UNIVERSITÀ DI PISA

Dipartimento di Informatica

Corso di Laurea in Data Science and Business Informatics

TESI DI LAUREA MAGISTRALE

Polarizzazione Politica & Echo Chamber: una metodologia per l'identificazione e analisi su Reddit

Relatori:

Prof. Giulio Rossetti

Dr.ssa Laura Pollacci

Candidato:

Virginia Morini

Sessione estiva

Anno Accademico 2019/2020

Sommario

Heterogeneity in content and freedom of expression provided by Social Networking sites can cause to their users Cognitive Dissonance. Such a discomfort leads users to selectively expose themselves to information that supports their personal beliefs or values, i.e., confirmation bias. This trend is further reinforced by the reccomendation algorithms of the Social Networks which leads to phenomena such as Polarization and Echo Chambers. This thesis, focusing on a political context of the first two and a half years of Donald Trump presidency, aims to identify Echo Chambers and measure Polarization on Reddit. Initially, we define a methodology to measure the political polarization of a sumbission (i.e., post on Reddit). Once a Ground Truth is created, we leverage two different algorithms/ approaches on this task and evaluate their results. The first consits in identifying via Latent Dirichlet Allocation highly disinctive terms from both that chacterize the rethorical languages of both parties and subsequently training a Support Vector Machine to classify submission accoridng to such extracted features. The latter is a neural approach. It leverages Word Embeddings and Long Short-Term Memory architecture in order to measure the polarization score of a submission by quantifying its agreement with both ideologies. Next, we verify the existence of polarized system, namely Echo Chamber, across three different topics concenrning socio-political issues. For all of them, we define the user interaction network, labeled by their polarization score. By applying algorithms of Community Detection, we extract communities that we further analyze by evaluating their structural and idelogical cohesion. The approaches deployed enable to identify potential politically polarized Echo Chamber both with respect to the democratic and republican ideologies.

Indice

Introduzione

Nella società contemporanea, le piattaforme di Social Networking (SNS[1]) sono ormai parti integranti della nostra quotidianità. Più della metà della popolazione mondiale[2] usa i SNS in media tre ore al giorno[3]. Queste nuove realtà virtuali proiettano gli utenti in un flusso continuo di informazioni, opinioni, immagini senza alcun limite spazio temporale o di espressione. L'eterogeneità dei contenuti e la libertà di pensiero offerte da SNS, porta gli utenti ad entrare in contatto con informazioni e ideologie che talvolta si discostano o addirittura contraddicono le proprie convinzioni. Lo stato di disagio provato in tali situazioni, noto come dissonanza cognitiva, porta gli individui ad adottare atteggiamenti di esposizione selettiva e bias di conferma. Gli utenti sono cioè più propensi a selezionare e condividere contenuti che rinforzano le proprie opinioni preesistenti, evitando quindi quelli contrastanti. Questa tendenza, insita nella natura dell'essere umano, viene ulteriormente rafforzata dagli algoritmi dei Social Network. Le piattaforme, infatti, per adeguarsi ai gusti dei propri utenti, attuano complessi meccanismi di personalizzazione dei contenuti, presentando ad ognuno una stessa informazione in modo diverso.

La ricerca di similarità ideologica si riflette, in un contesto online, anche nella scelta degli utenti con cui interagire e connettersi. Questa attitudine porta alla formazione di gruppi polarizzati, ossia insiemi di utenti in cui una tendenza iniziale di un membro viene rafforzata ed estremizzata dal confronto con gli altri. Tali sistemi, in cui solamente alcune ideologie, valori e informazioni vengono condivise sono definite *Echo Chamber* (EC). Ne consegue che i membri di un EC sono così fossilizzati nelle proprie convinzioni da arrivare a pensare che l'unica

[1] Dall'inglese *Social Network Sites*

[2] https://datareportal.com/reports/digital-2020-global-digital-overview ultimo accesso: 01/07/2020

[3] https://blog.globalwebindex.com/trends/2019-in-review-social-media ultimo accesso: 01/07/2020

visione veritiera della realtà sia la loro.

Questo fenomeno rischia così di minare il flusso democratico delle opinioni, alimentando situazioni di odio e inciviltà online. Proprio per tali motivazioni, nell'ultimo decennio la comunità scientifica ha mostrato crescente interesse nello sviluppo di tecniche e approcci per l'identificazione di sistemi polarizzati sulle principali piattaforme di SNS, quali Facebook e Twitter.

Il presente lavoro nasce con lo scopo di apportare un contributo a questo indirizzo di ricerca, proponendo una metodologia per l'identificazione di sistemi polarizzati su *Reddit*, piattaforma di Social Networking mai sfruttata in questo senso. Abbiamo deciso inoltre di declinare tale fenomeno ad un contesto di tipo politico, identificando e caratterizzando EC di stampo democratico e repubblicano durante gli anni di presidenza di Donald Trump. Nella sfera politica, in particolare quella Americana, la polarizzazione trova infatti il suo habitat naturale incarnando la tensione delle ideologie binarie dei partiti che la compongono.

Poiché le EC sono sistemi fortemente polarizzati, la prima fase di questo lavoro si concentra sulla definizione di una metodologia per predire il grado di polarizzazione di un contenuto online. In primo luogo, abbiamo creato una *ground truth* annotando post estratti da *Subreddits* noti per essere particolarmente polarizzati rispetto alle ideologie democratiche e repubblicane. In seguito, abbiamo testato parallelamente due diversi approcci per risolvere questo task. Il primo prevede l'estrazione di keywords discriminanti dai corpus creati, tramite *Latent Dirichlet Allocation* (LDA), e la conseguente classificazione dei post in base alle features estratte, tramite *Support Vector Machine* (SVM). Il secondo approccio, di tipo neurale, sfrutta *Word Embeddings* e *Long Short-Term Memory* (LSTM) per catturare la polarizzazione veicolata da un post. In particolare, il grado di polarizzazione viene misurato quantificando l'allineamento di un post rispetto alle ideologie democratiche o repubblicane. Analizzando i risultati ottenuti da entrambi gli approcci, notiamo che le performances di LSTM sono nettamente migliori di quelle di SVM. Gli step successivi sono quindi svolti utilizzando l'approccio neurale. Il lavoro svolto in questa fase è stato presentato e accettato alla *SEBD Conference 2020* [74].

La seconda fase di questa tesi prevede l'effettiva identificazione di Echo Chamber in tre topic relativi alla sfera socio politica. Inizialmente, creiamo un Network di interazioni tra utenti annotati in base alla polarizzazione dei loro post. In seguito, applichiamo algoritmi di *Community Discovery* (CD) per estrarre comunità omogenee dai Network creati. Definiamo quindi

requisiti strutturali e ideologici per valutare se le comunità estratte possono essere definite EC, analizzandone le peculiarità. Tra gli algoritmi di CD testati, *Infomap* risulta essere il più performante, individuando in ogni topic potenziali Echo Chamber di stampo democratico e repubblicano.
Il seguente lavoro di tesi è cosi suddiviso:

- Nel **Capitolo 1** viene presentata la letteratura esistente relativa agli argomenti presentati. In particolare, ci soffermiamo sugli aspetti psicosociali che portano alla formazione di Echo Chamber su SNS e sulle tecniche e metodologie utili alla loro identificazione.
- Nel **Capitolo 2** viene discusso il caso di studio scelto per l'identificazione di sistemi polarizzati.
- Il **Capitolo 3** si concentra sul processo di estrazione e manipolazione dei dati che ha portato alla creazione dei Dataset finali.
- Nel **Capitolo 4** sono discussi i due approcci testati per predire la polarizzazione di un post e i risultati ottenuti.
- Nel **Capitolo 5** viene presentata la metodologia utilizzata per identificare Echo Chamber a partire da utenti annotati in base al grado di polarizzazione politica. Le EC estratte sono state analizzate in base alla loro coesione strutturale e ideologica.
- Nel **Capitolo 6** viene riassunto l'elaborato e ipotizzate le possibili integrazioni future.

Capitolo 1

Stato dell'Arte

Questo capitolo, suddiviso in due sottocapitoli, presentata la revisione della letteratura. In primo luogo, nella Sezione 1.1, sono esplorati i fattori psicologici e sociologi che, con la diffusione dei Social Network, portano alla definizione di fenomeni come Polarizzazione ed Echo Chamber. Successivamente, nella Sezione 1.2, sono discussi gli studi che mirano alla loro identificazione e misurazione su SNS e Social Media. Inoltre tale Sezione fornisce una panoramica generale sulle tecniche di *Natural Language Processing*, *Machine Learning* e *Social Network Analysis* utilizzate in questo lavoro di tesi.

1.1 Contesto Psicosociale

Per definire fenomeni sociali e digitali quali Polarizzazione ed Echo Chamber, è opportuno studiare congiuntamente due aspetti: da un lato, la natura e le peculiarità delle piattaforme di Social Networking, strumenti chiave attraverso i quali le Echo Chamber nascono, crescono e persistono nel tempo; dall'altro le teorie di dissonanza cognitiva, considerate in letteratura come la principale causa della nascita e diffusione della Polarizzazione.

1.1.1 Social Network e Dissonanza Cognitiva

Nell'ultima decade, la nascita e l'incredibile diffusione di SNS ha drasticamente cambiato il modo in cui le persone interagiscono fra loro, diventando una parte integrante della loro quo-

tidianità. Statistiche[1] dimostrano che non esiste un limite di età nell'uso di SNS: il 90.4% dei *Millenials*, nati tra 1980-1994, il 77.5% della *Generation X*, nati tra 1965-1979, il 48.2% dei *Baby Boomers*, nati tra 1944-1964, usano attivamente SNS.

L'avvento di SNS ha quindi portato il consueto modo di comunicare ad evolversi in nuove realtà dominate da molteplicità, interattività, assenza di barriere spazio temporali e libertà di espressione. In [18] gli autori forniscono una definizione formale di SNS, sottolineando gli aspetti di molteplicità e partecipazione che differenziano la comunicazione digitale da quella tradizionale. Rendendo pubblico un proprio contenuto, l'individuo tende infatti a trascurare la comunicazione uno a uno favorendo piuttosto il dialogo e lo scambio interattivo di opinioni con un bacino di utenti potenzialmente infinito.

Come asserito da Brundidge in [19], il successo di tali piattaforme è senz'altro dovuto ai loro confini sfumati, che di fatto si traduce in una totale assenza di barriere spazio-temporali nella diffusione di una qualsiasi informazione. L'interazione diventa quindi immediata, globale ed eterogenea [27].

Ed è proprio l'eterogeneità dei contenuti diffusi online, che consegue dalla libertà di opinione concessa dai SNS, l'aspetto su cui si vuole concentrare l'attenzione. Quotidianamente, gli utenti di tali piattaforme vengono inondati da innumerevoli notizie, opinioni e informazioni che talvolta si discostano dalla loro esperienza tangibile. In letteratura, è stato ampiamente studiato come la mentalità e personalità di un individuo reagisce alla cosiddetta eterogeneità percepita, portando alla luce risvolti sia positivi che negativi.

Lo studio di Boulianne et al. [17] dimostra come l'uso di SNS sia positivamente correlato alla partecipazione politica e civile dell'individuo, che è quindi sempre più incoraggiato a condividere le proprie ideologie online. Anche gli autori di [98] giungono alla stessa conclusione, sottolineando come la condivisione di contenuti di interesse culturale porti l'utente ad avere una migliore percezione di se stesso, più attiva, impegnata e matura. In questo senso, un altro aspetto cruciale messo in luce in [63] riguarda l'interdipendenza tra la distanza fisica tra gli utenti, imposta dai SNS, e la maggiore libertà di espressione in tematiche di pubblico interesse.

Tuttavia, come illustrato in seguito, l'illimitata libertà di espressione e la conseguente eterogeneità dei contenuti percepita dagli utenti possono avere ripercussioni negative per uno

[1]https://www.emarketer.com/chart/226029/social-media-users-by-generation ultimo accesso: 01/07/2020

scambio costruttivo, critico e razionale di opinioni [93].
Sin dai primi anni 50, esperti di psicologia e scienze cognitive si sono interrogati su come gli esseri umani reagiscono a ideologie, esperienze e valori fattuali contrastanti al loro modo di pensare e potenzialmente capaci di cambiarlo. I loro sforzi si sono concentrati nel rispondere a tre domande:

1. Un individuo, di fronte a un'argomentazione fattuale in disaccordo con la propria, cambia opinione? In caso affermativo, in che misura?

2. Un individuo, nel processo cognitivo che porta al formarsi di un'opinione, è influenzato dal pensiero altrui? In caso affermativo, in che misura?

3. Un individuo definisce le proprie convinzioni basandosi sull'evidenza? Oppure, al contrario, sfrutta le proprie ideologie per giudicarla?

Un grande passo avanti in questo studio viene compiuto dallo psicologo statunitense Leon Festinger che nel 1957 in [32] introduce la teoria della Dissonanza Cognitiva, nel tentativo di rispondere ai quesiti sopra citati. Tale costrutto, nonostante revisioni e ampliamenti, è tutt'oggi uno dei pilastri della psicologia sociale. Festinger definisce la dissonanza cognitiva come la sensazione di disagio provata da un individuo di fronte a due ideologie incompatibili fra loro. Con ideologia si intende qualsiasi valore, attitudine, esperienza, comportamento o conoscenza. Due ideologie risultano quindi dissonanti fra loro quando una implica l'opposto dell'altra. L'impatto della dissonanza sull'individuo cresce in modo proporzionale al numero di ideologie dissonanti e all'importanza attribuita a ciascuna di esse.
Un efficace esempio di dissonanza cognitiva, ampiamente citato da Festinger, è quello del fumatore assiduo di sigarette che continua a fumare nonostante sia a conoscenza delle conseguenze deleterie del fumo per la sua salute.

Lo stato di disagio provocato da tale situazione, muove l'individuo a cercare di ridurre la discrepanza tra le due cognizioni dissonanti per avvicinarsi ad uno stato mentale di consistenza e coerenza tra le proprie opinioni, detto *consonanza cognitiva*. A questo proposito, Festinger [32] individua tre modalità per ridurre la dissonanza tra due cognizioni:

1. Ridurre l'importanza delle ideologie dissonanti;

2. Aggiungere ideologie consonanti che sminuiscono quelle dissonanti;

3. Modificare le ideologie dissonanti in modo che risultino coerenti e consistenti.

Tali strategie di riduzione della dissonanza hanno importanti conseguenze nel comportamento dell'individuo, ampiamente studiate nell'ambito della psicologia cognitiva e sociale. Tra queste, per i fini di questo studio è necessario citare la teoria dell'*esposizione selettiva*, anche conosciuta come *bias di conferma*. Con tale terminologia viene definita la tendenza degli individui a favorire, selezionare e diffondere informazioni che rinforzano le loro ideologie preesistenti, evitando quelle che risultano contraddittorie [64, 68].

Tra gli esperimenti sociali condotti per supportare tale teoria sono citati di seguito i lavori svolti da Fischer et al. in [35] e da Jonas et al. in [61]. In entrambi gli esperimenti la dissonanza viene indotta da una decisione triviale che i partecipanti sono incoraggiati a prendere (e.g., scelta tra due beni di consumo ugualmente attraenti, tra due strategie vantaggiose di investimento, tra due dibattute ideologiche politiche). Successivamente, i partecipanti ricevono informazioni supplementari di cui metà a supporto e metà in contraddizione con la decisione presa e sono invitati a selezionare le informazioni che vogliono leggere in modo più approfondito. Nella grande maggioranza dei casi, i partecipanti selezionano solamente gli articoli a supporto della loro idea iniziale.
Entrambi gli studi giungono quindi alle stesse conclusioni: l'individuo è più propenso a dimostrare con qualsiasi mezzo che le sue argomentazioni sono corrette, piuttosto che a metterle in discussione in favore ad un approccio più moderato e costruttivo.

Le teorie della dissonanza cognitiva e della conseguente esposizione selettiva sono state applicate non solo all'ambito socio-psicologico ma anche a contesti online quali i Social Media e, più nello specifico, i SNS. Numerose ricerche [57, 58] hanno condotto, interamente sul Web o su SNS, esperimenti analoghi a quelli sopraelencati. Gli autori hanno concluso che gli utenti, quando sono esposti a contenuti incongruenti con le loro ideologie, sono soggetti a percepire un conflitto interiore strettamente connesso a sentimenti di tensione e disagio. Come evidenziato in [16], tali contrasti vengono superati cercando informazioni, persone o discussioni che confermano la propria ideologia oppure abbandonando la situazione di dissonanza.

Concludendo, risulta quindi evidente come l'eterogeneità dei contenuti offerta dai SNS può portare il singolo utente a trovarsi in situazioni di dissonanza. Esso cerca quindi di risolvere la problematica adottando atteggiamenti di esposizione selettiva e di bias di conferma.

1.1.2 Definizione di Polarizzazione e Echo Chamber

Fino ad ora abbiamo analizzato il rapporto tra SNS e teorie di dissonanza cognitiva dal punto di vista individuale. Tuttavia, essendo i primi uno strumento di connessione tra un numero potenzialmente infinito di utenti [18], risulta necessario spostare tale analisi su una dimensione collettiva.

Polarizzazione di gruppo e Polarizzazione politica

L'area della psicologia che si concentra sullo studio delle comunità si basa sull'assunzione che una delle principali fonti di motivazione dell'essere umano nasca dal bisogno di appartenere ad un gruppo [71]. Baumeister e Leary [9] hanno dimostrato come ogni individuo senta l'impellenza di avere quotidianamente una certa quantità di interazioni sociali. L'incapacità di soddisfare questo bisogno si traduce in solitudine, disagio mentale e in un forte desiderio di formare nuove relazioni.
La scelta dell'essere umano di appartenere ad un gruppo piuttosto che ad un altro dipende solitamente dal grado di similarità con altri individui già facenti parte di tale comunità [13]. In altre parole, l'uomo tende ad aggregarsi con individui che gli permettono di raggiungere uno stato mentale di consonanza cognitiva [36].

Partendo da queste assunzioni, è possibile definire il concetto di Polarizzazione di Gruppo, ampiamente studiato da Sunstein in [94]. Con questa terminologia si definiscono le situazioni in cui una tendenza iniziale di singoli membri verso una data direzione viene rafforzata e resa più estrema dal confronto con gli altri membri [76]. Turner, in [97], asserisce che i membri di un gruppo possono essere paragonati alle molecole polarizzate che si allineano in modo più deciso nella direzione che stavano già favorendo.
In questo senso, gli esperimenti di Asch [3], relativi alle tecniche di condizionamento messe

in atto dalle dinamiche di gruppo, dimostrano che i singoli membri sono disposti a negare l'evidenza diretta dei propri sensi piuttosto che contraddire quanto asserito dagli altri membri.

Il fenomeno della polarizzazione di gruppo è osservabile in numerosi contesti come ad esempio quello sociale [75, 79], giuridico [55] e politico. Questo lavoro si focalizza sull'ultimo dominio.

Con *polarizzazione politica* si definisce la divergenza delle preferenze politiche verso estremi ideologici [34, 5]. Sia nei sistemi bipartitici che multipartitici, la polarizzazione trova il suo habitat naturale, incarnando la tensione delle ideologie e identità binarie/multiple che li compongono [29, 1].

Politologi hanno individuato due livelli di polarizzazione politica:

1. *Polarizzazione di Elite:* con tale terminologia si intende la polarizzazione esistente tra il/i partito/i al governo e il/i partito/i all'opposizione. I partiti politici polarizzati sono internamente coesi, programmatici e ideologicamente distinti. Solitamente gli studi su questo tipo di polarizzazione si focalizzano sugli organi legislativi e deliberativi [70, 42];

2. *Polarizzazione di Massa:* tale fenomeno si verifica quando l'atteggiamento dell'elettorato nei confronti delle questioni politiche e delle persone coinvolte si divide nettamente secondo le ideologie di partito. Ne consegue la messa in discussione della legittimità morale del partito opposto [23, 66]. Con "massa" si intende l'elettorato e, più in generale, le persone comuni non direttamente coinvolte nelle scelte attuate dalle istituzioni politiche.

In questo scenario, l'enorme libertà di espressione e l'assenza di confini spazio temporali offerta dai SNS porta ad amplificare repentinamente l'effetto di questo fenomeno [100].

Nella Sezione 1.2 saranno discusse ed analizzate tecniche e metodologie utilizzate in letteratura per identificare esempi di polarizzazione di massa sui SNS.

