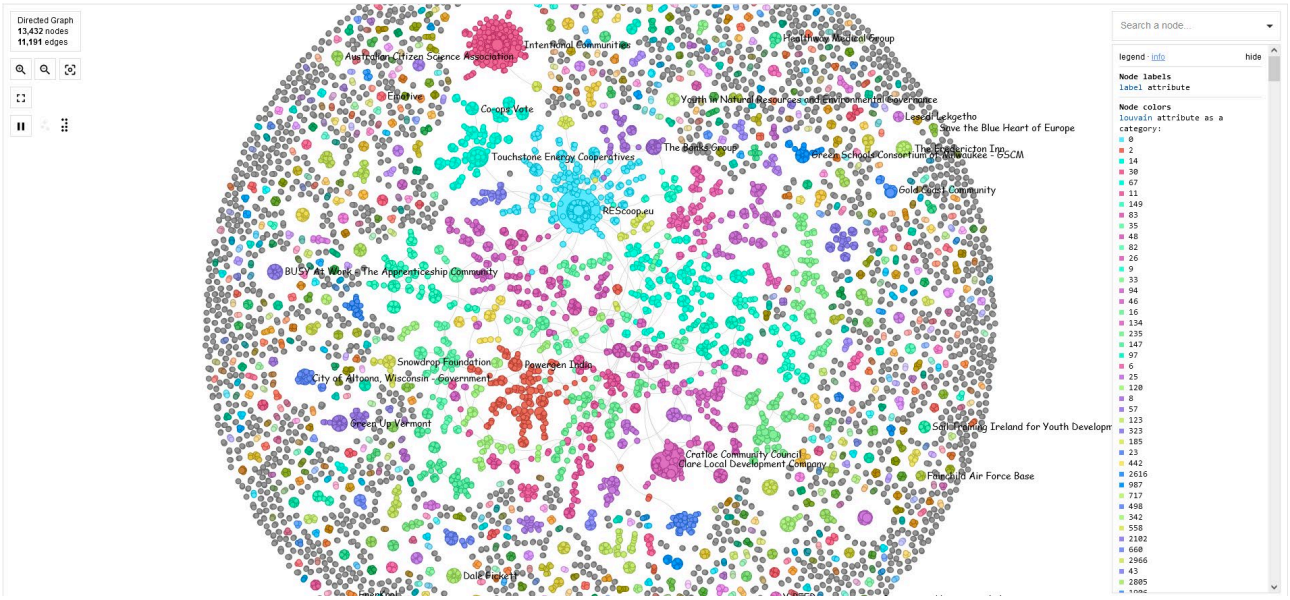

Figura 5 Grafo delle relazioni utente-utente per i messaggi raccolti da Meta

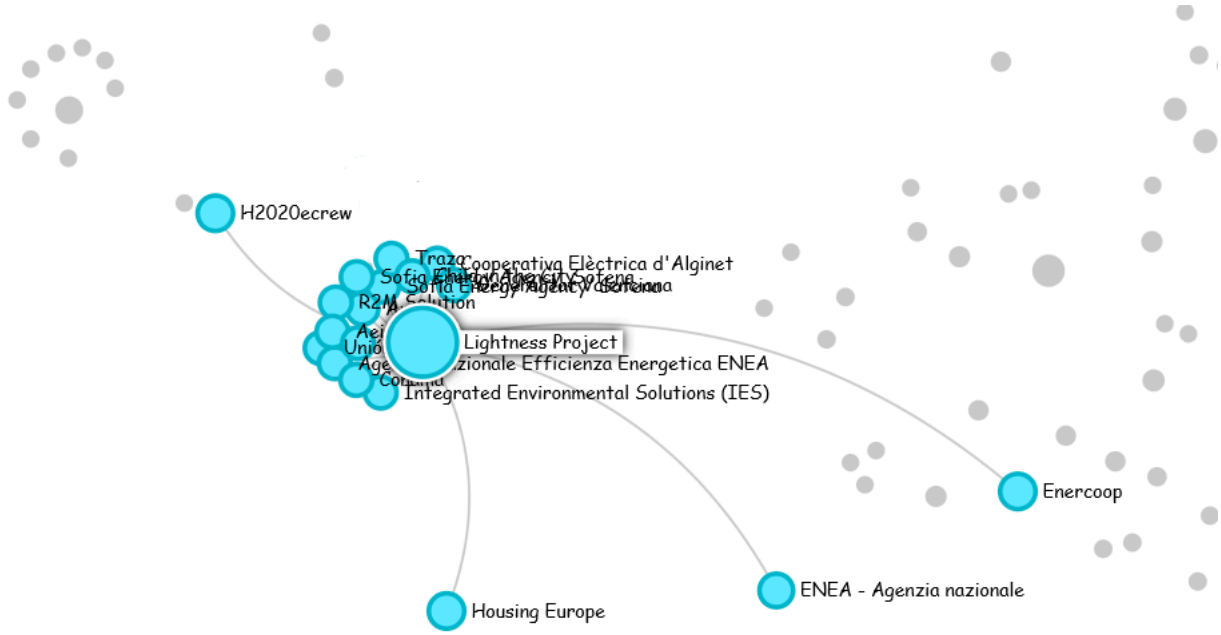

Figura 6 Particolare del grafo in Figura 5

Infine, nella terza tipologia di grafo, ogni nodo può rappresentare sia un utente che un hashtag. Gli archi, tuttavia, vengono creati esclusivamente tra hashtag e utenti, nel caso in cui l'utente utilizzi quel determinato hashtag all'interno di un proprio messaggio. Questa modellazione mira a evidenziare la relazioni e la vicinanza tra gli utenti, concentrandosi sul contenuto dei loro messaggi, poiché i messaggi con hashtag simili tendono ad avere contenuti affini. Un esempio per la piattaforma Meta è mostrato in Figura 7.

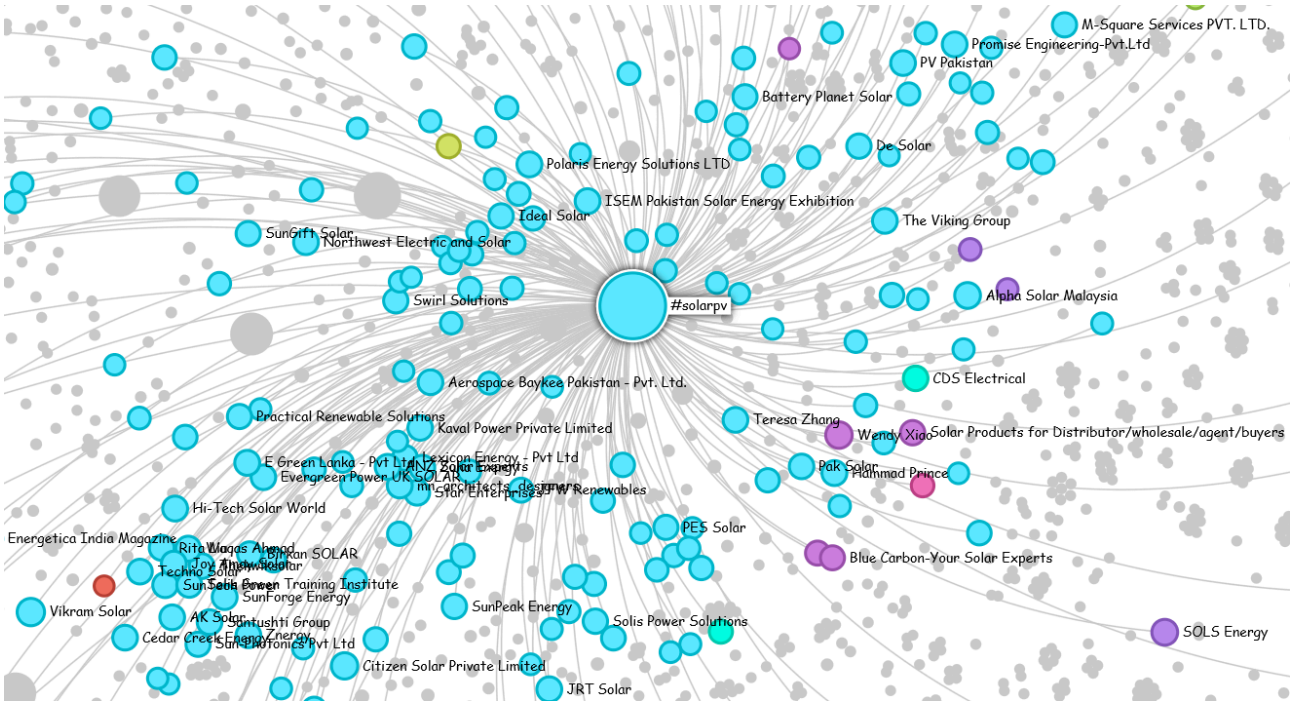

Figura 7 Grafo delle relazioni utenti-hashtag per i messaggi raccolti da Meta

7.5 Analisi del contenuto dei messaggi

L'analisi del contenuto dei messaggi mira ad ottenere informazioni sulla quantità e la tipologia degli argomenti discussi su Meta e X. Sono stati condotti due tipi di analisi: nella prima ci si è concentrati sull'attinenza dei vari messaggi alle keywords definite in §7.3. Nella seconda, invece, sono stati utilizzati algoritmi di clustering per classificare i messaggi in base al contenuto e analizzare con maggiore precisione gli argomenti trattati.

7.5.1 Attinenza alle chiavi di ricerca

Per determinare il punteggio di attinenza di un messaggio rispetto alla chiave di ricerca è stato utilizzato l'algoritmo di classificazione BM25 (cf. §7.1). Successivamente, per ogni messaggio è stata calcolata la somma dei punteggi di rilevanza relativi a ciascuna parola chiave al fine di ottenere un grafico che mostri quali messaggi siano maggiormente pertinenti ai temi proposti. Dalle Figura 8 e Figura 9 si nota come la quasi metà dei messaggi sia molto attinente alle chiavi di ricerca e la rimanente parte poco rilevanti, con una discontinuità netta.