Echo Chamber

Tra le applicazioni del concetto di polarizzazione di gruppo, e in particolare di polarizzazione politica, il fenomeno sociale e digitale delle Echo Chamber è ampiamente accreditato e studiato.

Come suggerito dal nome, con la terminologia Echo Chamber viene definita metaforicamente una situazione in cui solamente alcune idee, informazioni e valori vengono condivise, rafforzandosi a vicenda. Sia Sunstein in [95] che Jamieson et al. in [56] sostengono che coloro che fanno parte di una EC entrano in contatto con ideologie e informazioni che già supportavano, arrivando quindi a pensare che l'unica visione veritiera della realtà sia la loro. Il concetto di EC, in letteratura, viene solitamente associato alla sua accezione digitale.

Come evidenziato da Dubois et al. [31], sono principalmente due i fattori che portano Internet e le tecnologie ad esso correlate a incoraggiare la nascita e la diffusione di EC in un contesto digitale. Il primo aspetto riguarda la natura intrinseca del web che, basandosi sull'eterogeneità dei contenuti offerti e sulla libertà di espressione, permette agli utenti di selezionare e condividere informazioni provenienti da una miriade di fonti diverse, rafforzando le loro ideologie preesistenti [83]. I fattori sociali e psicologici che portano gli individui a far parte di un contesto digitale altamente polarizzato sono state discusse in Sezione 1.1.1.

Di seguito viene approfondito il secondo fattore di diffusione e formazione di EC, prettamente legato alla struttura degli algoritmi che regolano l'interazione tra utente e contenuto web. Come sottolineato in [51], gli utenti delle piattaforme di Social Media e Social Networking condividono le loro ideologie, esperienze e valori sulle stesse piattaforme che usano per informarsi (e.g., Facebook, Twitter e Reddit). Di conseguenza gli amministratori di tali piattaforme hanno potenzialmente a disposizione un'incredibile mole di informazioni riguardanti le preferenze espresse dagli utenti durante le loro attività online. Considerando che l'obiettivo delle piattaforme è quello di massimizzare il livello di soddisfazione dei loro utenti, che interesse avrebbero a mostrare loro contenuti dissonanti? Le piattaforme attuano un complesso meccanismo di personalizzazione dei contenuti, presentando ad ogni utente una stessa informazione in modo diverso per incontrare i loro gusti [96].

Il risultato di tali meccanismi di personalizzazione e filtraggio dei contenuti porta al diffondersi del fenomeno digitale delle "bolle di filtraggio", teorizzato da Parisier in [80]. Formalmente, Parisier definisce le bolle di filtraggio (in lingua originale *Filter Bubble*) come quel personale ecosistema di informazioni fornito ad ogni utente e soddisfatto da alcuni algoritmi. Che si tratti di motori di ricerca, SNS, o Social media, gli utenti sono meno esposti a punti di vista conflittuali, confinandosi intellettualmente nella propria bolla di informazioni (Figura

1.1).

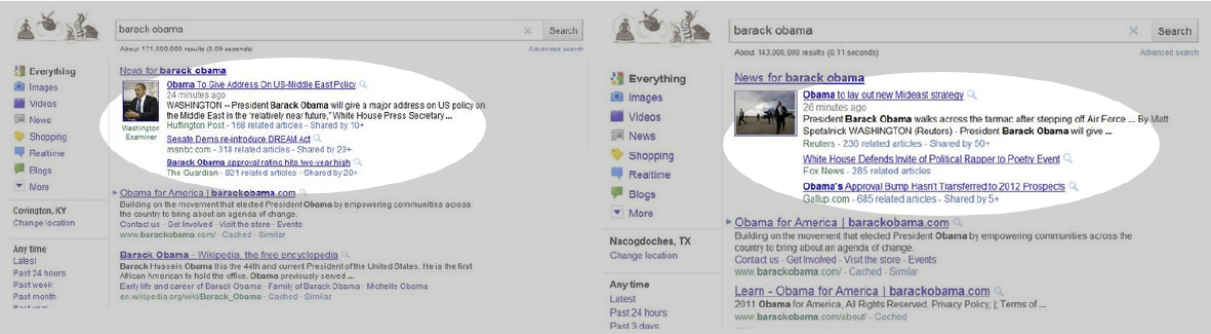


Figura 1.1: Esempio di Bolla di Filtraggio su Google.

Le conseguenze di tale bias algoritmico aprono una delicata parentesi, ampiamente discussa in ambito etico e sociologico, legata alla legittimità di tali algoritmi [90, 11, 24]. La tendenza degli individui a interagire e confrontarsi con persone aventi ideologie simili, rafforzata ulteriormente dalle piattaforme web, rischia di minare il flusso democratico delle informazioni, favorendo episodi di odio e inciviltà online [99, 85].

1.2 Studi riguardanti Polarizzazione e Echo Chamber

Il problema della polarizzazione online e del conseguente fenomeno delle EC viene affrontato in letteratura seguendo due approcci complementari. Il primo mira a verificare l'esistenza di tali fenomeni e, in caso affermativo, a studiarne la struttura e le peculiarità. Il secondo ha l'obiettivo di proporre tecniche e algoritmi per mitigarne l'effetto e la diffusione. Questo lavoro si colloca nella prima area di ricerca.

Dato che non esiste una definizione formale di EC, è necessario basarsi sulla letteratura esistente per comprendere a pieno quali sono gli aspetti da considerare nella modellazione del problema. Come discusso nelle sezioni precedenti, dal punto di vista sociologico, le EC sono definite come un sistema ideologicamente polarizzato. Per verificarne l'esistenza diventa quindi necessario definire una metodologia per stabilire il suo grado di polarizzazione [39]. Garimella, nella sua tesi di dottorato [40], suddivide le ricerche mirate all'individuazione della polarizzazione in due gruppi, basandosi sulla tipologia dell'approccio usato. Lo stato del-

l'arte è presentato di seguito seguendo questa tassonomia (i.e., Polarizzazione del Contenuto, Polarizzazione del Network).

Polarizzazione del Contenuto

Un contenuto viene definito polarizzante quando induce sentimenti contrastanti. Di conseguenza, gli individui che si confrontano con tale contenuto avranno nella loro totalità opinioni conflittuali (e.g., giusto/sbagliato, pro/contro). Questo approccio si basa sull'assunzione che la scelta di un contenuto controverso porti all'identificazione di sistemi altamente polarizzati.

Gli autori di [4] applicano questa metodologia a Facebook. Costruiscono un Dataset composto da oltre 10 milioni di utenti americani, che hanno espresso la loro ideologia politica, e dai relativi URL condivisi. Tramite l'uso di SVM classificano tali contenuti in altamente polarizzanti (e.g., politica, affari mondiali) e meno polarizzanti (e.g., sport, intrattenimento). Concentrandosi solo sulla prima macro-categoria, definiscono una metrica per misurarne il grado di affiliazione politica (i.e., Conservatori, Liberali). Infine, per ogni individuo, costruiscono la rete dei contenuti con cui viene in contatto, differenziando quelli provenienti dalla rete di amici e quelli proposti dal *feed* di notizie di Facebook.
Concludono che la prima, rispetto agli algoritmi di Facebook, gioca un ruolo decisivo nel limitare l'esposizione dell'individuo a contenuti aventi un ideologia diversa dalla propria e alla conseguente formazione di Echo Chamber.

Un lavoro simile è stato svolto in [44] sui blog. Gli autori utilizzano un Dataset composto da oltre 1,000 commenti fatti a contenuti presenti nei 33 blog più famosi al mondo. Anche in questo scenario, i blog sono raggruppati in macro-categorie (e.g., politica, tecnologia, tempo libero). I commenti sono classificati manualmente in 'negativi', 'positivi' e 'neutri' in base al loro livello di accordo con il contenuto del blog. Infine utilizzano tale *training set* in modo congiunto a tecniche di Natural Language Processing (NLP) (e.g., funzioni di peso quale *Tf-idf* e analisi grammaticale) e a tecniche di Machine Learning (ML) (e.g., classificatori quali *Bagging*, *Naive Bayes*) per predirre il livello di accordo di nuovi commenti estratti dai blog. Gli autori concludono che in argomenti meno polarizzanti il livello di accordo è in rapporto di 3 a 1 con quello di disaccordo e di 9 a 1 in argomenti altamente polarizzanti.

In questo progetto di tesi utilizzeremo questo tipo di approccio.

Polarizzazione del Network

Questo approccio si basa sull'assunzione che contenuti polarizzanti siano caratterizzati da un Network avente una struttura altamente clusterizzata. Con il termine *cluster* si definisce un insieme di elementi aventi caratteristiche simili. Inversamente alla metodologia basata sul contenuto, la polarizzazione del Network si basa sull'analisi della struttura dello stesso per identificare sistemi polarizzati.

Lo studio proposto in [2] è un esempio di questa tipologia di approccio. Glance et al. analizzano il grado e la struttura delle connessioni esistenti tra 40 blog americani. Selezionano blog che si occupano della condivisione di notizie e argomenti pertinenti alla sfera politica liberale e conservativa. A partire dai post presenti nei vari blog, costruiscono una rete di citazioni: si parla di citazione quando un URL presente all'interno di un blog rimanda ad un altro blog politico. Attuano tecniche di *Community Discovery* per verificare se il Network presenta insiemi di nodi caratterizzati da un numero molto alto di link interni rispetto a quelli esterni. Dai risultati ottenuti emergono due Echo Chamber principali connesse sia ideologicamente che strutturalmente, identificabili con la sfera conservativa e liberale. Infine, gli autori concludono che questo contesto predilige la formazione di Echo Chamber di tipo politico, ma che, mentre i Conservatori sono densamente connessi fra loro, i Liberali presentano una struttura meno uniforme e coesa.

Un'altra piattaforma di Social Networking che è necessario citare in questo scenario è Twitter. Gli autori di [41] si pongono l'obiettivo di misurare il grado di polarizzazione di diversi argomenti basandosi sulla struttura della rete che li caratterizza (e.g., legalizzazione delle armi, aborto, *Obamacare*). L'approccio presentato si basa su tre passi sequenziali. In primis, per ogni argomento viene costruita la sua rete conversazionale, ossia l'insieme delle connessioni esistenti tra utenti che ne parlano. Successivamente, ogni rete viene partizionata per identificare gruppi ideologicamente opposti che contribuiscono a creare un sistema polarizzato (e.g., testano algoritmi come *METIS*, *Spectral clustering*, *Label Propagation*). Infine, il grado di polarizzazione viene misurato testando diverse metriche, tra cui *Random Walk Controversy*, *Betweeness Centrality*, *Embedding Controversy*). Gli autori concludono che, nonostante questa metodologia possa essere applicata esclusivamente ai Dataset di Twitter, riesce efficacemente

a identificare e quantificare la polarizzazione di un argomento. Tra tutti i domini analizzati, i più polarizzanti sono risultati essere quelli legati alla sfera politica che, come sottolineato dagli autori, ben si prestano ad uno studio mirato all'individuazione di Bolle di Filtraggio e Echo Chamber.

Nonostante le principali piattaforme di Social Networking e Social Media (e.g., Facebook, Twitter, blog) siano state utilizzate in analisi di questo tipo, al meglio delle nostre conoscenze, in letteratura non sono attestate ricerche che mirano all'individuazione di EC su Reddit.

Lo stato dell'arte presentato, nonostante le differenze sia nelle piattaforme analizzate che nelle metodologie utilizzate, mette in luce alcuni aspetti cruciali:

1. Necessità di trasformare i contenuti testuali presi in considerazione (i.e., dati non strutturati) in una forma che si adatti efficacemente (i.e., dati strutturati) alle analisi successive. In seguito, definire una metodologia per misurare la similarità ideologica e il grado di polarizzazione di un insieme di contenuti/utenti.

2. Necessità di costruire un Network di connessioni fra gli utenti per studiare la struttura delle loro interazioni; necessità di utilizzare tecniche che permettono il partizionamento del Network in gruppi ideologicamente omogenei e polarizzati (i.e., Echo Chamber).

3. Tra i vari domini analizzati, la sfera politica risulta essere la più polarizzante.

4. La piattaforma di Social Networking scelta per l'analisi deve poter soddisfare le necessità indicate nei primi due punti e deve, preferibilmente, incoraggiare la presenza di discussioni e dibattiti.

Nelle Sezioni 1.2.1 e 1.2.2 sono presentate le tecniche e metodologie usate per strutturare i dati, misurare la similarità e polarizzazione dei contenuti (punto 1.). In Sezione 1.2.3 sono discusse tecniche di *Social Network Analysis* per la creazione e la successiva partizione di Network di utenti (punto 2.). Infine, gli ultimi due aspetti (punti 3. e 4.) sono affrontati nel Capitolo 2 e declinati al caso di studio scelto per questo lavoro.

1.2.1 Text Preprocessing e Feature Extraction

Le piattaforme di Social Networking permettono agli utenti di creare una quantità potenzialmente infinita di contenuti testuali come post e commenti. Il linguaggio utilizzato su tali piattaforme è solitamente non strutturato o semi-strutturato: gli utenti, nella formulazione di una frase, tendono a non tenere di conto dell'ortografia o delle regole grammaticali, esprimendosi spesso con espressioni colloquiali o dialettali [91]. Tali ambiguità, sintattiche e semantiche, rischiano di riflettersi sulle tecnologie che vi vengono applicate, impattando negativamente sulle loro performances finali [38].
Le aree di ricerca di *Natural Language Processing* e *Text Mining* offrono tecniche e metodologie per ridurre tali problemi. Con l'espressione Natural Language Processing si definiscono le tecniche mirate al trattamento informatico del linguaggio naturale. La terminologia Text mining si riferisce all'utilizzo di tecnologie per estrarre automaticamente informazioni da contenuti testuali. [50].

In questo scenario, il primo step riguarda il *text pre-processing*. Tale concetto si riferisce a tutte le operazioni di pre-elaborazione e pulizia dei dati testuali non strutturati, mirate, sia a rilevare e rimuovere anomalie, sia a ridurre le dimensioni del testo analizzato [54]. La tradizionale pipeline di text pre-processing prevede le seguente operazioni:

1. *Eliminazione di stop words:* le stop words sono parole lessicalmente vuote che, data la loro elevata frequenza in una lingua, sono di solito ritenute poco significative al fine dell'estrazione di informazioni rilevanti (e.g., in inglese *the, a, be*). Come affermato in [20], la rimozione di queste parole migliora l'efficienza e l'efficacia del pre-processamento del testo, in quanto ne riduce notevolmente la dimensione.

2. *Lowercasing:* con tale terminologia si intende la conversione di tutte le parole di un testo in minuscolo. Principalmente viene adottata per motivi di consistenza e coesione dell'input.

3. *Tokenizzazione:* tale espressione definisce il processo di frammentazione di un testo nelle parole che lo costituiscono, dette *token*. Generalmente, in questa fase viene rimossa la punteggiatura [77].

4. *Lemmatizzazione:* la lemmatizzazione si riferisce al processo di riduzione della forma flessa di una parola alla sua forma canonica, detta lemma. Questa fase si pone l' obiettivo di ridurre la sparsità dell'output.

Un altro aspetto centrale nell'elaborazione testuale riguarda la fase di *Feature Extraction*. Tale concetto, applicato a dati testuali, indica la creazione di una rappresentazione vettoriale di una parola tale che questa ne catturi i tratti salienti e risulti interpretabile dai vari algoritmi che vi saranno applicati. La scelta della tipologia di rappresentazione di una parola dipende dall'algoritmo che si vuole utilizzare per risolvere un *task* (i.e., compito). Focalizzando l'attenzione sul nostro scenario, è necessario sottolineare che i dati testuali estratti dai SNS, come post e commenti, sono sostanzialmente sequenze di parole che nel loro insieme veicolano un messaggio.

I due principali approcci utilizzati per rappresentare vettorialmente una parola sono i modelli di *Bag of Words* (BoW) e i *Word Embeddings* (WE). I modelli Bow sono l'approccio più semplice alla Feature Extraction. Questa tecnica codifica ogni parola distinta presente nel testo come un *one-hot-encoded vector* (i.e., un vettore binario con valori posti tutti a 0 tranne uno). Tale vettore ha dimensionalità pari alla grandezza del vocabolario e l'indice con valore 1 distingue la parola codificata. Un testo viene quindi rappresentato dalla somma dei vettori dei suoi costituenti. Tali vettori possono essere basati sulla semplice frequenza della parole di un testo oppure su altre metriche di peso di un termine, come *Tf-idf*.
Nonostante questo approccio sia piuttosto efficiente e facile da implementare, presenta però alcune limitazioni [65]. Il modello ignora la struttura sintattica di una frase (e.g., "cane morde uomo" è rappresentato vettorialmente allo stesso modo di "uomo morde cane"). Inoltre, data la natura dei vettori BoW, la valenza semantica di una parola non viene catturata a pieno (e.g., "comprare auto usate" e "acquistare vecchi veicoli" nonostante abbiano lo stesso significato sono rappresentati in BoW come vettori totalmente differenti).

A questo proposito, negli ultimi anni, i Word Embeddings sono diventati di fatto l'approccio standard per codificare la semantica di una parola. Con tale terminologia si fa riferimento generalmente ai metodi derivati dal lavoro di Mikolov, Word2Vec [72]. Questo approccio si basa sull'ipotesi di semantica distribuzionale secondo la quale il significato di una parola è dato dal suo contesto di utilizzo [49]. Più tecnicamente, definiamo i WE come dei vettori di

numeri reali, continui e densi tali che preservino le caratteristiche algebriche dello spazio da cui derivano. Questo significa che i WE riescono a catturare la semantica di una parola.
Oltre a Word2Vec, è importante citare un altro metodo per generare i WE, ossia GloVe [82]. Questa metodologia, come la precedente, tiene in considerazione la valenza semantica di una parola, in quanto nella creazione dei vettori la distanza di un termine da un altro è determinata dal loro grado di similarità semantica. Tuttavia, GloVe si differenzia da Word2Vec in quanto la fase di training viene effettuata a partire dalla matrice di co-occorrenze estratte dal testo.

1.2.2 Estrazione di ideologie

Con *ideologia* si intende l'insieme di credenze, opinioni e rappresentazioni che orientano le scelte del singolo individuo. Nel seguente lavoro di tesi questo concetto viene declinato ad un contesto politico: con la terminologia *ideologia politica* ci si riferisce all'insieme di esperienze, opinioni e valori che portano l'individuo a credere in determinate ideologie/orientamenti politici piuttosto che ad altri.

Nell'ambito di NLP e di ML sono numerosi i task aventi come obiettivo quello di, dato un contenuto testuale, estrarne l'ideologia che ne traspare e conseguentemente misurarne il livello di polarizzazione (i.e., capire quanto è forte e radicata tale ideologia). In questa tesi, il problema dell'estrazione delle ideologie da contenuti testuali, quali post di Reddit, è stato modellato testando due diversi approcci: *Supervised Text Classification* e *Topic Modeling*. Risulta quindi necessario dare una panoramica generale dei concetti alla base di tali metodologie.

Supervised Text Classification

Supervised Text Classification (TC) è un task di NLP, in cui, dato un insieme di testi di training, categorizzati in un numero finito di classi, si addestra un classificatore che assegni nuovi testi, non categorizzati, alla classe o alle classi a cui si presume appartengano (i.e., *target*), in base al contenuto dei documenti [89].
Formalmente, Text Classification viene definito come il task di addestramento di un classificatore $h : \mathcal{D} \rightarrow 2^{\mathcal{C}}$, dove $\mathcal{D}$ indica un insieme di documenti testuali e $\mathcal{C} = \{c_1, \ldots, c_{|\mathcal{C}|}\}$ un insieme di classi predefinite, che assegna ad ogni documento $\mathbf{x}_i \in \mathcal{D}$ un insieme di classi

appartenenti a $\mathcal{C}$ (i.e., un sottoinsieme di $\mathcal{C}$). Tale sottoinsieme di $\mathcal{C}$ può essere rappresentato come un vettore $v_h(\mathbf{x}_i)$ di lunghezza $|\mathcal{C}|$, in cui il valore +1 (i.e., al contrario: -1) nella j-esima posizione del vettore indica che la classe c_j è stata assegnata (i.e., al contrario: non assegnata) a $\mathbf{x}_i$.