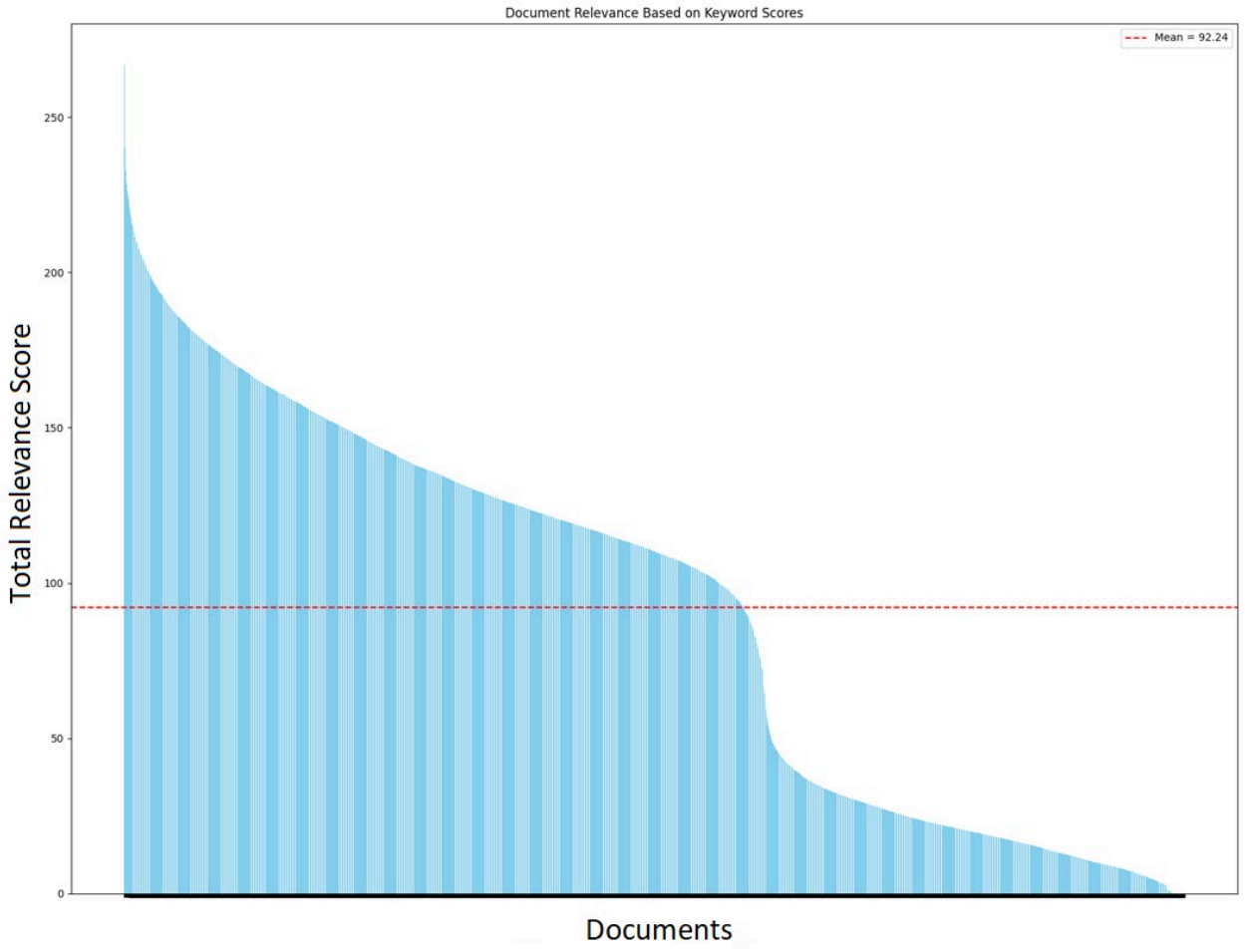

Figura 8 Attinenza dei messaggi alle chiavi di ricerca per i messaggi di X

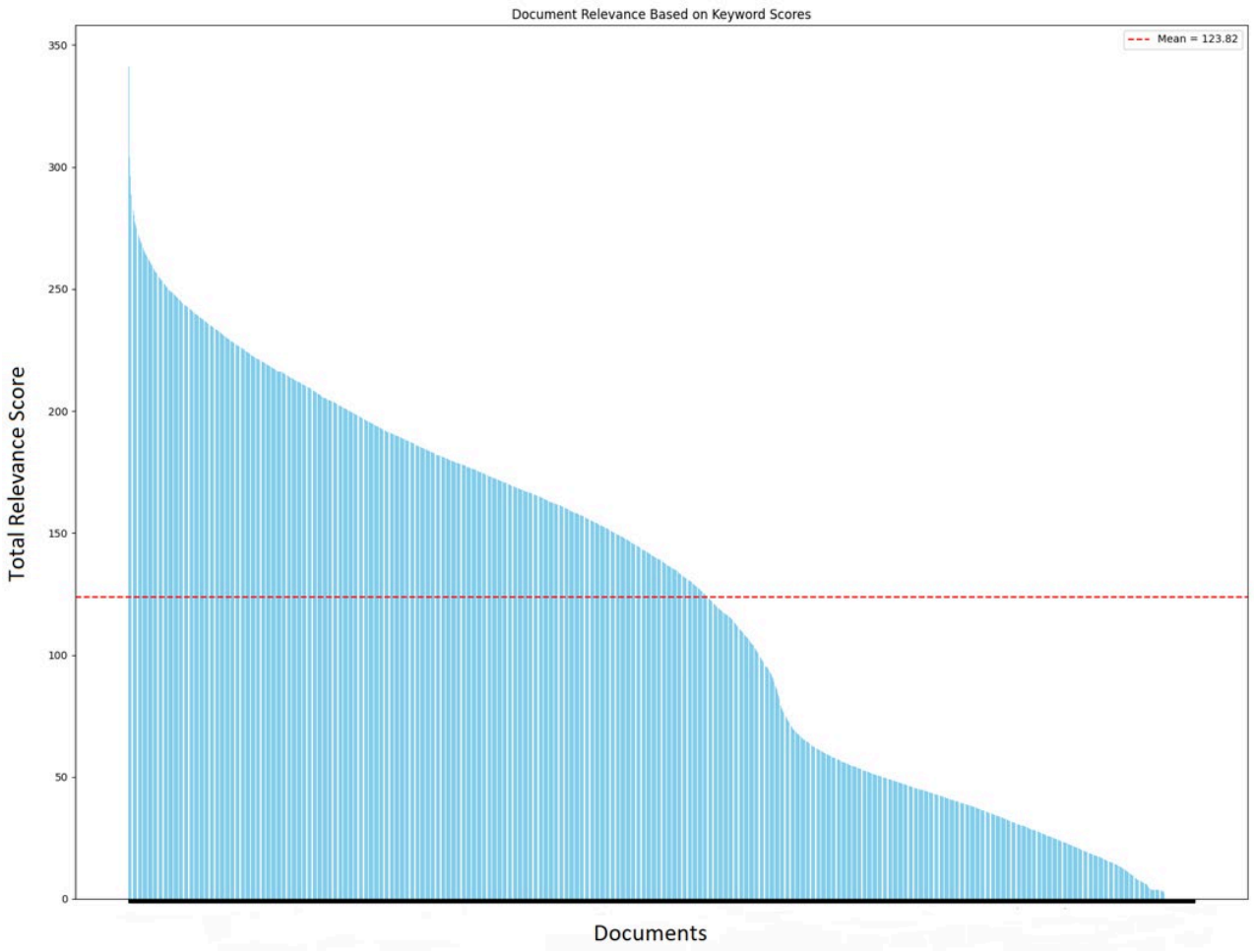

Figura 9 Attinenza dei messaggi alle chiavi di ricerca per i messaggi di Meta

Successivamente, per ogni chiave di ricerca, sono stati sommati i punteggi del messaggio contenente quella determinata chiave. Questa analisi permette di comprendere a quali chiavi di ricerca, e quindi a quali argomenti, facciano riferimento i messaggi ottenuti. Dalle Figura 10 e Figura 11 si nota come per entrambi i social ci siano poche chiavi di ricerca molto rilevanti, un gruppo di chiavi rilevanti allo stesso modo, ed un gruppo di chiavi poco rilevanti. Ad esempio, per i messaggi di X ci sono cinque chiavi (8.33%) molto rilevanti e che totalizzano un punteggio compreso tra 70000 e 80000, trentacinque chiavi (58.33%) il cui punteggio varia tra 35000 e 50000, e le rimanenti venti chiavi di ricerca (33.33%) con un punteggio inferiore a 20000. Le cinque chiavi più rilevanti sono, nell'ordine: *energy sustainability and community benefits*, *democratic access to energy*, *energy policy and consumer participation*, *energy vulnerability and poverty*. Per quanto riguarda i messaggi di Meta, le chiavi di ricerca con un punteggio superiore a 70000 sono dieci (16.66%), con la chiave *energy sustainability and community benefits* che risulta estremamente popolare, quelli con punteggio maggiore di 35000 sono trentadue (53.33%), e le rimanenti (30%) hanno un punteggio inferiore a 20000. Si nota anche come l'ordine di importanza delle chiavi differisce tra i messaggi di Meta e quelli di X.

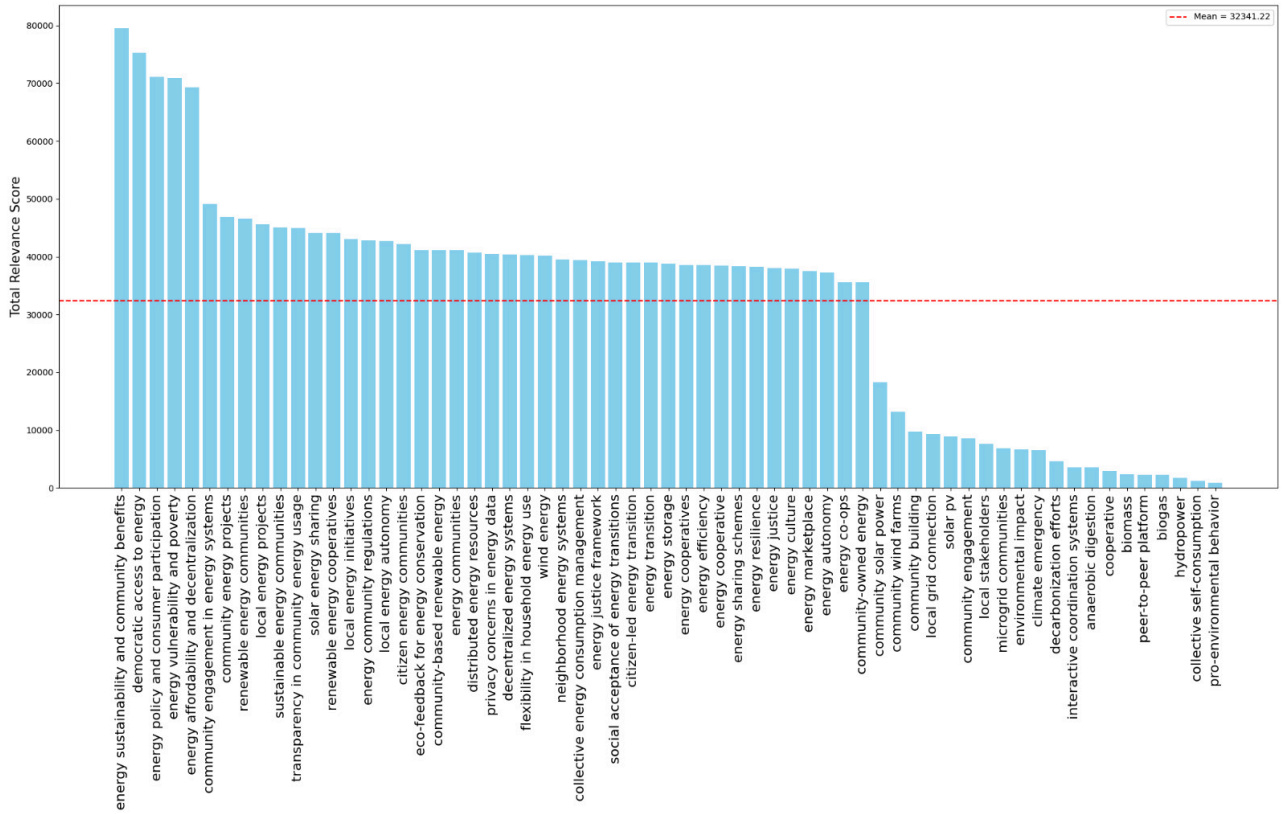

Figura 10 Rilevanza delle chiavi di ricerca per i messaggi di X

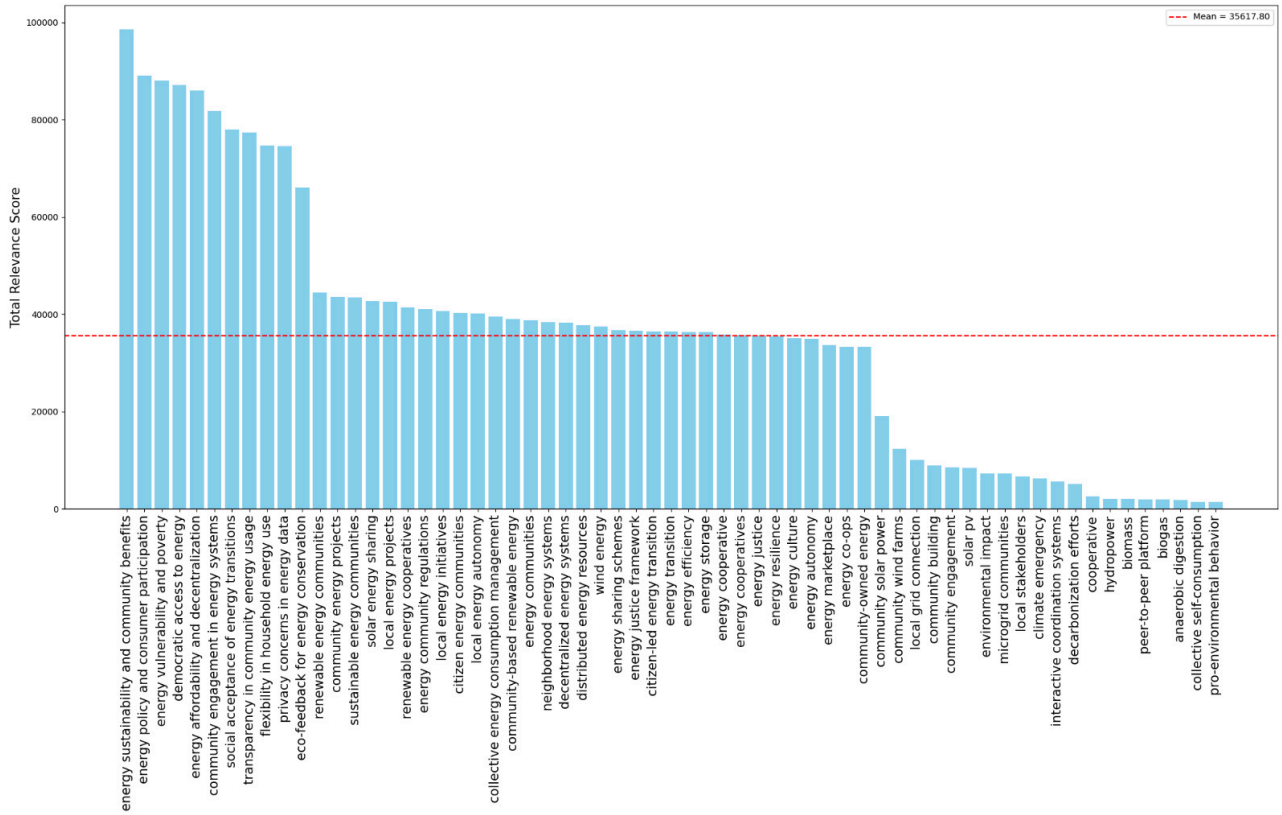

Figura 11 Rilevanza delle chiavi di ricerca per i messaggi di Meta

7.5.2 Analisi dei contenuti dei messaggi

Per analizzare la distribuzione dei messaggi in base al loro contenuto, sono state utilizzate due tecniche di clustering. La prima prevede la trasformazione del testo in vettori multi-dimensionali e la loro clusterizzazione, usando l'algoritmo k-means, consentendo così di capire come sono distribuiti gli argomenti trattati nei messaggi. La seconda tecnica implementa una pipeline di topic-modeling per identificare i principali argomenti discussi nei messaggi. Vediamo ora in dettaglio come sono state realizzate le due analisi.

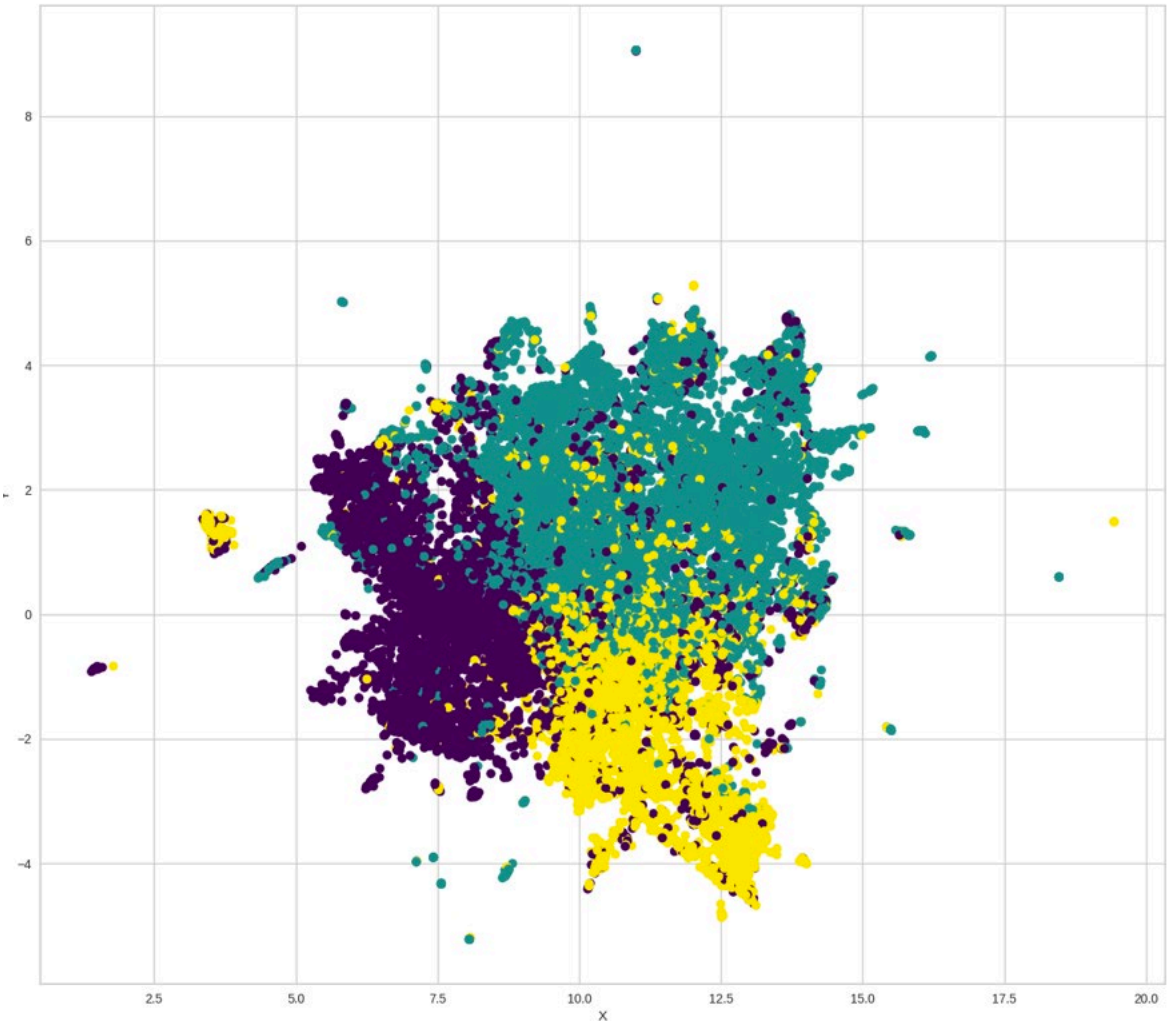

Figura 12 Rappresentazione bidimensionale dei cluster di messaggi di X

Per il calcolo degli embedding del testo è stata utilizzata la libreria *sentencetransformers* SBERT, ed in particolare il modello *all-MiniLM-L6-v2*, attraverso il quale sono stati generati, per ciascun post, vettori rappresentativi di 384 dimensioni. Il modello *all-MiniLM-L6-v2* è stato scelto in quanto offre il miglior compromesso tra velocità di esecuzione e precisione rispetto al modello che offre ma migliore qualità, ovvero all-mpnet-base-v2[xv]. Nel caso in cui il testo superasse la lunghezza gestibile dal modello, è stato suddiviso in blocchi utilizzando una tecnica di divisione ricorsiva. Una volta ottenuti i vettori di embedding, è stato applicato l'algoritmo di clustering k-means per suddividere i post in base alla vicinanza semantica del loro contenuto. Successivamente, per rappresentare visivamente i cluster in un grafico, è stata impiegata una tecnica di riduzione della dimensionalità. In particolare, è stato utilizzato UMAP, un algoritmo che eccelle nel preservare sia la struttura locale che quella globale dei cluster, catturando anche relazioni non lineari tra i dati. Il risultato della trasformazione è mostrato

nelle Figura 12 e Figura 13. Si nota come in entrambi i casi ci sono tre principali gruppi di argomenti trattati nei messaggi. Bisogna però osservare che la separazione tra i cluster non è netta.