I task di TC possono essere suddivisi in due gruppi in base al numero di classi target che vogliamo classificare. Si parla di *single-label text classification* quando esattamente una classe deve essere assegnata ad un documento testuale $\mathbf{x}_i \in \mathcal{D}$. Si parla invece di *multi-label text classification* quando qualsiasi numero di classi, in un range che va da 0 a $|\mathcal{C}|$, può essere assegnato allo stesso documento testuale $\mathbf{x}_i \in \mathcal{D}$. Un caso speciale di *single-label* è la *binary text classification*, in cui ogni $\mathbf{x}_i \in \mathcal{D}$ deve essere assegnato o alla classe c_j o al suo complemento $\bar{c}_j$. Nel seguente lavoro di tesi affronteremo un task di *Binary* TC.
Risulta importante sottolineare che, trattandosi di un task supervisionato, è necessario avere una *ground truth*. Tale terminologia applicata a questo scenario, si riferisce alla presenza di un insieme di documenti correttamente categorizzati su cui successivamente addestrare il classificatore.
Il task di TC è un componente fondamentale di molte applicazioni, come ad esempio categorizzazione di testi, filtraggio di spam e *Sentiment Analysis* [48, 67]. Sono altrettanto numerosi gli algoritmi utilizzati per risolvere task di TC. Per brevità, di seguito sono illustrati solamente quelli oggetto di questo lavoro.

L'algoritmo supervisionato *Support Vector Machine* è stato introdotto da Joachims in [59, 60] ed è ampiamente usato in task di TC. Dati un insieme di dati testuali di training, SVM cerca di trovare l'iperpiano che meglio li separa, basandosi sulle loro classi target. Tale approccio viene attuato secondo un criterio detto "margine di separazione massimo": l'iperpiano viene calcolato in modo che le distanze tra i punti (i.e., dati testuali) più vicini all'iperpiano e l'iperpiano stesso siano, su entrambi i lati, massimizzate. SVM può essere utilizzato sia per classificare dati separabili linearmente che non separabili in modo lineare. La seconda tipologia di SVM è diventata attualmente lo standard. Ciò è reso da possibile dall'introduzione delle cosiddette *slack variables* , indicate da ξ, che consentono di superare il problema della non-separabilità dei dati, permettendo che un dato numero di punti venga classificato erroneamente. Quindi, dato un numero sufficiente grande ξ, l'iperpiano risultante è convesso ed esiste

sempre. Formalmente, gli autori in [25] definiscono il problema di addestrare un classificatore tramite SVM come:

$$\begin{aligned} \text{minimizzare: } & \frac{1}{2}\mathbf{w}^2 + C\sum_{i=1}^{n}\xi_i \\ \text{per: } & \mathbf{w}, b, \xi_1, ..., \xi_n \\ \text{soggetto a: } & y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 - \xi_k \\ & \xi_i \geq 0, i = 1, ..., n \end{aligned} \tag{1.1}$$

dove **w** e b sono gli iperpiani e il bias, ξ_k sono le *slack variables* per i dati categorizzati (x_k, y_k), e C è l'iper-parametro che controlla il *trade-off* tra errore sul training e margine.

Un altro approccio largamente utilizzato per risolvere task di *supervised text classification* è quello neurale; di seguito ne illustriamo i tratti salienti.

Una Rete Neurale (NN[2]) è un modello di apprendimento automatico, ampiamente studiato in letteratura, che raggiunge ottimi risultati in una vasta gamma di task di ML, sia supervisionati che non supervisionati. Generalmente, una Rete Neurale è composta da un insieme di neuroni, comunemente chiamati nodi o unità, e da un insieme di archi diretti che li collegano. Ad ogni neurone *j* è associata una funzione di attivazione $l_j(\cdot)$. Ad ogni arco che va dal nodo j' al nodo *j* è associato un peso $w_{jj'}$. Il valore v_j di ogni neurone j viene calcolato applicando la sua funzione di attivazione alla somma pesata dei valori dei suoi nodi di input:

$$v_j = l_j\left(\sum_{j'} w_{ij'\cdot v_{j'}}\right) \tag{1.2}$$

Le funzioni di attivazione scelte sono solitamente *sigmoid* e *tanh*; Inoltre, la funzione scelta per i nodi di output dipende strettamente dal task che si sta performando (e.g., *softmax* per problemi di classificazione su più classi). Una limitazione di questo modello riguarda il fatto che è necessario specificare l'ordine in cui le varie computazioni devono essere svolte.

Le *Feedforward Neural Network* (FNN), un'altra tipologia di NN, risolvono questo problema impedendo la presenza di cicli nel grafo di nodi. Data l'assenza di cicli, tutti i nodi possono quindi essere disposti in *layer* (livelli) e di conseguenza l'output di ogni layer può essere calcolato dati gli output dei layer precedenti. L'input $\boldsymbol{x}$ di un FNN è quindi dato dall'impostazione

[2]Dall'inglese *Neural Network*

dei valori del primo layer. Ogni layer successivo è quindi calcolato in sequenza fino a che non viene generato l'output dell'ultimo livello $\hat{\mathbf{y}}$. FNN sono ampiamente utilizzate per task supervisionati di classificazione e regressione.

Nonostante l'uso di una FNN presenti numerosi vantaggi, è necessario evidenziare le limitazioni che non la rendono appropriata per il lavoro svolto in questa tesi. In una FNN, dopo che ogni esempio (i.e., dato) viene processato, l'intero stato della rete viene perso: questo non sarebbe un problema se ogni esempio fosse generato in modo indipendente ma se invece gli esempi sono connessi fra loro nel tempo o nello spazio diventa inaccettabile. Scenari di questo tipo possono riguardare l'elaborazione di video, audio, immagini e contenuti testuali. Concentrandosi sull'ultimo aspetto, un testo può essere rappresentato mediante una sequenza di parole dotata di una valenza semantica. Per codificare tale informazione, diventa necessario che un modello dopo aver processato una parola non si dimentichi di quella precedente ma che anzi la utilizzi per arricchire le informazioni a disposizione.

A questo fine vengono introdotte le *Recurrent Neural Networks* (RNN) [53, 62]. Le RNN sono delle FNN con una particolare caratteristica: alla struttura di una FNN sono aggiunti archi, detti *recurrent edges*, che si estendono su step temporali adiacenti, aggiungendo quindi la nozione di tempo nel suo funzionamento. Durante uno step temporale *t*, i nodi aventi *recurrent edges* ricevono un input sia dal dato corrente $\mathbf{x}^t$, sia dai valori del nodo nascosto $\mathbf{h}^{(t-1)}$ nello stato precedente della rete. L'output $\hat{\mathbf{y}}^{(t)}$ in ogni step temporale t viene calcolato dato il valore del nodo nascosto $\mathbf{h}^{(t)}$ allo step temporale *t*. L'input $\mathbf{x}^{(t-1)}$ allo step temporale *t-1* può influenzare l'output $\hat{\mathbf{y}}^{(t)}$ allo step temporale t e in quelli successivi grazie ai *recurrent edges*.
Questo tipo di approccio presenta delle limitazioni legate alla lunghezza delle dipendenze che la rete deve essere capace di 'ricordare' e propagare [12]. Tali problemi avvengono durante l'allenamento del modello quando per ricordare una dipendenza è necessario tornare indietro nella rete fino al *layer* iniziale. Queste problematiche vengono definite con la terminologia *Vanishing and Exploding Gradient problem* (i.e., scomparsa e esplosione del gradiente).

In [52], Hochreiter and Schmidhuber hanno introdotto il modello *Long Short-Term Memory* principalmente per superare il problema della scomparsa del gradiente delle RNN. Questo modello si basa infatti su una RNN con un *hidden layer* (i.e., livello nascosto): la differenza consta nel fatto che ogni nodo "originale" nell' hidden layer viene sostituito da una cella di memoria.

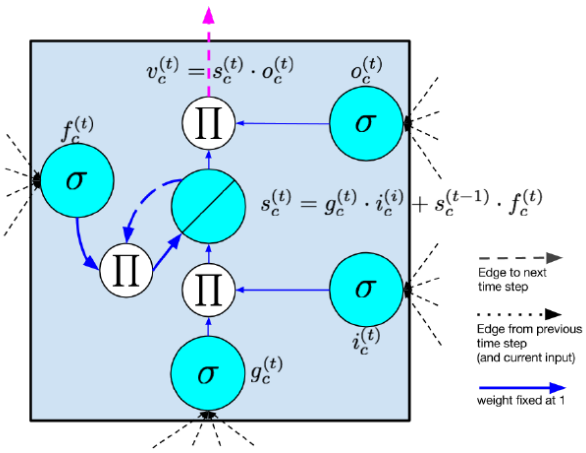


Figura 1.2: Rappresentazione grafica della cella di memoria di LSTM [43].

Una cella di memoria è un'unità composita, derivata a partire da nodi più semplici. Tale cella, come mostrato in Figura 1.2, è così composta:

1. *Input node:* Questo nodo, g_c, viene attivato in modo standard dal *layer* di input $\mathbf{x}^t$ nello step temporale corrente e dall' hidden layer allo step temporale precedente $\mathbf{h}^{(t-1)}$;

2. *Input gate:* i *gates* sono la caratteristica che distingue LSTM dagli approcci precedenti. Un *gate* è un'unità sigmoidale che, come il nodo di input, viene attivato dal dato corrente $\mathbf{x}^t$ e dall' hidden layer allo step temporale precedente. Un gate è così chiamato perché il suo valore viene utilizzato per moltiplicare il valore di un altro nodo: se il suo valore è uguale a 1 allora tutte le informazioni continuano a fluire all'interno della rete, altrimenti no (i.e., *input gate*= 0). Il valore dell' *input gate*, i_c, viene moltiplicato al valore dell' *input node*.

3. *Internal state:* All'interno della cella di memoria è presente un nodo, s_c, attivato in modo lineare, che viene definito in [52] come lo stato interno della cella. Tale nodo ha un *recurrent edge* connesso a se stesso con un'unità di peso fissa. Poiché quest'arco si espande, lungo gli step temporali, con peso costante, l'errore fluisce nei vari step successivi senza svanire o esplodere. L'aggiornamento dello stato interno viene indicato come segue:

$$s^{(t)} = \mathbf{g}^{(t)} \odot \mathbf{i}^{(t)} + \mathbf{s}^{(t-1)} \tag{1.3}$$

4. *Forget gate*: Questa tipologia di *gate* è stata introdotta in [43]. Gli autori forniscono un modo per far sì che la rete 'impari' il contenuto dello stato interno. Se l'output del *forget gate* è 0 l'informazione non rimane nello stato della cella, se invece è uguale a 1 può rimanere. L'aggiornamento dello stato interno, con l'aggiunta del *forget gate*, viene indicato come segue:

$$s^{(t)} = \mathbf{g}^{(t)} \odot \mathbf{i}^{(t)} + \mathbf{f}^{(t)} \odot \mathbf{s}^{(t-1)} \tag{1.4}$$

5. *Output gate:* L'ultimo valore, v_c, prodotto da una cella di memoria è il valore dello stato interno, s_c, moltiplicato al valore dell' *output gate*, o_c.

Topic Modeling

Con la terminologia *Topic Modeling* (TM) si definisce il task non supervisionato di ML che, dato un insieme di documenti testuali, si pone l'obiettivo di estrarre gli argomenti e i pattern che caratterizzano quella collezione. Essendo un task di tipo non supervisionato, non è necessario avere un training set o una ground truth per utilizzare questa tipologia di algoritmi. Gli algoritmi di TM, essendo modelli statistici e probabilistici, non si basano sulla semantica di una parola, o di un insieme di parole, per estrarre un argomento. Questi algoritmi si basano sull'assunzione che ogni parte di un documento viene combinata selezionando le parole da probabili *baskets of words* (i.e., insiemi di parole), in cui ognuno corrisponde a uno specifico argomento. Secondo tali algoritmi, le parole di un documento sono selezionate e combinate fino a che non viene trovata la distribuzione di parole più probabile nel formare un determinato *basket*. I task di TM ben si prestano a varie applicazioni, come analisi del sentimento [69], ingegneria del software [92] e indicizzazione nei motori di ricerca [28].

Tra i vari algoritmi di TM , di seguito viene illustrato il metodo utilizzato in questa tesi: *Latent Dirichlet Allocation*. LDA, introdotto in [14], è un modello statistico generativo e non supervisionato, nato con l'obiettivo di modellare un'insieme di documenti, detti *corpus*. Con il termine generativo si fa riferimento ad un tipo di modello che, dato per assunto che i dati vengono generati in determinate circostanze, cerca di impararne le distribuzioni. L'assunzione alla base di LDA è la seguente: un documento viene considerato come un insieme di argomenti, la cui distribuzione segue il principio di Dirichlet in tutti i documenti. A sua volta, ogni

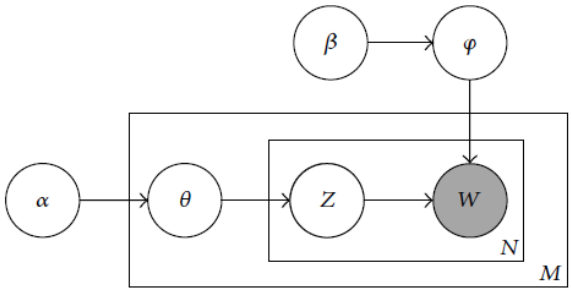


Figura 1.3: Rappresentazione grafica del modello generativo LDA [30].

argomento viene considerato come un insieme di termini, i quali seguono la distribuzione di Dirichlet. Più tecnicamente, dato un corpus *D*, costituito da *M* documenti, in cui il documento *d* è costituito da N_d parole $(d \in 71, ..., M)$, LDA modella *D* secondo il seguente processo generativo (Figura 1.3):

(a) Viene scelta una distribuzione multinomiale φ_t per l'argomento $t(t \in \{1, ..., T\})$ da una distribuzione di Dirichlet con parametro β;

(b) Viene scelta una distribuzione multinomiale Θ_d per il documento $d(d \in \{1, ..., M\})$ da una distribuzione di Dirichlet con parametro α;

(c) Per ogni parola $w_n(n \in \{1, ..., N_d\})$ in un documento *d*,

   i. Viene selezionato un argomento z_n da Θ_d;

   ii. Viene selezionato una parola w_n da φ_{zn}.

1.2.3 Social Network Analysis

La seconda fase di questo lavoro di tesi mira all'individuazione di sistemi polarizzati sulla piattaforma Reddit. Come discusso in Sezione 1.1.2, con l'espressione sistemi polarizzati o Echo Chamber si definisce un insieme di utenti che condividono una stessa ideologia e che tendono ad interagire principalmente fra di loro. Risulta evidente da un lato la necessità di costruire una rete di interazioni fra gli utenti, dall'altro l'esigenza di trovare tecniche che consentono di partizionare tale rete in gruppi ideologicamente omogenei e polarizzati. Di seguito, sono illustrati i concetti principali su cui si basa l'analisi dei Social Network, seguiti da un'analisi

degli approcci esistenti in letteratura per estrarre comunità da un Network (i.e., *Community Discovery*).

Cenni di Social Network Analysis

Con la terminologia *Social Network Analysis* (SNA) viene definita l'area di ricerca che mira a indagare la struttura e le peculiarità delle relazioni sociali attraverso l'uso della nozione di Network e dei concetti alla base della Teoria dei Grafi. Le tecniche di SNA hanno l'obiettivo di modellare sistemi complessi (i.e., sistemi composti da numerosi componenti che interagiscono fra loro) come Networks rappresentanti le interazioni tra gli elementi presi in considerazione.

Di seguito sono discusse le nozioni di SNA, basate sugli studi di Barabasi [7] e Newman [78], utili per la comprensione di questo lavoro. I componenti di un Network vengono definiti nodi, mentre le interazioni tra i nodi sono dette link. Dati due nodi *i,j* il link che li connette viene indicato con la notazione *(i,j)*. La dimensione di un Network viene definita dal numero di nodi presenti, indicata con *N*. Il numero totale di link esistenti tra i nodi viene indicato con *L*. I link di un Network possono essere diretti o indiretti. Un link viene definito indiretto quando tra i due nodi connessi esiste una relazione di simmetria (e.g., su Facebook se un utente *i* è amico di utente *j* allora anche *j* è amico di *i*). Un link viene definito diretto quando tra i due nodi connessi esiste una relazione di asimmetria (e.g., le telefonate: un individuo *i* chiama un individuo *j*). Un Network è definito indiretto se tutti i link presenti sono indiretti; viceversa, è definito diretto se tutti i suoi link sono diretti. Inoltre, un Network viene definito pesato se ad ogni link *(i,j)* è associato un peso distinto w_{ij} (e.g., numero di interazioni tra nodo *i* e *j*).

Ogni nodo ha un grado, definito in un Network indiretto come il numero di link adiacenti a quel nodo (e.g., se consideriamo il Network delle amicizie su Facebook, il grado di un nodo utente è dato dal numero di amici a cui è connesso). Il grado medio di un Network indiretto è dato dal rapporto fra la somma dei gradi dei nodi che lo compongono e il numero totali di nodi.

Community Discovery

Con l'espressione Community Discovery si definisce la ricerca e individuazione di comunità nascoste all'interno di Network complessi. La letteratura non fornisce una definizione condi-

visa del concetto di comunità. In generale, è possibile definire una comunità come un insieme di entità in cui ognuna è più vicina alle altre all'interno della comunità rispetto a quelle al di fuori di essa. Più formalmente, si intende un insieme di nodi strettamente connessi tra loro piuttosto che con i nodi appartenenti ad altri insiemi di uno stesso Network.

Oggigiorno, il task di CD è tra i più discussi nell'ambito di SNA. L'identificazione di gruppi di entità che condividono particolari proprietà o che hanno lo stesso ruolo in un dato fenomeno è infatti di fondamentale importanza per svariate applicazioni come l'identificazione di pagine Web che trattano lo stesso argomento [37] e di cicli metabolici aventi le stesse funzionalità [47]. Tale area di ricerca, inoltre, si presta perfettamente all'ambito dei Social Network e all'individuazione di gruppi di utenti strutturalmente o ideologicamente connessi fra loro [45, 81].

Poiché non esiste una definizione condivisa di comunità, esistono vari approcci di CD ciascuno basato su requisiti che un insieme di nodi deve soddisfare per essere considerato una comunità. Di seguito viene illustrata la tassonomia proposta in [26]:

1. *Feature Distance:* questo approccio comprende tutti gli algoritmi di CD che si basano sull'assunzione che una comunità sia composta da entità che condividono un insieme definito di features che assumono valori simili. Una feature ampiamente utilizzata può essere un arco o qualsiasi attributo connesso ad un'entità. Per esempio, se viene definita una misura di distanza basata sui valori delle features scelte, due entità apparterranno alla stessa comunità se risultano essere molto vicine l'una all'altra.

2. *Internal Density:* questa categoria comprende tutti gli algoritmi che definiscono una comunità come un insieme di entità densamente connesse fra loro. Ogni comunità deve quindi avere un numero di link decisamente superiore rispetto a quello di un *random graph*. Questi algoritmi si basano quindi sulla definizione di una funzione di qualità che miri a misurare la densità di una comunità e a massimizzarla. Una delle funzioni di qualità più utilizzate in questo contesto è la modularità. Questa funzione misura la qualità della partizione di un Network in comunità, tenendo in considerazione sia la densità interna di una comunità, sia l'assenza di link tra comunità diverse. Una modularità alta indica che il Network è suddivisibile in comunità e viceversa.

3. *Bridge Detection:* con tale terminologia si definiscono gli approcci di CD che identificano una comunità basandosi sul concetto che le comunità sono dei sottoinsiemi densi di un Network connesse fra loro da un piccolo numero di link (i.e., *bridge*) che se rimossi dividono il grafo in parti ben distinte.

4. *Diffusion:* questo approccio comprende tutti gli algoritmi di CD che si basano sull'idea che le comunità sono gruppi di nodi che possono essere influenzati dalla diffusione di una certa proprietà o informazione all'interno di un Network. Ai fini di questa tesi è necessario citare *Label Propagation*, uno dei metodi di Diffusion più utilizzati. Si tratta di un algoritmo semi supervisionato che assegna delle etichette (i.e., *label*) ai nodi di un Network. Inizialmente ogni nodo ha una label distinta che viene scambiata, nella prima iterazione dell'algoritmo, con quella di uno dei nodi vicini (stabiliti all'inizio). Nelle iterazioni successive, ad ogni nodo viene assegnata la label condivisa dalla maggioranza dei suoi vicini. L'algoritmo termina quando ogni nodo ha la stessa label dei suoi vicini.

5. *Closeness:* questa categoria di approcci di CD definisce una comunità come un insieme di entità vicine fra loro (i.e., pochi link separano nodi di una stessa comunità). Ne consegue che le entità di comunità diverse sono distanti l'una dall'altra. Un concetto largamente utilizzato per estrarre comunità i cui membri sono molto vicini l'uno all'altro è *Random Walk*: dato un Network e un nodo iniziale, selezioniamo un suo vicino in modo casuale e ci spostiamo su questo nodo, quindi selezioniamo un suo vicino sempre in modo casuale e ci spostiamo su tale nodo e così via. La sequenza di nodi identificati secondo questa metodologia viene definita un random walk nel Network.

6. *Structure:* seguendo queste metodologie di CD una comunità viene definita tramite una struttura precisa e immutabile di link. Gli algoritmi che si basano su questa idea definiscono tali tipologie di strutture e cercano di individuarle all'interno di un Network.

7. *Link Clustering:* questo approccio si differenzia dagli altri discussi in quanto il clustering non si basa sui nodi di un Network, ma sui suoi link. In questo caso, una comunità viene definita dalle relazioni esistenti fra i nodi che la compongono.

Nella scelta di un approccio di CD è necessario considerare il tipo di comunità da estrarre dalla rete. Le comunità sono definite: pesate, se i link che connettono i nodi forniscono delle informazioni aggiuntive; dirette, se le relazioni tra due nodi non sono simmetriche; sovrapposte, se i nodi possono contemporaneamente appartenere ad una o più comunità.

Infine, risulta utile introdurre due metriche utilizzate per valutare la qualità delle comunità estratte:

- *Internal Edge Density:* una comunità c viene definita densa se il numero di link interni alla comunità è vicino al numero massimo di link possibili. Dato l'insieme dei link L_c e l'insieme dei nodi N_c:

$$IED_c = \frac{2|L_c|}{|N_c|(|N_c| - 1)} \tag{1.5}$$

- *Conduttanza:* la conduttanza di una comunità è data dal volume dei link che escono da tale comunità. Dato l'insieme dei link interni a una comunità L_c e l'insieme dei link uscenti L_{oc}:

$$Conduttanza_c = \frac{2|L_{oc}|}{2|l_c| + |L_{oc}|} \tag{1.6}$$

Capitolo 2

Caso di studio

L'espressione Echo Chamber fa emergere, letteralmente, due componenti imprescindibili di questo fenomeno. Con *Echo* (i.e., eco) viene indicata l'opinione condivisa e polarizzata dei membri, che riecheggia all'interno del sistema. Con *Chamber* (i.e., camera) viene definito il sistema che permette alle opinioni dei membri di "echeggiare", rafforzandosi a vicenda.
Di seguito, viene definito il caso di studio scelto per la seguente tesi. Sono quindi spiegate le motivazioni che hanno portato alla scelta di Reddit come *chamber* d'analisi e delle ideologie democratiche e repubblicane durante la presidenza di Donald Trump come *echo*.

Reddit

Reddit[1], come asserito dal suo slogan "La prima pagina di Internet", non è una piattaforma di Social Networking tradizionale, ma piuttosto una sorta di grande forum che consente agli utenti di condividere link, opinioni, contenuti e notizie in tempo reale e su qualsiasi argomento. Reddit, fondato nel 2005 da Steve Huffman, è ad oggi il ventesimo sito web più visitato al mondo, il terzo se consideriamo solamente SNS[2].

Reddit ha una struttura atipica se confrontata con quella degli altri SNS. La piattaforma è organizzata in *subreddit*, ossia forum dedicati ad argomenti specifici (e.g., videogiochi, politica, programmazione, satira, scienza, cibo) ognuno identificato dal prefisso *r/*. Gli utenti, detti *Redditors*, se iscritti ad un subreddit possono condividere link, immagini o testo oppure

[1] https://www.reddit.com/

[2] https://www.alexa.com/topsites ultimo accesso: 01/07/2020

commentare post pubblicati da altre persone. Inoltre, ogni utente ha la possibilità di votare in modo positivo (i.e., *upvote*) o negativo (i.e., *downvote*) un post, determinandone il livello di visibilità.
Oltre a quelli sopraelencati, sono tre gli aspetti principali che rendono Reddit differente dalle altre piattaforme:

1. *Anonimato degli utenti:* per registrarsi alla piattaforma è necessario scegliere uno pseudonimo. A differenza di SNS come Facebook e Twitter, nessun'altra informazione personale come genere, età o immagine del profilo viene richiesta. Ne consegue che i *redditors* non hanno un proprio profilo personale dove condividere i propri contenuti. Selezionando lo pseudonimo di un utente, è possibile vedere solamente le attività, post o commenti, che ha svolto sui subreddits a cui è iscritto.

2. *Assenza di un sistema di amici/seguaci:* Reddit non prevede la possibilità di creare una connessione diretta tra due utenti, da cui deriva il concetto di amico o seguace, né prevede un sistema di messaggistica interno alla piattaforma. L'unica interazione che si crea tra gli utenti è data quindi dallo scambio di commenti relativi ad un post.

3. *Presenza di moderatori:* ogni subreddit ha uno o più moderatori che si occupano della loro manutenzione, i.e., impostare lo stile del subreddit, rispondere alle mail degli iscritti, rimuovere commenti o post giudicati inopportuni o spam.

Il motivo principale per cui abbiamo scelto Reddit come SNS d'analisi riguarda il suo rapporto con la sfera politica. Reddit è infatti una delle piattaforme più attive nelle discussioni politiche. L'anonimato degli utenti e il fatto che post e commenti non sono limitati in lunghezza, come in altre piattaforme (e.g., Twitter), spinge gli utenti a esprimere le proprie ideologie senza filtri [46]. Inoltre, la struttura organizzata in subreddits, oltre a semplificare la ricerca di contenuti relativi ad un argomento specifico, facilita la nascita e la crescita di comunità ideologicamente polarizzate. Un altro aspetto da tenere in considerazione riguarda la trasparenza della piattaforma: i dati di Reddit (e.g., post, commenti, utenti) sono pubblici e facilmente reperibili. Infine, questa scelta è stata determinata anche dalla volontà di apportare un contributo agli studi già esistenti sull'individuazione e l'analisi di EC online. Come discusso nella Sezione 1.2, in letteratura non esistono ricerche che svolgono analisi di questo tipo su Reddit.

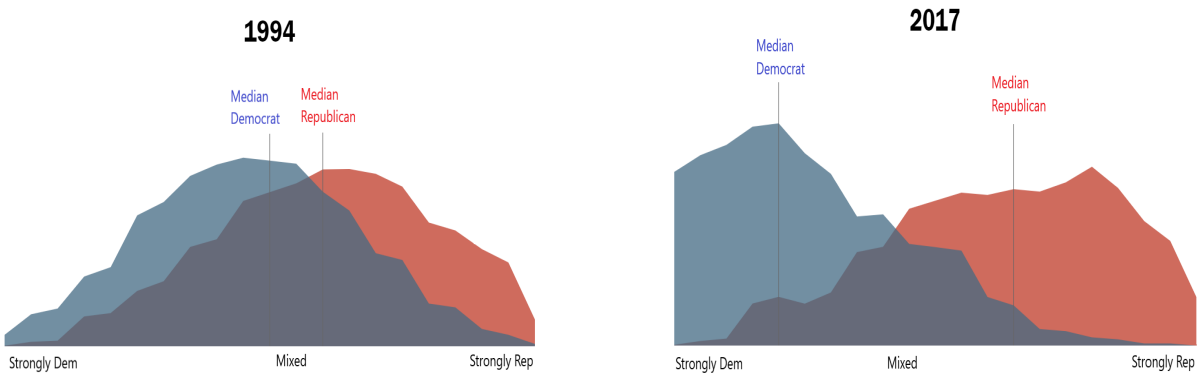


Figura 2.1: La crescita della polarizzazione nella società americana secondo una ricerca del *Pew Research Center*. L'area blu rappresenta la distribuzione delle ideologie dei Democratici, quella rossa dei Repubblicani.

Presidenza di Donald Trump

La scena politica americana, grazie alla sua struttura bipartitica (i.e., democratici e repubblicani) incarna, da decenni, la tensione delle ideologie binarie che la compongono favorendo la nascita di situazioni polarizzate [29].
Come mostrato in Figura 2.1, l'ascesa politica di Donald Trump, personalità altamente polarizzante, ha portato questa tendenza all'estremo, rendendo il dibattito politico tra i suoi sostenitori, tendenzialmente repubblicani, e i suoi oppositori, generalmente democratici, ancora più incivile e controverso [8]. Per queste considerazioni, come argomento di analisi di questa tesi abbiamo selezionato i primi due anni e mezzo di presidenza di Trump (i.e., Gennaio 2017-Settembre 2019).

Questa decisione ben si coniuga con la precedente, in quanto il 49.6% di traffico su Reddit proviene proprio dagli Stati Uniti[3]. Infine, per quanto riguarda la scena politica americana, gli studi mirati all'individuazione di polarizzazione e EC su SNS si concentrano principalmente sul periodo delle elezioni presidenziali del 2016 [33, 73], piuttosto che sugli anni di presidenza di Donald Trump.

[3]https://www.statista.com/statistics/325144/reddit-global-active-user-distribution/
ultimo accesso: 01/07/2020

Capitolo 3

I Dati

Il seguente Capitolo è dedicato alla descrizione delle fasi di estrazione e manipolazione dei dati necessari alla creazione dei Dataset finali.

Come discusso nel Capitolo 2 abbiamo scelto di utilizzare Reddit come piattaforma di analisi poiché a differenza di altri SNS (e.g., Facebook, Twitter) Reddit offre libero accesso ai suoi dati. Tutti i dati utilizzati per risolvere i vari task di questo studio sono stati estratti utilizzando la piattaforma Pushshift, creata nel 2015 da J. Baumgartner [10]. Quest'ultima offre, oltre a un *dump* mensile di tutte le attività avvenute su Reddit dal 2005 ad oggi, API [1] (i.e. *Pushshift Reddit API*) per permettere ai ricercatori di estrarre i dati tramite *queries* al database, senza la necessità di scaricare l'intero dump. Abbiamo scelto di usare tali API al posto delle Reddit API ufficiali in quanto i limiti nella dimensione di una singola query imposti da Pushshift sono molto meno stringenti rispetto a quelli di Reddit. Inoltre, le API di Pushshift offrono ai loro utenti l'accesso diretto a degli *endpoint*[2], aggregati rispetto al Dataset di cui si vuole usufruire (i.e., *subreddit, submissions, comments*).
In questa tesi sono stati utilizzati gli endpoint *Submissions* (Sezione 3.1) e *Comments* (Sezione 3.2), contenenti, rispettivamente, tutti i post e i commenti condivisi sulla piattaforma.

[1] Dall'inglese *Application Programming Interface*

[2] Un endpoint è sostanzialmente un URL tramite il quale le API possono accedere alle risorse di cui hanno bisogno

3.1 Submissions Dataset

Il Submissions endpoint consente di accedere ai 651,778,198 post condivisi all'interno dei subreddits presenti sulla piattaforma e filtrarli secondo le proprie necessità. Ogni post viene restituito dalle API sotto forma di un oggetto JSON. In Tabella 3.1 sono mostrate e descritte le proprietà principali di ogni oggetto JSON (i.e., post).

I dati estratti da *Submissions* sono stati utilizzati per svolgere due differenti task:

1. Creare una ground truth su cui allenare sia gli algoritmi di Topic Modeling (LDA), sia quelli di Text Classification (LSTM) con il fine di definire un approccio per predire la polarizzazione politica di un post (Sezione 3.1.1);

2. Creare dei topic relativi a questioni socio-politiche, in cui verificare la presenza di sistemi polarizzati, quali Echo Chamber (Sezione 3.1.2).

Tabella 3.1: Oggetto JSON rappresentante un post: proprietà, descrizione e tipo di dato.

Proprietà	**Descrizione**	**Tipo**
id	Identificatore del post (e.g., 5wzmao)	Stringa
url	Eventuale URL condiviso all'interno del post. Se non presente, questo campo è uguale a *permalink*	Stringa
permalink	URL che rimanda al post condiviso	Stringa
author	Nome dell'utente che ha scritto il post	Stringa
created_utc	Data e orario di condivisione del post (formato UNIX)	Intero
subreddit	Nome del subreddit in cui è stato condiviso il post	Stringa
subreddit_id	Id del subreddit in cui è stato condiviso il post (e.g., t5_3zfta)	Stringa
selftext	Testo del post condiviso	Stringa
title	Titolo del post condiviso	Stringa
num_comments	Numero dei commenti associati al post	Intero
score	Score del post, dato dal numero di *upvotes* meno quello di *downvotes*	Intero
is_self	Se il post contiene link (False) o solo contenuto testuale (True)	Booleano
over_18	Se il post è non adatto ai minorenni (True) o no (False)	Booleano
distinguished	Se il post è stato scritto da moderatori (True) o no (False)	Booleano
edited	Se il post è stato modificato (True) o no (False)	Booleano
stickied	Se il post è fissato nella parte alta del subreddit (True) o no (False)	Booleano
quarantine	Se il subreddit è in quarantena (True) o no (False)	Booleano

Tabella 3.2: Per ogni subreddit scelto, numero di post, numero medio di parole per post, numero di iscritti e ideologia politica.

Subreddit	# post	# medio parole	# iscritti	Ideologia politica
r/The_Donald	151,395	92.02	745,267	pro-Trump
r/Fuckthealtright	78,200	82.09	141,986	anti-Trump
r/EnoughTrumpSpam	73,168	79.34	98,750	anti-Trump

3.1.1 Creazione della Ground Truth

Il primo ostacolo affrontato in questo lavoro di tesi riguarda l'assenza di una ground truth. Né sulla piattaforma né in letteratura è stato possibile trovare utenti, post o commenti annotati in base all'orientamento politico democratico o repubblicano.

Per questo motivo abbiamo deciso di costruire una ground truth estraendo post da subreddits noti per essere altamente polarizzati rispetto a tali ideologie. Abbiamo scelto, inizialmente, r/The_Donald (T_D) come ground truth repubblicana/pro-Trump e r/Fucktheal-tright (FAR) per quella democratica/anti-Trump. Poiché come mostrato in Tabella 3.2 il numero di post condivisi su T_D è due volte maggiore rispetto a FAR, abbiamo selezionato r/EnoughTrumpSpam (ETS), un ulteriore subreddit polarizzato rispetto alle ideologie democratiche, in modo da avere un training set bilanciato fra le due classi target. I due subreddits scelti risultano infatti altamente correlati sia per gli utenti iscritti sia per la terminologia utilizzata[3]. La decisione di selezionare tali subreddits come ground truth è stata presa per diverse motivazioni. In primis, ci siamo basati sulla descrizione delle community fornita da ogni subreddit[4]. Inoltre, su Reddit sono stati trovati numerosi post in cui utenti affermano di essere stati rimossi da questi subreddits o di aver ricevuto *downvotes*, solamente per avere postato

[3]Statistiche per ogni subreddit:

https://subredditstats.com/r/EnoughTrumpSpam

https://subredditstats.com/r/Fuckthealtright

[4]Subreddit scelti:

https://www.reddit.com/r/The_Donald/ : "The_Donald is a never-ending rally dedicated to the 45th President of the United States, Donald J. Trump."

https://www.reddit.com/r/Fuckthealtright/ : "A subreddit dedicated to shitting on the racist, misogynist, anti-Semitic, adolescent clusterfuck known as the 'Alt-Right' "

https://www.reddit.com/r/EnoughTrumpSpam/ : "Because the amount of Trump spam is too damn high!"

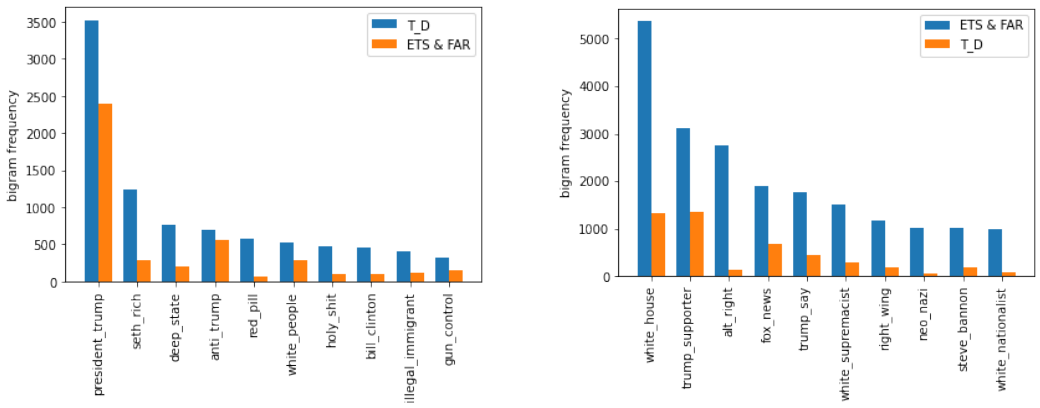


Figura 3.1: Confronto tra i 10 bigrammi più frequenti di ogni subreddit e la loro frequenza nel subreddit di ideologia opposta.

un'opinione divergente rispetto all'ideologia generale delle community. Infine, per verificare ulteriormente la validità della nostra scelta abbiamo estratto da entrambe le ground truth i 10 bigrammi più frequenti, analizzando la loro frequenza nella ground truth di ideologia opposta. Come illustrato in Figura 3.1, i bigrammi sembrano essere discriminanti e semanticamente correlati ai subreddits di appartenenza: espressioni quali *red pill* e *illegal immigrant* sono infatti largamente utilizzate dai sostenitori di Trump; così come *white supremacist* e *neo nazi* dai suoi oppositori.

Dunque, abbiamo estratto da tali subreddit tutti i post condivisi tra gennaio 2017 e settembre 2019. In una prima fase di filtraggio abbiamo rimosso tutti i post aventi un numero di *downvotes* maggiore di 50, per le motivazioni sopracitate. In seguito, osservando la distribuzione del numero di parole di ogni post, abbiamo notato che la gran parte dei post composti da un numero di parole minore di 6 presenta l'attributo *is_self* (in Tabella 3.1) uguale a False (i.e., post composto da uno o più contenuti multimediali). Per evitare di condizionare le performances dei modelli, con post ipoteticamente poco informativi o rilevanti, abbiamo preferito eliminarli.

Durante la fase di *Feature Selection* abbiamo selezionato gli attributi *id, title, selftext.* Gli ultimi due sono stati uniti in un unico campo *content*, poiché abbiamo osservato che spesso il campo *selftext* è vuoto o semplicemente un riferimento al titolo. Mediante questa procedura di selezione ci assicuriamo di fornire ai modelli utilizzati in seguito un input che rispecchia

tutto ciò che l'utente voleva veicolare. A tale Dataset è stato aggiunto un altro campo *Political_leaning* con le classi target (i.e., 0 per i post scritti nei subreddits anti-Trump e 1 per quelli pro-Trump). Infine, abbiamo applicato ai dati una pipeline standard di elaborazione del testo rimuovendo le *stop words*, la punteggiatura e trasformando il testo in minuscolo.
Questo Dataset, descritto in Tabella 3.2, è stato utilizzato come *Training Set* per l'approccio neurale descritto in Sezione 4.2 e come *corpus* per il task di *Topic Modeling* tramite LDA, descritto nella Sezione 4.1.1.

3.1.2 Scelta dei Topic d'Analisi

Il Dataset *Submissions* è stato utilizzato anche per la creazione di topic a cui applicare il modello scelto, con il fine di individuare e misurare il livello di polarizzazione dei post.
In questo lavoro, con il termine *topic* si intende un insieme di subreddits relativi ad uno stesso argomento socio-politico, ma caratterizzati da ideologie differenti. Abbiamo scelto tre topic ampiamente discussi in America, ognuno composto da sei subreddits:

1. *Gun Control*: Con tale terminologia ci riferiamo all'insieme di leggi e politiche che regolano la fabbricazione, la vendita, il possesso, la modifica o l'uso delle armi da fuoco da parte dei civili in America. Questo topic riguarda tutte le discussioni riguardanti la legittimità della legalizzazione delle armi e delle azioni che ne conseguono;

2. *Minority Discrimination*: Con tale terminologia ci riferiamo a quei gruppi sociali che all'interno di una società non costituiscono una realtà maggioritaria, divenendo quindi oggetto di discriminazione. Sono un esempio le minoranze etniche, religiose e di genere;

3. *Political Sphere*: Con tale espressione ci riferiamo a tutte le realtà politiche esistenti in America e alle ideologie che ne conseguono.

Nella scelta dei subreddits abbiamo cercato, per ogni topic, di selezionare *community* aventi ideologie differenti fra loro in quanto, per definire un sistema polarizzato, è necessario studiare come questo interagisce con comunità aventi opinioni divergenti.