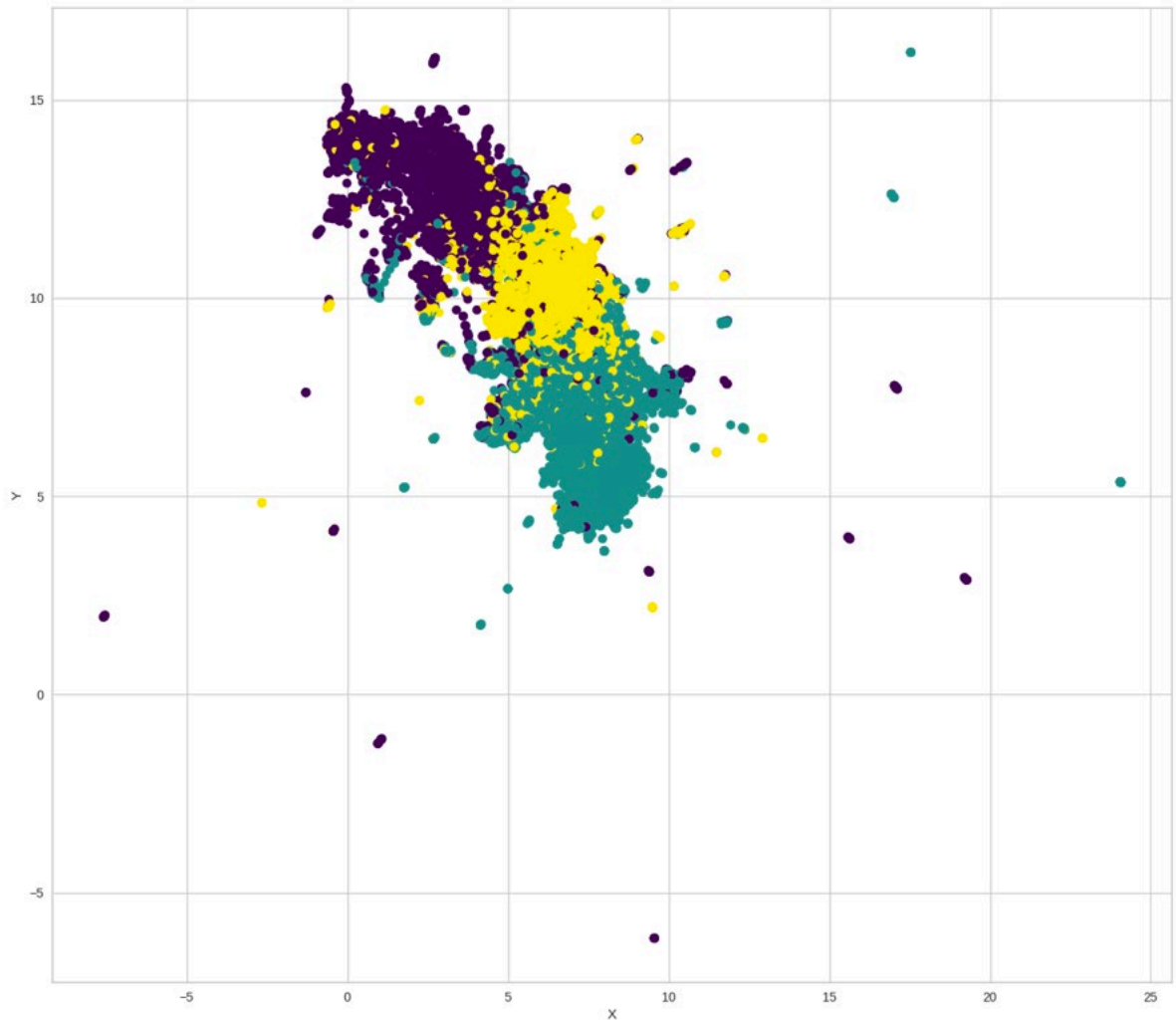

Figura 13 Rappresentazione bidimensionale dei cluster di messaggi di Meta

Al fine di individuare gli argomenti discussi all'interno dei cluster è stata implementata la pipeline di modellazione dei topic facendo uso della libreria BERTopic. Inizialmente, sono stati calcolati i vettori di embedding usando il modello sentence transformer, a cui è stata successivamente applicata la riduzione delle dimensionalità mediante l'algoritmo UMAP. Il terzo passo ha previsto il calcolo dei cluster mediante l'algoritmo HDBSCAN, con il vincolo che ogni cluster contenesse almeno 75 messaggi. Questo accorgimento è stato adottato per evitare la creazione di cluster troppo piccoli, che risulterebbero poco informativi. Per identificare le parole più rappresentative di ciascun cluster, è stato impiegato l'algoritmo c-TF-IDF, una variante del tradizionale TF-IDF progettata per lavorare a livello di cluster o categorie, anziché di singoli documenti. In questo modo, sono state ottenute le parole più rappresentative per ciascun cluster. Per creare una breve descrizione rappresentativa dell'argomento trattato in ogni cluster, questa è stata ottenuta utilizzando un LLM. In particolare, le parole più rappresentative di ogni cluster sono state utilizzate per costruire un prompt di interrogazione del modello LLM gpt-3.5-turbo adattato come spiegato in §7.3

I cinque principali temi trattati in X e in Meta sono rispettivamente riportati nelle Tabella 1 e Tabella 2. È interessante notare come solo il 2.5% dei messaggi di X ed il 3% dei messaggi di

Meta si polarizza intorno ai cinque argomenti principali, e che tutti gli altri sono dispersi su discussioni di interesse ristretto, ovvero che coinvolgono meno dello 0.2% dei messaggi scambiati. Degli argomenti di maggior interesse, per quanto riguarda X, il tema del "supporto finanziario alla realizzazione di progetti per le energie rinnovabili" polarizza il 41.6% della discussione, mentre seguono a pari livello di interesse (14.4% circa) le tematiche legate allo sviluppo di progetti per l'uso dell'energia solare ed eolica, o in generale energie rinnovabili e pulite.

Topic id	Descrizione	Numero post
1	Community Energy Generation Fund supports local renewable projects	2160 (41.6%)
2	Community solar energy projects and benefits	769 (14.9%)
3	Community-owned wind farms and offshore wind energy sustainable development	752 (14.5%)
4	Renewable Energy Cooperatives for community-owned clean electricity	751 (14.5%)
5	Anaerobic digestion of waste-activated sludge	749 (14.5%)

Tabella 1 Le cinque principali topic trattate sulla piattaforma X

Su Mata invece la discussione si accentra su due temi principali: la rivoluzione dell'energia pulita promossa dal basso (dagli utenti) (43%), e lo sfruttamento dell'energia solare (32.8%). Seguono al margine le discussioni su energia eolica e biogas (8.7%).