Per ogni topic sono stati estratti i post condivisi tra gennaio 2017 e settembre 2019 costruendo tre Dataset corrispondenti (in Tabella 3.3). Ai Dataset sono stati applicati gli stessi

Tabella 3.3: Per ogni subreddit appartente ad un topic: numero di post, numero di utenti unici e numero di iscritti.

Topic	Subreddit	# post	# utenti	# iscritti
Gun Control	r/guncontrol	4,561	1,581	4,561
	r/antiwar	4,753	514	4,753
	r/liberalgunowners	8,436	3,140	8,436
	r/gunpolitics	9,491	3,172	9,491
	r/Firearms	26,128	9,794	26,128
	r/guns	20,204	10,694	113,204
Minority Discrimination	r/MensRights	59,516	15,786	266,662
	r/KotakuInAction	43,594	8,208	120,633
	r/metacanada	60,008	4,650	37,642
	r/racism	34,094	4,471	28,278
	r/AgainstHateSubreddits	16,314	7,099	103,812
	r/Anarchism	7,462	9,221	174,958
Political Sphere	r/esist	58,287	10,711	125,938
	r/democrats	41,042	7,517	133,297
	r/MarchAgainstTrump	34,998	6,976	80,450
	r/Conservative	149,473	21,696	365,619
	r/Libertarian	111,674	21,445	395,817
	r/Republican	22,777	5,256	119,991

step di filtraggio, feature selection e text pre-processing descritti in dettaglio nella Sezione 3.1.1, ad eccezione dell'aggiunta delle classi target per ogni post.

A questo Dataset sono state applicate le metodologie mirate all'identificazione di sistemi polarizzati presentate nel Capitolo 5.

3.2 Comments Dataset

Il *Comments endpoint* consente di accedere ai 5,601,331,385 commenti fatti ai post dei *subreddits* presenti su Reddit. Come nel caso dei post, ogni commento viene restituito dalle API sotto forma di un oggetto JSON. In Tabella 3.4 sono mostrate e descritte le proprietà principali di ogni oggetto JSON (i.e., post).

Tabella 3.4: Oggetto JSON rappresentante un commento: proprietà, descrizione e tipo di dato.

Proprietà	**Descrizione**	**Tipo**
id	Identificatore del commento (e.g., dbglmaz2)	Stringa
author	Nome dell'utente che ha scritto il commento	Stringa
link_id	Identificatore del post a cui è associato il commento (e.g., t3_51954r)	Stringa
parent_id	Identificatore del commento (o post) a cui il commento in analisi ha risposto (e.g., t1_dbu5bpp)	Stringa
created_utc	Data e orario di condivisione del commento (formato UNIX)	Intero
subreddit	Nome del subreddit in cui è stato condiviso il commento	Stringa
subreddit_id	Id del subreddit in cui è stato condiviso il commento (e.g., t5_3zfta)	Stringa
body	Testo del commento	Stringa
score	Score del commenti, dato dal numero di *upvotes* meno quello di *downvotes*	Intero
distinguished	Se il commento è stato scritto da moderatori (True) o no (False)	Booleano
edited	Se il commento è stato modificato (True) o no (False)	Booleano
stickied	Se il commento è fissato nella parte alta del subreddit (True) o no (False)	Booleano

I dati estratti da *Comments* sono stati utilizzati per costruire il grafo delle interazioni fra gli utenti al fine di verificare l'esistenza di Echo Chamber (Sezione 5.2).
Per ogni post presente nel Dataset *Submissions* sono stati estratti tutti i commenti ad esso associati. Tra gli attributi a disposizione (Tabella 3.4) abbiamo utilizzato *link_id* per identificare il post a cui è associato il commento e *parent_id* per identificare il commento/post a cui il commento estratto si sta effettivamente riferendo.

Capitolo 4

Identificazione della Polarizzazione Politica

Un aspetto centrale nella definizione di Echo Chamber è l'essere un sistema (i.e., insieme di utenti) fortemente polarizzato rispetto ad una determinata ideologia. Risulta quindi necessario definire una metodologia per determinare la polarizzazione politica di un utente. Gran parte degli studi esistenti hanno risolto questo task su altri SNS considerando caratteristiche demografiche dell'utente, la sua rete di amicizie e talvolta il suo orientamento politico (i.e., personaggio pubblico noto per l'appartenenza ad un partito). Su Reddit, essendo una piattaforma anonima, non è possibile seguire un ragionamento di questo tipo. Perciò definiamo la polarizzazione di un utente in base alla polarizzazione dei post che condivide.
In questo capitolo sono discussi i due approcci testati per predire la polarizzazione politica di un post. In Sezione 4.1 è presentata la prima metodologia composta, a sua volta, dalle fasi di Topic Modeling (Sezione 4.1.1) e di Text Classification (Sezione 4.1.2). Il secondo approccio, di tipo neurale, è descritto in Sezione 4.2. Infine, la Sezione 4.3 è dedicata alla discussione dei risultati ottenuti e alla scelta del modello finale.

4.1 Approccio "Corpus as a proxy"

Il primo approccio testato per classificare un post rispetto al suo grado di polarizzazione prevede due step sequenziali. Nella prima fase il problema viene modellato come un task di Topic

Modeling con il fine di estrarre da corpus polarizzati le keywords che caratterizzano il linguaggio di individui aventi ideologie democratiche o repubblicane. La seconda fase si pone l'obiettivo di classificare, tramite SVM, l'ideologia di un post in base alle keywords precedentemente estratte. Abbiamo scelto di definire questa procedura "Corpus as a proxy" in quanto i corpus polarizzati si comportano come un *proxy*: veicolano la polarizzazione e diventando il mezzo per ritrovarla in altri documenti testuali.

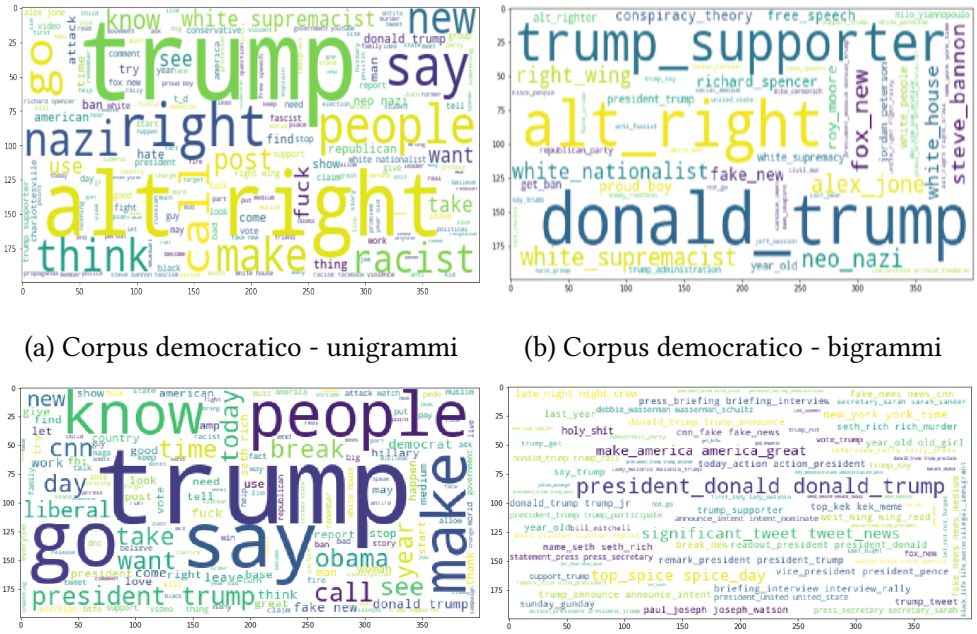
4.1.1 Estrazione di Keywords Discriminanti

Gli algoritmi di Topic Modeling sono dei modelli statistici, non supervisionati che, data una collezione di documenti testuali, tentano di estrarre i topic principali che li caratterizzano. Ogni topic estratto consiste in un insieme di keywords.

L'approccio utilizzato in questa tesi si basa su quest'ultimo aspetto. Dati dei documenti che assumiamo essere di ideologia democratica o repubblicana, abbiamo estratto gli insiemi di keywords che li distinguono, utilizzando LDA come algoritmo di Topic Modeling.
Abbiamo utilizzato il Dataset descritto in Sezione 3.1.1 per creare i due corpus su cui allenare LDA. I 151,368 post aventi classe target 0 costituiscono il corpus democratico, mentre i 151,395 con target 1 compongono il corpus repubblicano. Ogni post del Dataset è stato pre-processato, tokenizzato (i.e., diviso nelle parole che lo compongono) e lemmatizzato (i.e., la forma flessa di ogni parola è stata sostituita dalla sua forma canonica) seguendo la procedura descritta in Sezione 1.2.1.

Per processare ed interpretare una collezione di documenti LDA richiede che ogni testo sia rappresentato come un vettore numerico avente una lunghezza fissa. Per questo motivo abbiamo utilizzato il modello *Bag of Words* (BoW), trasformando ogni testo in una lista di tuple. Ogni tupla è composta da un indice univoco, che rappresenta la parola, seguito dal numero di volte che quella parola compare nel testo. Ad esempio la frase "america television say trump want america great" è rappresentata vettorialmente come [(0,2), (1,1), (2,1), (3,1), (4,1), (5,1)].

Inoltre, per eseguire LDA è necessario specificare il numero di topic da estrarre dai corpus creati. In letteratura [86], tale valore viene stimato attraverso metriche di *Coherence*. Tali metriche assegnano ad ogni topic un punteggio basato sul grado di somiglianza semantica tra le keywords che lo compongono: maggiore è la loro somiglianza più alto sarà il punteggio.

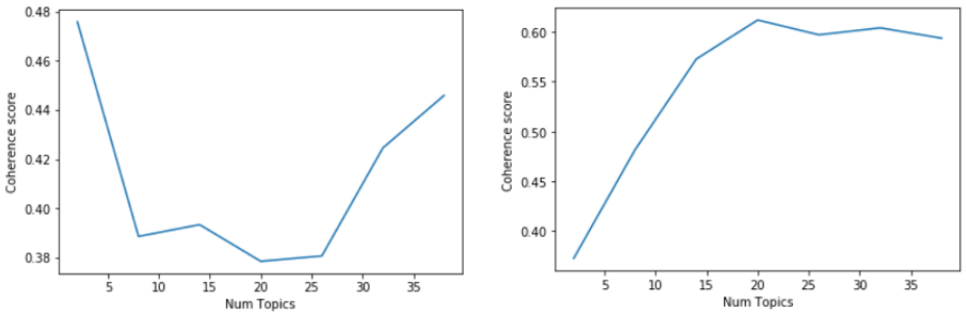


(a) Corpus democratico - unigrammi (b) Corpus democratico - bigrammi

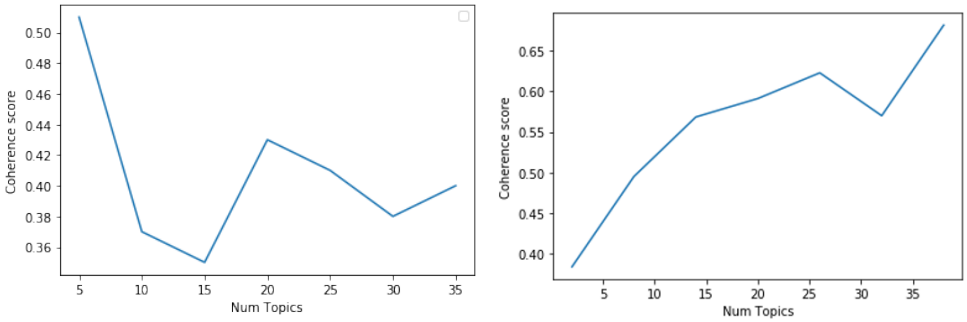
(c) Corpus repubblicano - unigrammi (d) Corpus repubblicano - bigrammi

Figura 4.1: Wordclouds raffiguranti gli unigrammi e i bigrammi più frequenti del corpus democratico (a,b) e repubblicano (c,d).

In questo lavoro abbiamo utilizzato il $C_V score$ che assegna un punteggio al topic tramite 4 step consecutivi: segmentazione dei dati in coppie di parole, calcolo della probabilità di una parola o di una coppia di parole, calcolo di una *Confirmation Measure* che quantifica il grado di correlazione tra due insiemi di parole e infine aggregazione delle singole Confirmation Measures in un unico Coherence score. Al fine di estrarre l'insieme di keywords più informativo, abbiamo deciso per entrambi i corpus di allenare LDA sia sul testo standard, ossia composto da un insieme di unigrammi (i.e., singola parola), sia sul testo inteso come insieme di bigrammi, ovvero due parole contigue in un testo. Per scegliere l'approccio migliore, abbiamo estratto da entrambi i corpus i primi 50 unigrammi e bigrammi più frequenti. Dalle *Wordclouds* in Figura 4.1 possiamo notare che, rispetto agli unigrammi, i bigrammi risultano decisamente più discriminanti nell'ideologia veicolata. Espressioni come *alt right, white supremacist* e *neo nazi* sono infatti tipiche del linguaggio democratico (Figura 4.1b), mentre unigrammi come *trump, say, people, make* risultano troppo generici per rappresentare il linguaggio dei due partiti (Figure 4.1a,4.1c).



(a) Corpus democratico - unigrammi (b) Corpus democratico - bigrammi



(c) Corpus repubblicano - unigrammi (d) Corpus repubblicano - bigrammi

Figura 4.2: Analisi del Coherence Score al variare del numero di topic di LDA per il corpus democratico (a,b) e repubblicano (c,d).

Inoltre, abbiamo allenato LDA sui due corpus rappresentati sia come insiemi di unigrammi che di bigrammi, analizzando il valore del Coherence Score al variare del numero di topic nel range [2,40]. Come mostrato dai grafici in Figura 4.2, anche in questo caso l'uso dei bigrammi porta ad un netto miglioramento delle performances del modello: i valori del Coherence Score si aggirano intorno al 50% per gli unigrammi (Figure 4.2a,4.2c) mentre raggiungono il 60-65% per i bigrammi (Figure 4.2b,4.2d).

In base a queste considerazioni abbiamo deciso di allenare i due modelli finali con i corpus rappresentati da bigrammi. Considerando l'andamento del Coherence Score al variare del numero di topic (Figure 4.2b,4.2d) abbiamo scelto un numero di topic pari a 20 per il corpus democratico e pari a 26 per il corpus repubblicano. LDA è stato implementato utilizzando la libreria *Gensim*[1].

[1]https://radimrehurek.com/gensim/models/ldamodel.html

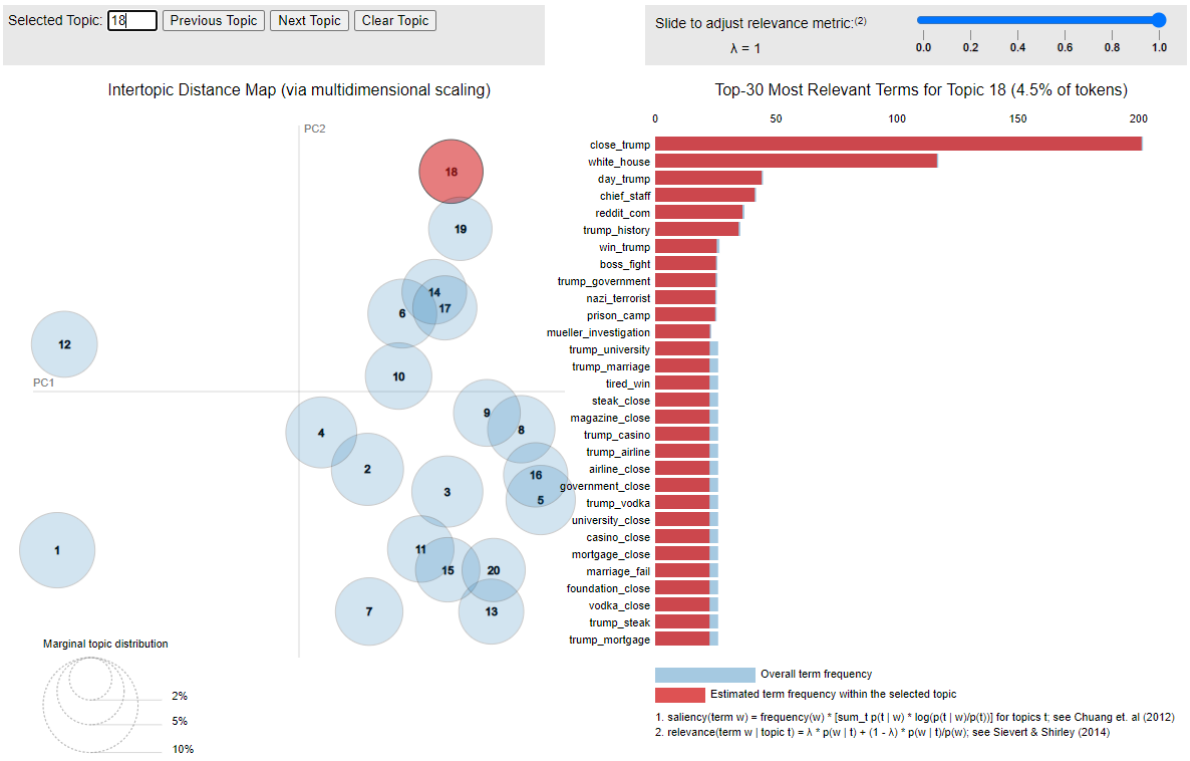


Figura 4.3: Topic e keywords estratte dal corpus democratico tramite LDA.

La Figura 4.3, realizzata con *pyLDAvis*[2], mostra uno screenshot della visualizzazione interattiva dei topic estratti dal corpus democratico (a destra) e delle relative keywords più rilevanti (elencate a sinistra). Ogni cerchio rappresenta un topic e la grandezza della sua aerea è determinata dall'importanza del topic nell'intero corpus. La distanza tra due topic indica il loro grado di similarità. L'interfaccia permette di selezionare un topic (in rosso) e di visualizzare le 30 keywords più rilevanti, sia dal punto di vista del singolo topic, sia del corpus in generale.

4.1.2 Classificazione tramite Keywords

Questa fase si basa sull'idea di classificare un post, come polarizzato rispetto alle ideologie democratiche o repubblicane, in base alla presenza delle keywords estratte da LDA.

Abbiamo identificato 400 keywords per il partito democratico e 520 per quello repubblicano. Al fine di ottenere due insiemi disgiunti, la loro intersezione è stata eliminata (e.g.,

[2]https://github.com/bmabey/pyLDAvis

bigrammi quali *donald_trump, president_trump, justice_departement*). I due insiemi risultanti sono composti, rispettivamente, da 360 e 480 bigrammi.

Abbiamo quindi costruito il Dataset per allenare il classificatore SVM. Ogni post utilizzato nella fase precedente, inteso come insieme di bigrammi, è stato trasformato in un vettore di occorrenze di lunghezza pari al numero totale di keywords 840 (i.e., ogni indice i del vettore contiene il numero di occorrenze della i-esima keyword nel post). Ad ogni post è stata assegnata la classe target: 0 per corpus democratico, 1 per corpus repubblicano.

Al fine di trovare la configurazione di SVM che massimizzi le performances, abbiamo utilizzato *3-fold Cross Validation*, testando i valori [0.1,1,10,100] di *C*. Quest'ultimo è un iperparametro che controlla il trade-off tra gli errori nel training e nel test set, evitando che il modello sia troppo specifico rispetto ai dati osservati (i.e., *Overfitting*). Tali procedure sono state implementate grazie alla libreria *scikit-learn*[3]

Il classificatore migliore è quello ottenuto impostando *C* uguale a 10, avente un Accuracy del 60.8%, una Precision del 72.2% e una Recall del 20.7%. Riteniamo che le performances risultanti non sono del tutto soddisfacenti. Questo può essere dovuto alla sparsità dei vettori in input: il 66% dei vettori aventi target 0 e il 68% dei vettori con target 1, hanno infatti ogni valore a 0 (i.e., non contengono nessuna keyword). Per migliorare le performances, abbiamo rimosso i vettori nulli, sia dal training set che dal test set, per poi riallenare SVM. Nonostante le procedure messe in atto non osserviamo significativi incrementi nei risultati (i.e., Accuracy 62.3%).

4.2 Approccio Neurale

Di seguito, il problema di misurare la polarizzazione politica di un post viene modellato come un task di Text Classification seguendo un approccio di tipo Neurale. Tale metodologia prevede le fasi di *Model Selection* e *Model Evaluation*, che saranno discusse nelle sezioni successive.
Per allenare e testare il modello abbiamo utilizzato il Dataset bilanciato descritto in Sezione 3.1.1. Sono stati selezionati 242,763 esempi per il training set e 60,000 per il test set (i.e. il 20% della totalità dei dati).

[3]https://scikit-learn.org/stable/modules/svm.html

4.2.1 Architettura e Selezione del Modello

In un approccio di tipo neurale, la scelta dell'architettura del modello dipende strettamente dal tipo di input che vogliamo utilizzare (e.g., numerico, testuale, visuale). In questo lavoro di tesi, l'input consiste in una serie di testi, ognuno composto da una sequenza di parole (i.e., *data points*) aventi una valenza semantica.
Come discusso in Sezione 1.2.2, questa tipologia di input per essere correttamente interpretata da un rete neurale richiede una rappresentazione vettoriale che riesca a catturarne i tratti salienti. I tradizionali modelli di *Bag of Words* (BoW) non risultano adatti a questo scenario in quanto, oltre a ignorare la struttura sintattica di una frase, non riescono a catturarne a pieno la semantica (i.e., due frasi che esprimono lo stesso concetto utilizzando parole differenti sono rappresentate in BoW da due vettori completamente diversi). Abbiamo quindi preferito usare i *Word Embeddings* (WE) per rappresentare vettorialmente le parole che compongono un post. I WE riescono a codificarne il significato, le relazioni semantiche e i differenti tipi di contesto in cui vengono utilizzate.

Anche nella scelta del modello è opportuno considerare quanto appena detto. Le reti neurali standard, *Feed Forward Neural Networks*, non risultano adatte a data points correlati fra loro nel tempo o nello spazio. Questo perché, ogni volta che un data point viene processato, l'intero stato della rete viene azzerato, rendendo impossibile imparare le dipendenze esistenti tra le parole di un testo. Per questo motivo abbiamo scelto di utilizzare *Long Short-Term Memory*, un particolare tipo di *Recurrent Neural Network*, capace di modellare il significato e le dipendenze che intercorrono tra le parole di una frase, grazie alla conformazione delle celle di memoria alla base della sua struttura.

Prima di descrivere l'architettura scelta per il modello è necessario sottolineare che i WE richiedono che i dati di input siano codificati come interi, ossia che ad ogni parola sia associato un indice che la rappresenti in modo univoco. Perciò, abbiamo creato un vocabolario basato sulla frequenza della parole: gli indici sono assegnati in ordine decrescente di frequenza (l'indice 0 non viene utilizzato). Ogni testo viene trasformato in un vettore di interi, sostituendo ogni parola con l'indice corrispondente (e.g., la frase "president trump says" diventa [5,1,39]). Abbiamo utilizzato l'intero vocabolario per allenare il modello. Infine, i vettori ven-

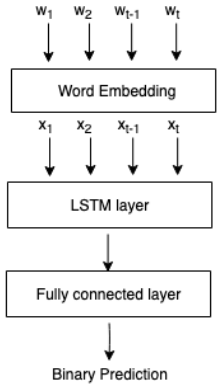


Figura 4.4: Ad alto livello, l'architettura utilizzata per l'approccio neurale.

gono trasformati in modo da avere tutti la stessa lunghezza di 100: le sequenze più lunghe sono troncate, mentre in quelle più corte viene inserito l'indice 0 al posto dei valori mancanti. L'architettura del modello (Figura 4.4) può essere suddivisa nelle seguenti componenti:

1. *Embedding Layer:* Questo layer prende come input le sequenze di interi w_1,..., w_t create precedentemente e le trasforma in un vettore x_t denso di dimensione pari a 100 (dove t indica la lunghezza della sequenza). Abbiamo deciso di analizzare le variazioni delle performances del modello al variare della tipologia di WE (i.e., *Pretrained WE GloVe* e *Learned WE* calcolati direttamente dalle sequenze di input).

2. *LSTM Layer:* Questo layer è composto da più unità LSTM, ognuna contenente una cella di memoria. Tali celle codificano le informazioni osservate, di volta in volta, dall'input attraverso il complesso meccanismo dei *gates* descritto in Sezione 1.2.2. Per evitare che il modello sia troppo specifico rispetto ai dati osservati (i.e., *Overfitting*), abbiamo utilizzato una tecnica di regolarizzazione detta *Dropout Regularization* che consiste nell'ignorare, durante l'allenamento del modello, un certo numero di neuroni selezionati casualmente.

3. *Ouput Layer:* Questo layer, totalmente connesso, è composto da un singolo neurone con il fine di eseguire predizioni binarie. Abbiamo quindi utilizzato la Sigmoide come funzione di attivazione per avere un output nel range [0,1]: ai post aventi un *Prediction Score* maggiore o uguale a 0.5 viene assegnata la classe target 1 (i.e., pro-Trump), agli altri 0 (i.e., anti-Trump). Come funzione di perdita (i.e., *loss*) abbiamo utilizzato *Binary Cross-Entropy* e come ottimizzatore *Adam*.

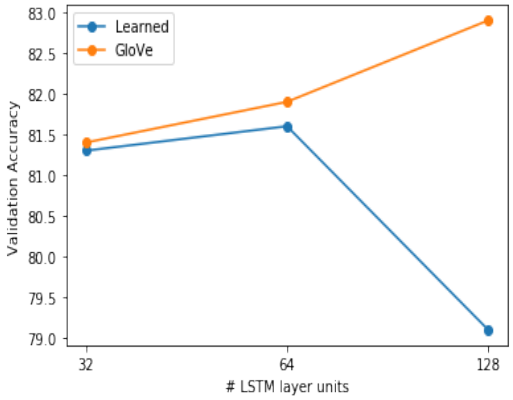


Figura 4.5: Per ogni tipologia di WE, analisi dell'Accuracy sul Validation set al variare del numero di unità di LSTM.

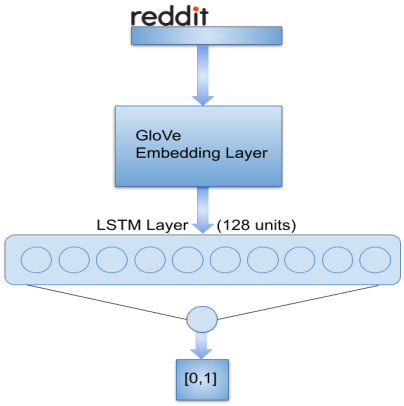


Figura 4.6: Architettura del miglior modello: GloVe Word Embeddings di 100 dimensioni e 128 unità di LSTM.

Per identificare il modello che raggiunge le migliori performances, abbiamo utilizzato *3-fold Cross Validation* e testato differenti configurazioni degli iper-parametri, variando il numero di unità di LSTM [32,64,128] e la tipologia di WE (i.e., *Pretrained WE GloVe* e *Learned WE*). La dimensione dei WE è stata fissata a 100. Per ogni configurazione, il modello è stato allenato per 10 epoche. La trasformazione dei post in vettori numerici e l'implementazione della rete neurale sono state svolte tramite la libreria *Keras*[4].

In Figura 4.5, per entrambe le tipologie di WE, viene analizzata l'accuracy del modello sul Validation Set al variare del numero di unità di LSTM utilizzate. Nonostante le due tipologie di WE con unità di LSTM pari a 32 e 64 mostrano risultati simili in accuratezza, impostando il numero di tali unità a 128 si osserva invece un miglioramento nell'accuratezza di GloVe e un peggioramento in quella dei Learned WE.

Il modello con le migliori performances è ottenuto utilizzando GloVe Pretrained WE e 128 unità di LSTM che, dopo 8 epoche, raggiunge un'accuratezza del 84.6% sul training set e del 82.9% sul validation set (Figura 4.6).

[4]https://keras.io/

Tabella 4.1: Performances del modello sul test set e sui topic Gun Control, Minority Disccrimination, Political Sphere.

	# post	Accuracy	Precision	Recall	F1-score
Test Set	60,000	0.843	0.841	0.841	0.841
Gun Control	2,411	0.712	0.751	0.692	0.720
Minority Discrimination	4,839	0.732	0.823	0.735	0.762
Political Sphere	46,339	0.721	0.701	0.732	0.711

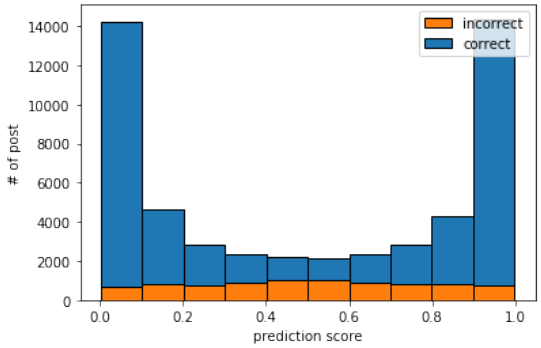


Figura 4.7: Distribuzione dei post classificati correttamente (blu) e erroneamente (arancione) in base al Prediction Score.

4.2.2 Valutazione del modello

In fase di *Model Evaluation* abbiamo deciso di valutare le performances del miglior modello sul *test set* di 60,000 esempi estratto dal Dataset iniziale. Come mostrato in Tabella 4.1, il modello raggiunge buoni risultati su tutte le metriche considerate, con un Accuracy del 84,3%. Dalla distribuzione in Figura 4.7, si nota che la densità delle predizioni sbagliate sia più alta nella parte centrale (i.e., nel range [0.3,0.7]), mentre i post più polarizzati in un estremo e nell'altro sembrano essere, in gran parte, predetti correttamente.

Questo approccio si pone l'obiettivo di misurare la polarizzazione di un post, quantificandone l'allineamento politico rispetto alle ideologie Democratiche e Repubblicane durante gli anni di Presidenza di Trump. Per questo, abbiamo deciso di validare il modello anche su argomenti, seppur sempre relativi alla sfera socio-politica, non apertamente correlati alla persona di Donald Trump. Come descritto in Sezione 3.1.2, sono stati scelti 3 topic su cui incentrare l'analisi: *Gun Control*, *Minority Discrimination* e *Political Sphere*. Per ogni topic sono stati se-

lezionati 6 subreddits, aventi ideologie divergenti, e i relativi post condivisi tra gennaio 2017 e settembre 2019.

Anche in questo scenario non è disponibile una ground truth per validare il modello scelto sui differenti topic. Per colmare questa mancanza, abbiamo deciso di validare il modello tramite gli utenti più polarizzati del Dataset utilizzato per l'allenamento. In dettaglio, abbiamo applicato il modello sia al Dataset utilizzato come training set sia al test set, in modo da avere per ogni post il suo Prediction Score (i.e., il suo livello di polarizzazione rispetto alle due ideologie politiche considerate). A questo punto è stato calcolato calcolato, per ogni utente, il *Polarization Score*, PS_u, definito come:

$$PS_u = \frac{\sum_{i=1}^{n} \text{Prediction Score}(P_i)}{n} \tag{4.1}$$

dove P_i indica un post e n è il numero totale di post condivisi dall'utente.

In questa fase, riteniamo interessante studiare la capacità del modello nel predire la polarizzazione di un post (i.e., le ideologie più estreme e radicate). Per questo motivo abbiamo selezionato gli utenti con PS_u <= 0.3 per l'ala democratica e PS_u >= 0.7 per quella repubblicana, escludendo tutti gli utenti con PS_u neutro.

Questa procedura ci ha permesso di selezionare 1,150 utenti: 550 repubblicani e 600 democratici. Abbiamo estratto i post scritti nei subreddits che compongono i 3 topic e selezionato quelli scritti dagli utenti selezionati. Ad ogni post è stata assegnata la classe target (i.e., 0 democratici, 1 repubblicani) in base all'orientamento dell'utente. È stato così possibile valutare le performances del modello sui 3 Dataset ottenuti, mostrate in Tabella 4.1.

Dai risultati ottenuti possiamo osservare che le performances del modello risentono sia del cambiamento di scenario effettuato nella scelta dei topic, sia della strategia di validazione utilizzata. Tale esito rispecchia le aspettative in quanto i topic sono volutamente composti da subreddits non per forza incentrati sulla persona di Trump ma piuttosto sulle conseguenze del suo operato. Inoltre, seguendo un approccio di validazione di questo tipo, non è stato possibile garantire il bilanciamento dei 3 Dataset sulle 2 classi considerate.

4.3 Analisi dei Risultati

Nelle sezioni precedenti sono stati discussi i due approcci testati per predire la polarizzazione politica di un post su Reddit.

La prima metodologia, in Sezione 4.1, con accuracy finale del 60.8% risulta non essere adatta a questa tipologia di task. La sparsità dei vettori di occorrenze, dati in input a SVM, dimostra che le keywords estratte da LDA risultano poco rappresentative del linguaggio tipicamente repubblicano e democratico.

Le performances dell'approccio neurale, in Sezione 4.2, sono decisamente migliori. Il modello finale, ottenuto utilizzando GloVe Pre-trained WE e 128 unità di LSTM, raggiunge sul test set un accuracy del 84.3%. Al meglio delle nostre conoscenze, non esistono studi analoghi svolti su Reddit con cui poter confrontare i risultati ottenuti. Tuttavia, su piattaforme differenti, quali Facebook [21] e Twitter [84], sono presenti ricerche mirate a predire l'orientamento politico di un post che ottengono un'accuratezza, rispettivamente, del 87% e del 85%, nonostante la presenza di una ground truth di utenti annotati. Inoltre abbiamo testato il modello su tre topic più generali pertinenti alla sfera socio-politica, raggiungendo un accuracy intorno al 72%.

I risultati ottenuti mostrano che il modello scelto riesce a generalizzare in modo sufficiente da permettere lo svolgimento degli step successivi di questo lavoro di tesi. Tuttavia, data l'assunzione fatta nella creazione della ground truth (Sezione 3.1.1) è possibile che il Dataset utilizzato per allenare il modello non sia totalmente rappresentativo. Infatti, abbiamo assunto che tutti i post contenuti in `r/The_Donald` siano polarizzati rispetto alle ideologie repubblicane mentre quelli in `r/Fuckthealtright` e `r/EnoughTrumpSpam` siano polarizzati rispetto alle ideologie democratiche. È però possibile che non tutti i post condivisi in tali subreddits abbiamo un contenuto polarizzato o connotato politicamente. Dalla distribuzione in Figura 4.7, notiamo che la densità delle predizioni errate effettuate dal modello è più alta nella parte centrale del grafico, ossia nel range $[0.3, 0.7]$. Inoltre, dagli esempi di predizioni mostrati in Tabella 4.2, notiamo che i post aventi un Prediction Score intorno a 0.5 non mostrano alcuna connotazione politica.

É importante sottolineare come, nell'ottica generale del progetto, non siamo interessati a co-

Tabella 4.2: Esempi di predizioni effettuate dal modello finale. Per ogni post: testo, topic e Prediction Score.

Post	**topic**	**score**
Tropical storm Barry: Obama has transformed his hatred for America into a new type~of treason	*Test set*	0.93
Trump's re-election crisis: His internal polls show it, national polls show it, and even a poll in reliably conservative Texas shows it.	*Test set*	0.02
Never forget: Hillary worked hard to keep wages of Haitian garment workers under 31 cents per hour	*Political sphere*	0.93
Today is a good day! Good morning community	*Political sphere*	0.52
Thanks everyone for supporting me in my fundraiser. I'm so glad	*Gun Control*	0.43
American soldiers aren't dying for our freedom in Syria, Iraq and Afghanistan. They're dying for nothing	*Gun control*	0.15
Poor Immigrants Are The Least Likely Group To Use Welfare, Despite Trump's Claims missing	*Discrimination*	0.03
Feminist deliberately acts like a condescending asshole to men. When they react like she's being an asshole, declares that men are an issue and Buzzfeed trends it.	*Discrimination*	0.90

struire un classificatore con risultati da stato dell'arte, bensì alla creazione di un modello che permetta di individuare post particolarmente polarizzati. Per tali considerazioni, dato un post, fissiamo due *threshold* (i.e., limiti) per definire la sua polarizzazione, limitando la presenza di falsi positivi e falsi negativi: i post con Prediction Score $<= 0.3$ sono considerati polarizzati rispetto alle ideologie democratiche mentre quelli con Prediction Score$>= 0.7$ sono considerati polarizzati rispetto alle ideologie repubblicane. I restanti vengono identificati come poco polarizzati (neutri).

Capitolo 5

Identificazione di Echo Chamber

Dopo aver definito una metodologia per predire la polarizzazione politica di un post è possibile procedere all'identificazione di sistemi polarizzati, quali le Echo Chamber. In questo scenario, abbiamo deciso di verificare l'esistenza di tale fenomeno in tre differenti topic, di argomento socio-politico, descritti in Sezione 3.1.2: *Gun Control, Minority Discrimination, Political Sphere*. Questo capitolo è così suddiviso: in Sezione 5.1 è presentato un approccio esplorativo volto a comprendere se i topic scelti si prestano a questo tipo di analisi, ossia se gli utenti tendono a polarizzarsi verso posizioni repubblicane o democratiche; in Sezione 5.2 presentiamo le metodologie utilizzate per l'effettiva identificazione di sistemi polarizzati; infine nella Sezione 5.3 sono illustrati e discussi i risultati ottenuti.

5.1 Analisi Esplorativa

Prima di seguire un approccio mirato all'identificazione di EC, abbiamo condotto un'analisi esplorativa su i tre topic selezionati per assicurarci che ben si prestassero ad un lavoro di questo tipo. Per questo motivo, abbiamo verificato la tendenza dei topic ad essere suddivisi in gruppi di utenti aventi una polarizzazione simile.

Per le motivazioni discusse nella Sezione 4.2.2, durante la validazione del modello abbiamo utilizzato un campione dei post appartenenti ai tre topic. In questa fase, abbiamo invece estratto tutti i post condivisi durante i primi due anni e mezzo di Presidenza di Trump. In Sezione 3.1.2, per ogni topic sono descritti i subreddits che li compongono, in termini di utenti

e post. In seguito, il modello è stato applicato a tali post in modo che ad ognuno sia associato il suo Prediction Score (i.e., il suo livello di polarizzazione rispetto alle ideologie democratiche e repubblicane).

Dato che ogni topic è composto da subreddits eterogenei, abbiamo utilizzato algoritmi di clustering per raggruppare gli utenti in base alla loro polarizzazione. In questa fase, le features disponibili per misurare la similarità tra gruppi di utenti sono il Prediction Score di un post e la data in cui il post è stato pubblicato. Dunque, per ogni topic abbiamo definito le *time series* (TS) di ogni utente che ha condiviso almeno un post e abbiamo applicato algoritmi di clustering alle TS. Ogni data point di una TS rappresenta il Prediction Score medio dei post condivisi dall'utente in un dato mese.
Inizialmente, abbiamo definito la TS di ogni utente considerando l'intero periodo temporale di analisi (Gennaio 2017 - Settembre 2019). La lunghezza massima di una TS è 30, tuttavia questo valore è ottenuto solo se un utente ha scritto almeno un post ogni mese nei due anni e mezzo considerati. Osservando le TS ottenute, è risultato evidente che i dati sono troppo sparsi per permettere un analisi efficace in un periodo di tempo così lungo. Nei topic analizzati, nessun utente ha una TS completa. Per questo motivo abbiamo ridotto la lunghezza della TS, selezionando gli utenti con $len(TS) >= 18$, ma anche in questo caso, siamo in grado di osservare solo l'1% degli utenti totali. Crediamo che sia piuttosto irrealistico che un utente condivida post in modo costante per due anni e mezzo.