Topic id	Descrizione	Numero post
1	Grassroots clean energy revolution in US communities	2795 (43%)
2	Harnessing solar power for a sustainable future	2122 (32.8%)
3	Importance of wind energy for a sustainable future	567 (8.7%)
4	Celebrating biogas innovation and sustainability in the energy sector	563 (8.7%)
5	Seeking connection in nature-centered communities	422 (6.5%)

Tabella 2 Le cinque principali topic trattate sulla piattaforma Meta

7.5.3 Analisi del sentimento

Infine, sfruttando il modello di sentiment analysis RoBERTa[xvi], sono stati analizzati tutti i messaggi raccolti ed è stato analizzato il sentimento degli utenti rispetto ai temi trattati. Il sentimento degli utenti viene classificato come Positivo, Neutrale e Negativo. Per quanto riguarda la piattaforma X il 50% di messaggi esprime un sentimento Positivo, il 42.5% un sentimento Neutro, ed il 7.5% un sentimento Negativo. Per la piattaforma Meta, invece, il sentimento negativo sale al 22,5%, mentre il 40% dei messaggi esprimono un sentimento Positivo ed il 37.5 un sentimento Neutro.

Andando ad analizzare il sentimento rispetto agli argomenti trattati, ovvero le parole chiavi, abbiamo l'andamento riportato in Figura 14 e Figura 15. Si può notare che non c'è un argomento che polarizzi un particolare sentimento.

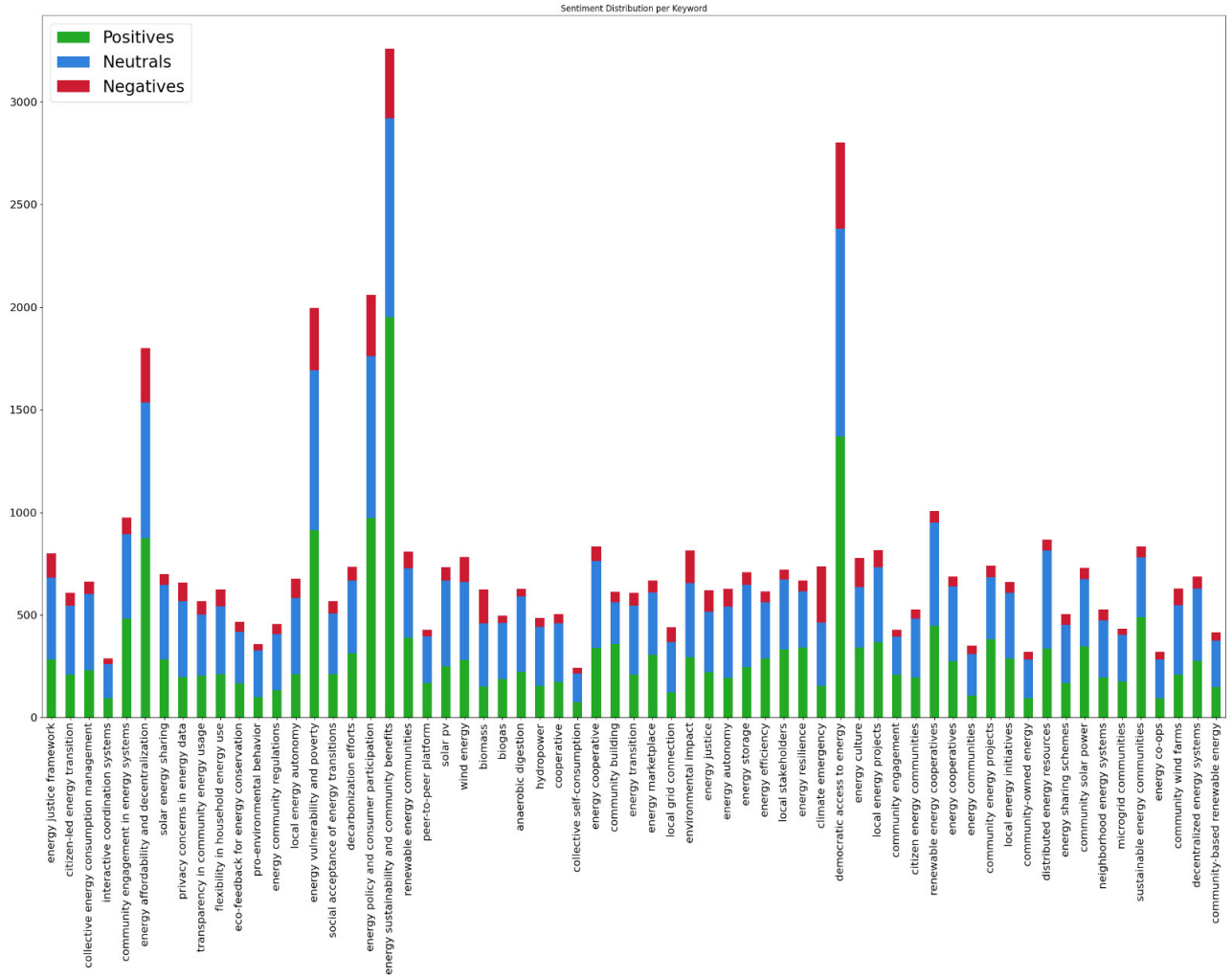

Figura 14 Distribuzione del sentimento degli utenti rispetto alle parole chiavi per la piattaforma X

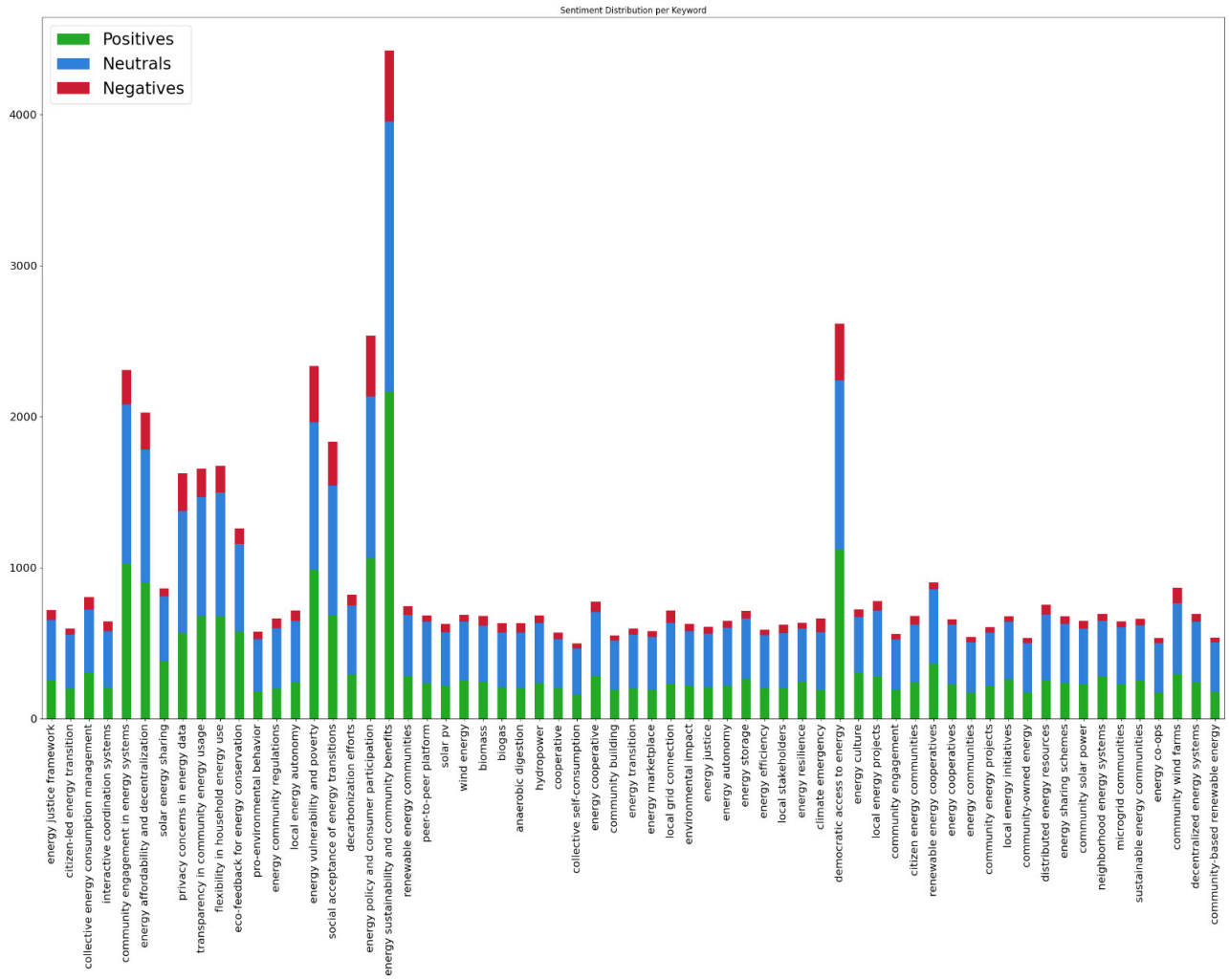

Figura 15 Distribuzione del sentimento degli utenti rispetto alle parole chiavi per la piattaforma Meta

8 Contributo delle eventuali consulenze alle attività sopra descritte

Non ci si è avvalsi di consulenze.

9 Pubblicazioni scientifiche

Ad oggi non sono state prodotte pubblicazioni scientifiche. Una pubblicazione scientifica è in fase di preparazione.

10 Eventi di disseminazione

Non sono stati svolti eventi di disseminazione in quanto non previsti dal capitolato.

[i] W. A. Qader, M. M. Ameen and B. I. Ahmed, An Overview of Bag of Words;Importance, Implementation, Applications, and Challenges, 2019 International Engineering Conference (IEC), Erbil, Iraq, 2019, pp. 200-204, https://doi.org/10.1109/IEC47844.2019.8950616

[ii] Nils Reimers, Iryna Gurevych, Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, 2019, https://doi.org/10.48550/arXiv.1908.10084

[iii] Leland McInnes, John Healy, James Melville UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, 2018. https://doi.org/10.48550/arXiv.1802.03426

[iv] Campello, R.J.G.B., Moulavi, D., Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In: Pei, J., Tseng, V.S., Cao, L., Motoda, H., Xu, G. (eds) Advances in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science(), vol 7819. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37456-2_14

[v] Grootendorst, Maarten. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure." *arXiv preprint arXiv:2203.05794* (2022).

[vi] BERTopic GitHub repository https://github.com/MaartenGr/BERTopic

[vii] The Selenium Browser Automation Project, https://www.selenium.dev/documentation/

[viii] Florian Hanke, Rachel Guyet, Marielle Feenstra Do renewable energy communities deliver energy justice? Exploring insights from 71 European cases, Energy Research & Social Science, Volume 80, 2021, https://doi.org/10.1016/j.erss.2021.102244.

[ix] Federico Gianaroli, Michele Preziosi, Mattia Ricci, Paolo Sdringola, Maria Alessandra Ancona, Francesco Melino, Exploring the academic landscape of energy communities in Europe: A systematic literature review, Journal of Cleaner Production, Volume 451, 2024,https://doi.org/10.1016/j.jclepro.2024.141932.

[x] Daniel Petrovics, Dave Huitema, Mendel Giezen, Barbara Vis, Scaling mechanisms of energy communities: A comparison of 28 initiatives, Global Environmental Change, Volume 84, 2024, https://doi.org/10.1016/j.gloenvcha.2023.102780.

[xi] Iraide López, Nerea Goitia-Zabaleta, Aitor Milo, Julen Gómez-Cornejo, Itxaso Aranzabal, Haizea Gaztañaga, Elvira Fernandez, European energy communities:Characteristics, trends, business models and legal framework, Renewable and Sustainable Energy Reviews, Volume 197, 2024, https://doi.org/10.1016/j.rser.2024.114403.

[xii] Neves C., Oliveira T., Sarker S. Citizens' participation in local energy communities: the role of technology as a stimulus, European Journal of Information Systems, 1–24. https://doi.org/10.1080/0960085X.2024.2302426

[xiii] OpenAI, GPT-3.5 Turbo fine-tuning and API updates, https://openai.com/index/gpt-3-5-turbo-fine-tuning-and-api-updates/

[xiv] Guillaume Plique. (2022). ipysigma, A Jupyter widget using sigma.js to render interactive networks. Zenodo. https://doi.org/10.5281/zenodo.7446059

[xv] SBERT, Pretrained models, https://www.sbert.net/docs/sentence_transformer/pretrained_models.html

[xvi] Liu, Yinhan. "Roberta: A robustly optimized bert pretraining approach." *arXiv preprint arXiv:1907.11692* 364 (2019).