Alla luce di queste considerazioni, abbiamo deciso di svolgere la nostra analisi su base semestrale, selezionando, per ognuno dei cinque semestri, tutti gli utenti che hanno scritto post in almeno quattro mesi su sei e definendone quindi la TS. In questo modo, riusciamo a selezionare circa il 24%, il 36% e il 41% degli utenti totali rispettivamente per *Gun Control, Minority Discrimination* e *Political Sphere*. Riteniamo che le percentuali ottenute su base semestrale siano sufficienti per condurre un'analisi preliminare su ogni topic. Poiché abbiamo selezionato anche le TS con lunghezza pari a 4, è stato necessario gestire i valori mancanti in modo che l'input dell'algoritmo di clustering sia composto da TS con uguale lunghezza. Per ogni topic ed ogni semestre abbiamo testato tre tecniche per gestire i valori mancanti:

1. *Last Observation Carried Forward (LOCF)* e *Next Observation Carried Backward (NOCB):* il data point mancante viene sostituito con il valore del dato che lo precede (LOFC) o, se

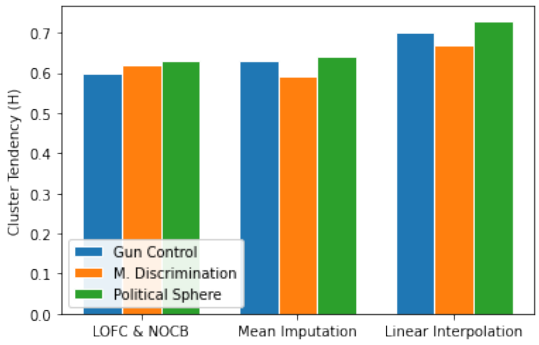


Figura 5.1: Cluster tendency dei tre topic in base alla metodologia di gestione dei valori mancanti utilizzata.

assente, con quello che lo segue (NOCB);

2. *Mean Imputaton:* il data point mancante viene sostituito dalla media dei valori degli altri dati presenti nella TS;

3. *Linear Interpolation:* il data point mancante viene sostituito dalla media tra i valori che lo precedono e quelli che lo seguono.

Per identificare la metodologia più adatta al trattamento delle TS, abbiamo osservato le variazioni nella tendenza dei dati ad essere suddivisi in cluster. Tale tendenza è stata misurata tramite la statistica di Hopkins (H), un indice statistico che mira a stimare la probabilità che i dati siano distribuiti in modo uniforme [6]. Se i dati sono ben clusterizzati H è uguale a 1; se H è uguale a 0 i dati sono distribuiti in modo regolare. Dalla Figura 5.1, è possibile osservare che i tre topic hanno una tendenza piuttosto alta ad essere clusterizzati e che il miglior metodo per gestire i valori mancanti risulta essere Linear Interpolation (i.e., $H \geq 0.65$).

Come algoritmo di clustering abbiamo scelto *K-means*. Questo algoritmo permette di suddividere i dati passati in input in k gruppi aventi simili caratteristiche seguendo tre step sequenziali: $a)$ sono create k partizioni e ad ognuna di esse sono assegnati i dati di input in modo casuale; $b)$ per ogni gruppo viene calcolato il suo centroide, dato dalla media dei punti assegnati a quel cluster; $c)$ viene definita una nuova partizione associando ogni dato al gruppo avente il centroide più vicino. Questi step vengono eseguiti fino a che non viene trovata la partizione migliore.

Per misurare la distanza tra due Time Series abbiamo utilizzato Dynamic Time Warping (DTW). Le operazioni di manipolazione delle TS e l'implementazione di K-means sono state svolte utilizzando la libreria *tslearn*[1].
Al fine di selezionare il numero ottimale di cluster (i.e., k), per ogni semestre di ogni topic, abbiamo analizzato i valori assunti dall'*Inertia* e dalla *Silhouette* al variare del numero di k nel range [2,20]. L'inertia è definita come la distanza quadratica media tra ogni istanza e il suo centroide più vicino e ha l'obiettivo di misurare la distanza tra le time series in uno stesso cluster. La silhouette, invece, misura la distanza tra time series appartenenti a cluster diversi.

Di seguito, per ogni topic e per i semestri 1,3 e 5, sono illustrati i *line charts* raffiguranti i centroidi dei cluster risultanti. Ogni grafico presenta due linee parallele all'asse delle ascisse: la linea blu rappresenta il limite (i.e., 0.7) sopra il quale il cluster viene considerato polarizzato rispetto alle ideologie repubblicane; la linea rossa definisce il limite (i.e., 0.3) sotto il quale il cluster viene considerato polarizzato rispetto alle ideologie democratiche.

In Figura 5.2 sono mostrati i centroidi risultanti per *Gun Control*. Il numero di cluster migliore è 3 per ogni semestre considerato. Dai line charts emerge che il topic risulta essere tendenzialmente più polarizzato rispetto alle ideologie repubblicane. In tutti i semestri, è presente un gruppo di utenti con TS piuttosto polarizzate verso le ideologie repubblicane e due gruppi le cui TS oscillano nell'area poco polarizzata.

Anche per il topic *Minority Discrimination*, in Figura 5.3, il miglior k è 3. In questo caso, i centroidi, nonostante siano meno stabili nel tempo, mostrano diversi livelli di polarizzazione. In ogni semestre è presente un cluster tendenzialmente più polarizzato rispetto alle idee repubblicane, uno più vicino a quelle democratiche e uno neutro con trend diversi in base al semestre.

Infine, per il topic *Political Sphere*, in Figura 5.4, il numero ottimale di cluster è 4. Ad eccezione del terzo semestre, notiamo la presenza di un cluster di stampo decisamente repubblicano e uno democratico. I restanti centroidi tendono invece a oscillare nell'area poco polarizzata.

In base ai risultati ottenuti, possiamo concludere che in tutti i topic i dati sono suddivisibili in gruppi polarizzati. Tuttavia, notiamo che i centroidi dei cluster non hanno un andamento

[1]https://tslearn.readthedocs.io/en/latest/index.html

costante all'interno dei singoli semestri: alcuni utenti non hanno un orientamento politico stabile durante l'arco temporale osservato.

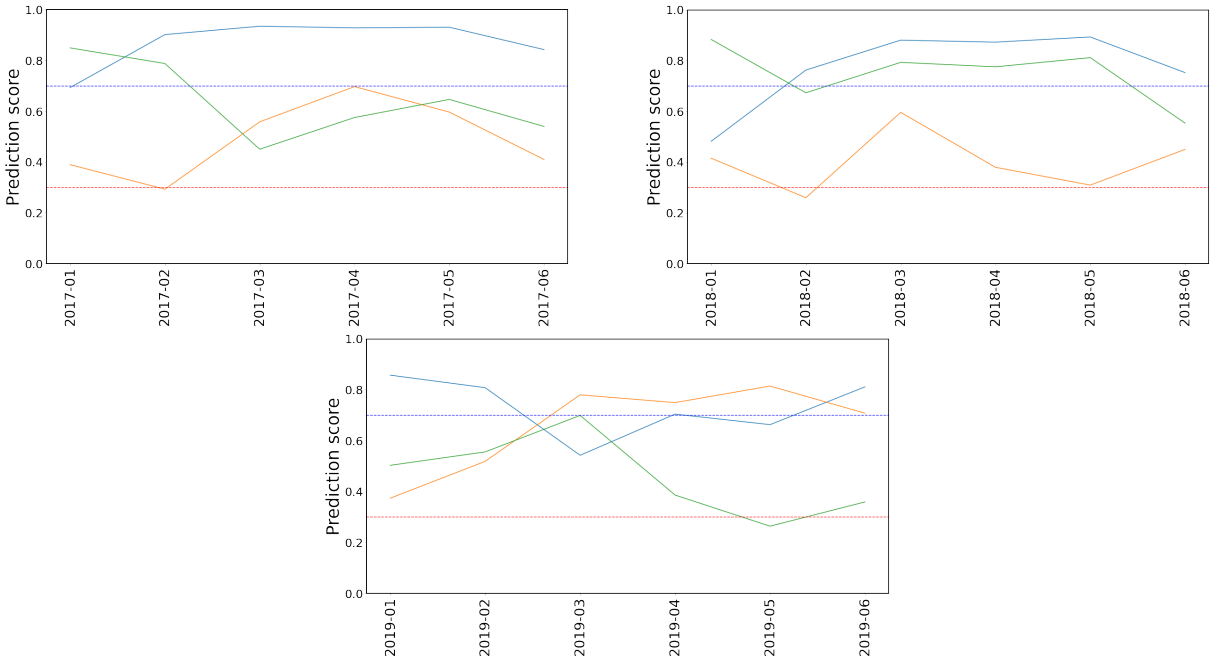


Figura 5.2: *Gun Control* - per i semestri 1,3,5 centroidi dei cluster risultanti

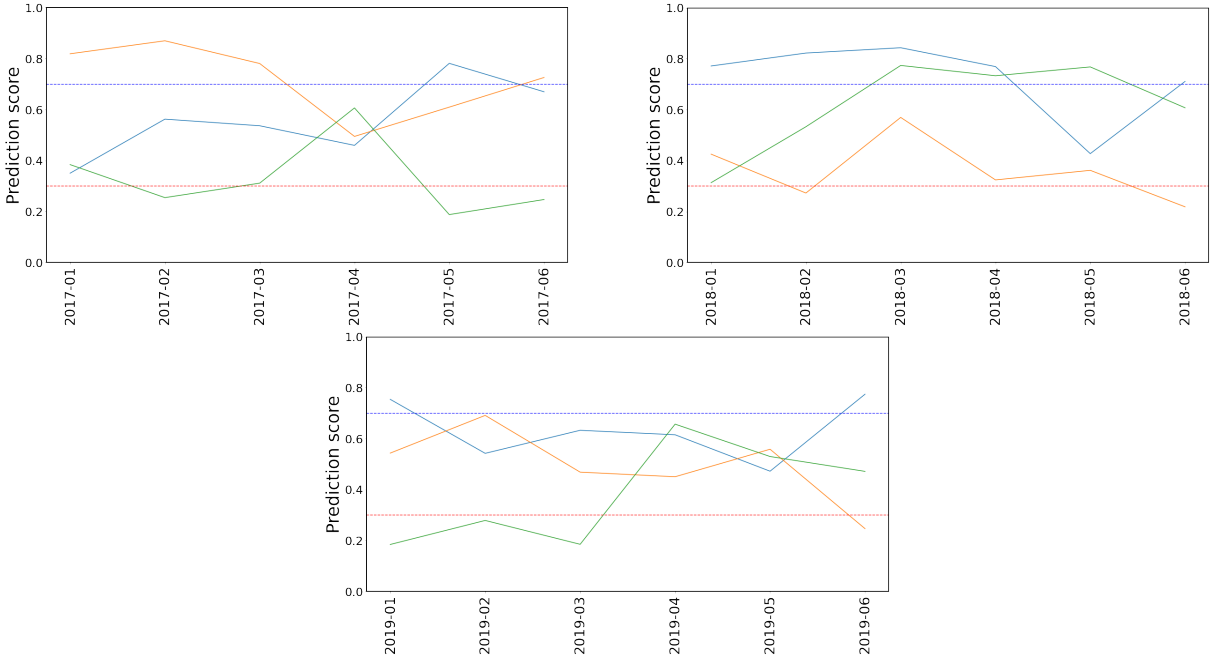


Figura 5.3: *Minority Discrimination* - per i semestri 1,3,5 centroidi dei cluster risultanti

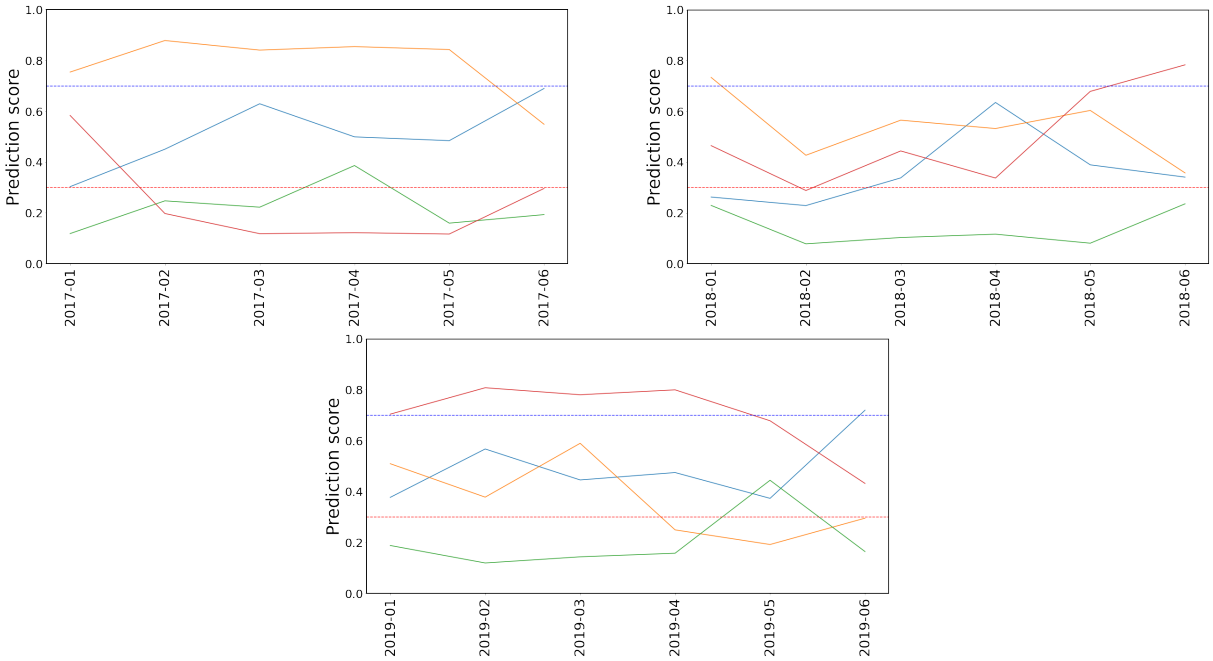


Figura 5.4: *Political Sphere* - per i semestri 1,3,5 centroidi dei cluster risultanti

5.2 Estrazione e selezione di sistemi polarizzati

L'analisi esplorativa effettuata sui topic scelti mette in luce la tendenza dei dati ad essere suddivisi in gruppi polarizzati. Sulla base ai questi risultati, abbiamo deciso di procedere all'effettiva identificazione di Echo Chamber. Inoltre, abbiamo constatato che l'analisi su base semestrale risulta più efficace e puntuale rispetto a quella annuale.

Il primo step affrontato riguarda la creazione di una rete di interazioni fra gli utenti, per ogni topic e semestre considerato. Per questo motivo, abbiamo: a) raggruppato gli utenti annotati in base al loro livello di polarizzazione; b) selezionato i post relativi a questi utenti; c) estratto i commenti effettuati a tali post dagli utenti annotati.

Per annotare gli utenti in base al loro grado di polarizzazione, per ognuno di essi abbiamo calcolato il *Polarization Score*, definito formalmente in 4.1. Per definizione, i membri di una stessa Echo Chamber condividono ideologie di cui sono forti sostenitori. I risultati del clustering hanno tuttavia messo in luce che non tutti gli utenti hanno un un grado di polarizzazione stabile nei mesi. Per questo motivo, abbiamo deciso di selezionare come utenti finali quelli con *Standard Deviation* dei Prediction Score ≤ 0.25. Seguendo questo procedimento, siamo in

Tabella 5.1: Per ogni topic e semestre, descrizione dei Network in termini di numero di commenti, nodi, link e grado medio.

Topic	Semestre	# commenti	# nodi	# link	grado medio
Gun Control	*1*	48,070	2,038	18,862	18.513
	2	51,242	2,279	20,143	17.677
	3	63,978	3,456	30,987	17.938
	4	56,657	3,374	26,286	15.581
	5	73,345	3,997	33,720	16.872
Minority Discrimination	*1*	21,853	1,497	10,104	13.498
	2	17,492	1,598	9,358	11.712
	3	18,207	1,657	8,577	10.352
	4	16,103	1,787	8,654	9.685
	5	21,169	2,136	11,641	10.899
Political Sphere	*1*	14,988	2,000	7,677	7.677
	2	16,515	1,875	8,388	8.947
	3	22,704	2,038	10,297	10.105
	4	29,15	2,493	12,739	10.219
	5	27,678	2,582	12,171	9.427

grado di selezionare un insieme di utenti con ideologia politica ben definita. Infine, abbiamo filtrato il dataset *Comments* (Sezione 3.2) selezionando i commenti effettuati dagli utenti finali a post condivisi da tali utenti.

I dati ottenuti ci hanno permesso di costruire la rete delle interazioni tra gli utenti finali. Un'interazione tra due utenti avviene quando, dati due utenti *i* e *j*, l'utente *i* condivide un commento in risposta ad un post o commento dell'utente *j* e viceversa. Abbiamo costruito un Network di tipo indiretto e pesato. Indiretto perché siamo interessati a sapere se c'è stata un'interazione tra *i* e *j*, piuttosto che alla direzione di tale interazione. Pesato in quanto, per ogni coppia di nodi *i* e *j*, vogliamo annotare il numero di interazioni esistenti w_{ij}. Inoltre, abbiamo eliminato tutti i *cappi* (i.e., link i cui estremi coincidono) in quanto non significativi ai fini del nostro lavoro. I network risultanti sono descritti in Tabella 5.1 in termini di numero di commenti, nodi, link e grado medio. In tutti i Network notiamo che il numero di commenti estratti è decisamente maggiore del numero di link pesati. Possiamo dunque evincere che gli utenti tendono ad interagire più volte con la stessa persona.

Per identificare sistemi polarizzati nei Network definiti, abbiamo applicato approcci di *Community Discovery*. Tali algoritmi sono direttamente comparabili con le tecniche di cluste-

ring esplorate nella sezione precedente. Entrambe sono metodologie non supervisionate che mirano a identificare gruppi omogenei in un contesto eterogeneo: gli approcci di CD partizionano i nodi di un Network in comunità, mentre le tecniche standard di clustering partizionano dati di vario tipo (e.g., time series) in cluster. Con il termine *comunità* ci riferiamo ad un insieme di nodi strettamente connessi tra loro piuttosto che con i nodi appartenenti ad altri insiemi di uno stesso Network. Come discusso in Sezione 1.2.3, non esiste una definizione condivisa di comunità e di conseguenza un unico approccio di CD. Per questo motivo, abbiamo testato differenti algoritmi di CD, descritti di seguito:

1. *Infomap* [88]: questo algoritmo appartiene agli approcci di CD che si basano sulla vicinanza tra nodi di un Network per definire una comunità (i.e., *Closeness*). Infomap si basa sulla combinazione di concetti di teoria dell'informazione (i.e., codifica di Huffman[2]) e di *random walks*. L'algoritmo usa il flusso di probabilità di random walks nel Network per determinare il flusso di informazioni in un sistema complesso reale. Il network viene partizionato in comunità tramite un'operazione di compressione di tale flusso di probabilità. Abbiamo scelto questo algoritmo per estrarre comunità ben separate tra di loro;

2. *Angel* [87]: questo algoritmo fa parte degli approcci di CD che identificano una comunità come un gruppo di nodi influenzabili dalla diffusione di proprietà o informazioni in un Network (i.e., *Diffusion*). Angel si basa sui concetti di *bottom-up partioning* (i.e., nodi con alta similarità vengono inclusi nella stessa comunità), di *Ego Network* (i.e., rete costruita su un nodo centrale i e composta da tutti i vicini di i e da tutti i collegamenti tra i e gli altri nodi), e di *Label propagation*. L'algoritmo restituisce in output macro-comunità ottenute dall'unione di micro-comunità simili fra loro. Tali comunità vengono definite sovrapposte, in quanto ogni nodo può contemporaneamente appartenere a due comunità diverse. Abbiamo deciso di utilizzare Angel in quanto, studiando interazioni sociali, vogliamo valutare il livello di sovrapposizione esistente tra le diverse comunità estratte;

3. *Louvain* [15]: questo algoritmo appartiene alla categoria di CD che definisce una comunità come un insieme di nodi densamente connessi tra loro (i.e., *Internal Density*). Lou-

[2]La codifica di Huffman, concetto alla base della teoria dell'informazione, è un algoritmo di compressione dei dati che codifica insiemi di stringhe basandosi sulla frequenza di ciascun carattere.

vain mira a massimizzare la modularità in due fasi: inizialmente, i singoli nodi vengono inclusi in una comunità valutando l'incremento della modularità; successivamente, viene creato un Network aggregato basato sulle partizioni ottenute nella fase precedente. Tali fasi vengono ripetute fino a che la modularità ha raggiunto il suo valore massimo. Abbiamo utilizzato Louvain in quanto uno dei criteri per definire un Echo Chamber è la densità delle interazioni esistenti;

4. *EVA* [22]: questo algoritmo è un' estensione di Louvain implementata per permettere di considerare, nell'estrazione di comunità, non solo la struttura di un Network ma anche gli attributi presenti sui singoli nodi. EVA mira ad ottimizzare due funzioni di qualità quali modularità e purezza con l'obiettivo di creare comunità dense e coese da un punto di vista di omogeneità degli attributi. La purezza di una comunità è data dal prodotto delle frequenze degli attributi più frequenti all'interno di tale comunità. Abbiamo testato questo approccio in quanto, oltre a creare comunità dense, ci permette di sfruttare l'annotazione degli utenti precedentemente calcolata.

Ogni algoritmo di CD testato è stato implementato tramite la libreria CDlib[3] e prende in input i Network definiti in Tabella 5.1. Poiché EVA richiede in input anche l'annotazione dei nodi, abbiamo assegnato ad ogni utente un attributo in base al suo *Polarization Score* (i.e. "Republican" se $PS_u \geq 0.7$, "Democratic" se $PS_u \leq 0.3$, e "Neutral" se $0.3 < PS_u < 0.7$).

Dopo aver estratto le comunità per ogni topic e semestre, è necessario stabilire quali caratteristiche deve soddisfare una comunità per essere considerata un Echo Chamber. Come discusso nella Sezione 1.2, in letteratura non esiste una definizione formale di EC. Gli studi effettuati in quest'area di ricerca [44, 41] ci suggeriscono che un EC, per essere considerata tale, deve essere composta da utenti con alto grado di similarità ideologica. Inoltre notiamo che, a livello strutturale, tale similarità si traduce in un alto numero di interazioni tra membri della stessa EC. I membri di EC diverse tendono invece a non connettersi fra loro.

In questa tesi, abbiamo dunque scelto di identificare Echo Chamber che mostrano una coesione interna sia ideologica che strutturale.
Per soddisfare il primo requisito, abbiamo calcolato il grado di polarizzazione di una comunità,

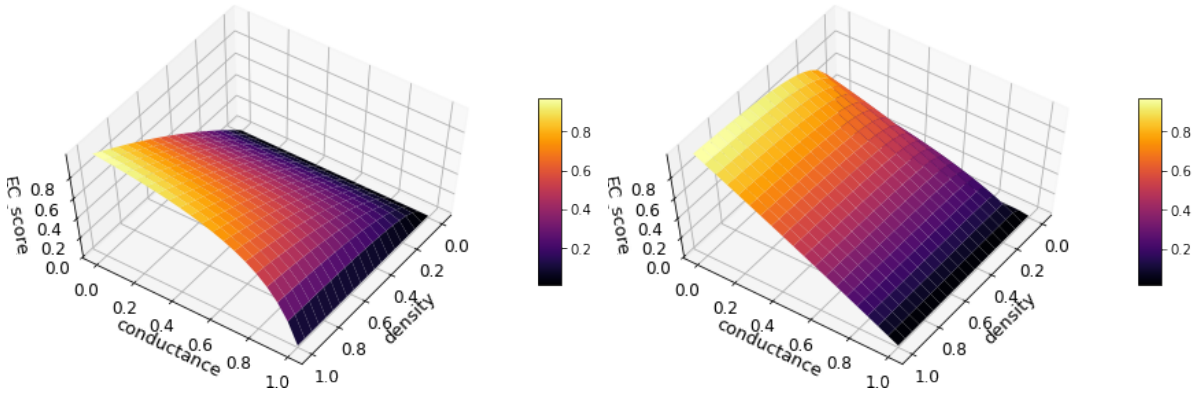
[3] https://cdlib.readthedocs.io/en/latest/index.html

come la media dei *Polarization Scores* dei suoi membri. Successivamente, abbiamo calcolato la *Standard Deviation* (std) dei Polarization Scores dei membri e fissato una threshold per tale valore pari a 0.25. Dunque, se una comunità ha una std dei polarization scores ≤ 0.25 soddisfa il requisito di coesione ideologica di un EC. Questa procedura è stata effettuata per tutti gli approcci di CD ad eccezione di EVA. Infatti, questo algoritmo restituisce in output comunità già annotate rispetto all'ideologia politica, perciò abbiamo valutato la coesione ideologia di queste comunità in termini di purezza: se la purezza risulta ≥ 0.7 la comunità è coesa ideologicamente.

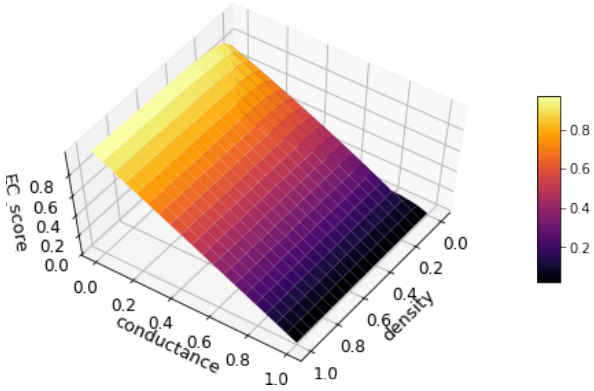
Per soddisfare il requisito di coesione strutturale di un EC, abbiamo valutato la densità dei link all'interno di una comunità (i.e., *Internal edge density*) e il volume dei link che escono dalla comunità (i.e., *Conduttanza*). Per una maggiore interpretabilità dei risultati, abbiamo definito una metrica per valutare se una comunità *c* può essere definita strutturalmente un Echo Chamber in termini di densità e conduttanza. Tale metrica, detta *EC_score*, è la media armonica pesata di densità e conduttanza:

$$EC\_score_c = (1 + \beta^2) \cdot \frac{densità_c \cdot (1 - conduttanza_c)}{(\beta^2 \cdot densità_c) + (1 - conduttanza_c)} \tag{5.1}$$

dove β è uguale a 4 e densità e conduttanza hanno valori nel range [0,1].
La metrica assume valore massimo, ovvero 1, quando la densità è massima e la conduttanza è minima: ciò significa che la comunità è molto densa e non presenta link che escono dalla comunità (i.e., soddisfa requisito strutturale EC). Viceversa, la metrica assume valore minimo, ossia 0, quando la conduttanza è massima e la densità è minima: ciò significa che la comunità non è coesa internamente. Inoltre, abbiamo assegnato un peso β alla conduttanza in quanto, definendo un EC, riteniamo più rilevante l'assenza di archi verso l'esterno della comunità piuttosto che la sua densità interna. Per scegliere il valore di β abbiamo osservato i valori assunti dalla funzione *EC_score* al variare di β (i.e., $0.5, 4, 10$). In Figura 5.5a, osserviamo che con un valore di β troppo piccolo la conduttanza e la densità influiscono in egual modo sulla funzione; viceversa con un valore troppo grande (Figura 5.5c) la densità diventa ininfluente e la funzione varia in base al variare della conduttanza. Per queste considerazioni abbiamo scelto un valore intermedio (β = 4) (Figura 5.5b).
Dopo aver definito il valore di β, osservando i valori assunti dalla funzione in Figura 5.5b riteniamo che il requisito strutturale di un EC sia soddisfatto se una comunità ha EC_score $\geq$



(a) $\beta = 0.5$ (b) $\beta = 4$



(c) $\beta = 10$

Figura 5.5: Funzione *EC_score* al variare dei parametri densità e conduttanza con β fissato.

0.5.

Le Echo Chamber sono quindi identificate selezionando le comunità estratte che soddisfano i requisiti strutturali e ideologici sopracitati.

5.3 Analisi dei Risultati

Dopo aver descritto la metodologia proposta per identificare Echo Chamber su Reddit (Sezione 5.2), in questa sezione analizziamo i risultati ottenuti. Per semplificarne la comprensione, definiamo le comunità estratte che soddisfano i requisiti strutturali e ideologici di un EC come `EC_community`, mentre quelle non idonee come `simple_community`.

Il processo di identificazione di `EC_community` a partire dalla totalità delle comunità estratte viene presentato, per ogni topic e semestre, tramite uno *Scatter Density plot*. In ogni visualizzazione sono quindi mostrate sia le `simple_community` (label 0) che le `EC_community` (label 1). Di seguito descriviamo le caratteristiche degli Scatter plot creati in base all'algoritmo di CD utilizzato.

Per gli algoritmi Infomap, Angel e Louvain abbiamo creato un'unica visualizzazione in quanto i risultati sono direttamente comparabili. Ogni comunità estratta è descritta in termini di EC_score (asse x) e Polarization Score (asse y). La dimensione di ogni comunità è espressa dall'area del cerchio. Le due linee parallele all'asse delle ascisse indicano le threshold di polarizzazione (*PS*). Se il *PS* di una comunità è ≥ 0.7, la comunità è polarizzata rispetto alle ideologie repubblicane; se la comunità ha $PS \leq 0.3$ viene definita polarizzata rispetto le ideologie democratiche; altrimenti non polarizzata e quindi neutrale. La linea parallela alle asse delle ordinate indica il limite oltre il quale la comunità soddisfa il requisito strutturale di un EC (i.e., EC_score ≥ 0.5). Per motivi di leggibilità, in queste visualizzazioni sono mostrate esclusivamente le comunità che soddisfano il requisito di coesione ideologica (i.e., std dei polarization score ≤ 0.25).

Le comunità estratte da EVA non possono essere descritte in termini di Polarization Score, in quanto, durante l'esecuzione dell'algoritmo, sono già annotate ideologicamente rispetto alla classe di maggioranza dei suoi nodi. Perciò nello Scatter Plot, ogni comunità ha due attributi: EC_score (asse x) e Purezza (asse y). La dimensione di ogni comunità è espressa dall'area del cerchio, la classe di maggioranza dal colore. La linea parallela all'asse delle ascisse indica il limite oltre il quale la comunità soddisfa il requisito di coerenza ideologica (i.e., purezza ≥ 0.7). Come nel caso precedente, la linea parallela all'asse delle ordinate indica il limite oltre il quale la comunità soddisfa il requisito strutturale di un EC.

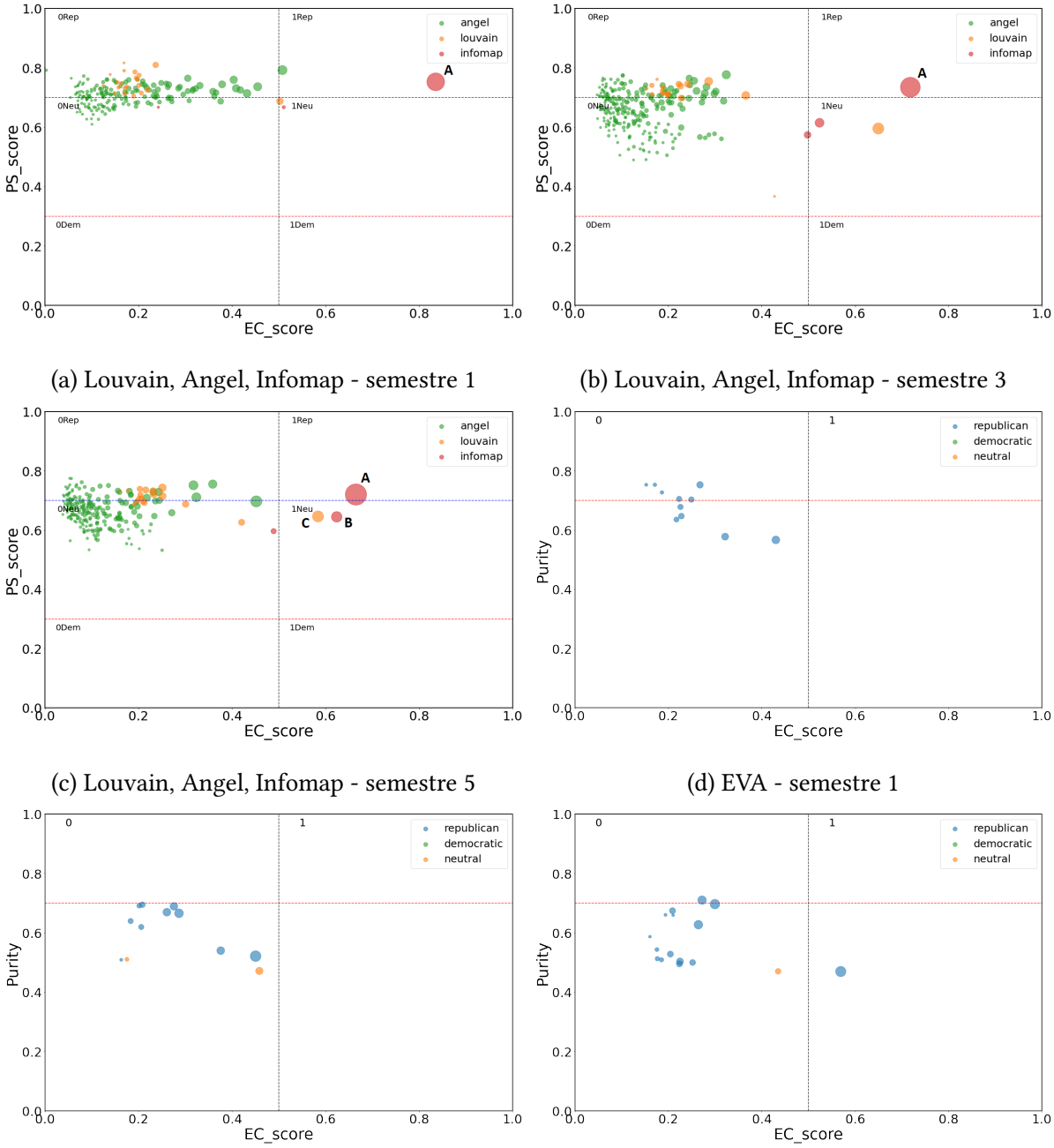
Per tutti gli algoritmi sono mostrate esclusivamente le comunità aventi una dimensione accettabile (i.e., numero di nodi/utenti ≥ 50). Per motivi di leggibilità presentiamo tali visualizzazioni per il primo, terzo e quinto semestre di ogni topic analizzato.

Inoltre, per ogni topic descriviamo, in forma tabellare, le EC più rilevanti in termini di algoritmo utilizzato, dimensione, standard deviation/purezza e EC_score. Ad ognuna di queste EC abbiamo assegnato una label in modo da poterla facilmente identificare all'interno dello Scatter Plot relativo.

Gun Control. L'unico algoritmo che riesce ad individuare Echo Chamber politicamente polarizzate è Infomap. Dalle visualizzazioni 5.6a, 5.6b, 5.6c emerge la presenza, in tutti i semestri, di una `EC_community` polarizzata rispetto alle ideologie repubblicane decisamente coesa internamente sia dal punto di vista ideologico (i.e., std nel range [0.17,0.18]) che strutturale (i.e., EC_score nel range [0.66,0.83]). Tali EC comprendono più della metà dei nodi totali dei Network considerati (Tabella 5.2). Sono inoltre individuate nel terzo e quinto semestre, sia da Infomap che Louvain `EC_community` poco polarizzate/neutre ma tendenti verso l'ala repubblicana (i.e., PS_score ≥ 0.60). Tali risultati sono in linea con quelli preliminari ottenuti clusterizzando le time series degli utenti (Sezione 5.1).

Tabella 5.2: *Gun Control* - Descrizione EC_community più rilevanti in termini di algoritmo utilizzato, semestre, label, dimensione, numero di nodi, std dei PS_score e EC_Score.

Algoritmo	semestre	label	Pol. Ideologica	# nodi	std	purezza	EC_score
Infomap	1	A	repubblicana	1814	0.173	-	0.835
Infomap	3	A	repubblicana	2247	0.182	-	0.718
Infomap	5	A	repubblicana	2580	0.185	-	0.664
Infomap	5	B	neutrale	645	0.235	-	0.623
Louvain	5	C	neutrale	709	0.227	-	0.582



(a) Louvain, Angel, Infomap - semestre 1

(b) Louvain, Angel, Infomap - semestre 3

(c) Louvain, Angel, Infomap - semestre 5

(d) EVA - semestre 1

(e) EVA - semestre 3

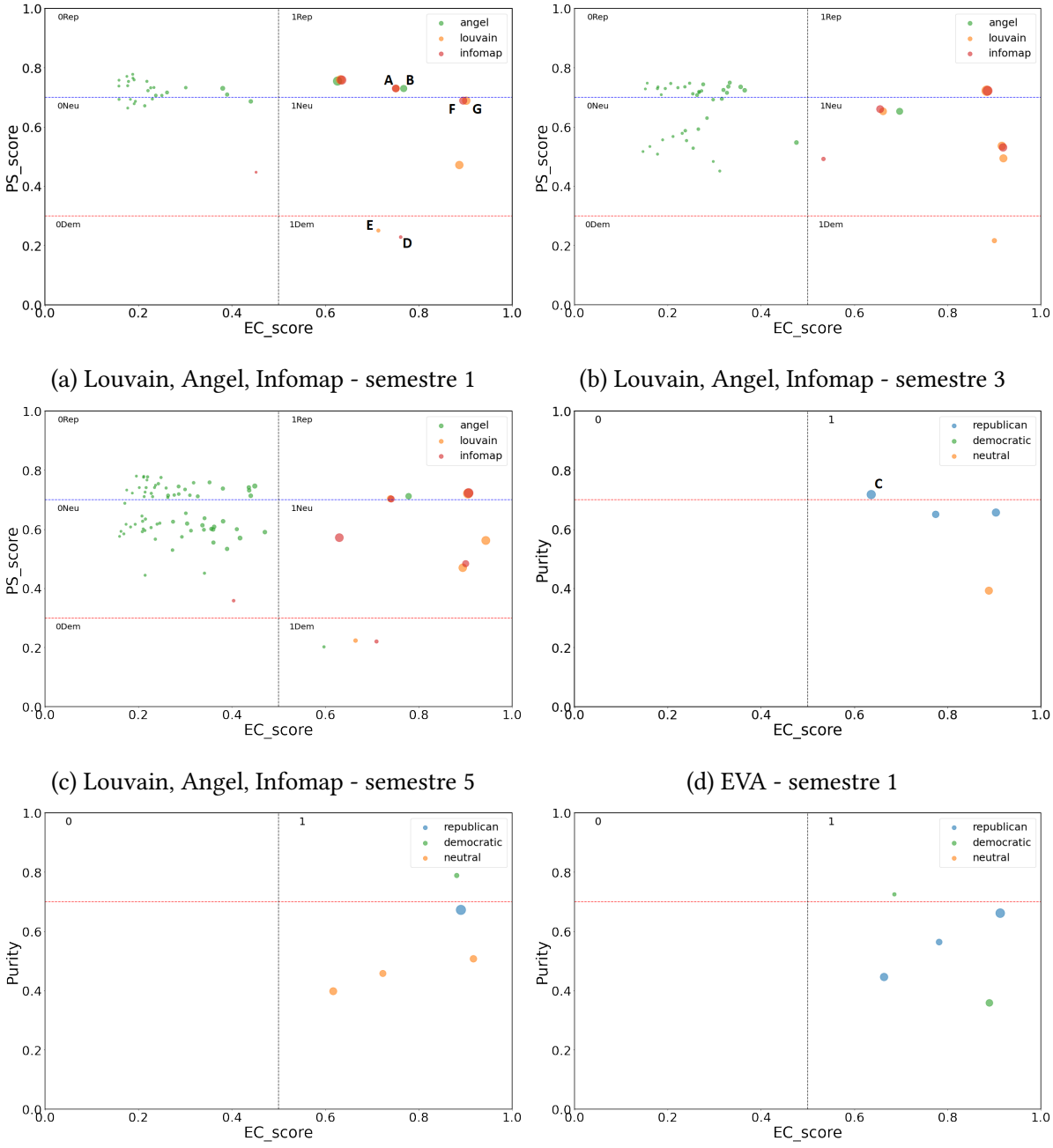
(f) EVA - semestre 5

Figura 5.6: *Gun Control* - simple_community e EC_community estratte in base all' algoritmo di CD utilizzato.

Minority Discrimination. In questo scenario tutti gli algoritmi riescono ad individuare EC politicamente polarizzate (Figura 5.7). Risulta inoltre interessante osservare che il topic presenta `EC_community` polarizzate sia rispetto alle ideologie democratiche che a quelle repubblicane. Le prime risultano essere però più piccole e con una coerenza ideologica interna meno definita rispetto alle seconde (Tabella 5.3). Notiamo inoltre una netta differenza nella coerenza strutturale tra le `EC_community` polarizzate (i.e., EC_score ≤ 0.76) e neutre (i.e., EC_score ≥ 0.89). Anche in questo caso i risultati ottenuti sono piuttosto simili a quelli forniti dal clustering: sono presenti gruppi omogenei di utenti in tutti e tre gli intervalli di polarizzazione presi in considerazione.

Tabella 5.3: *Minority Discrimination* - Descrizione EC_community più rilevanti in termini di algoritmo utilizzato, semestre, label, dimensione, numero di nodi, std dei PS_score e EC_Score.

Algoritmo	**semestre**	**label**	**Pol. Ideologica**	**# nodi**	**std**	**purezza**	**EC_score**
Infomap	1	A	repubblicana	287	0.173	-	0.750
Angel	1	B	repubblicana	249	0.166	-	0.770
Eva	1	C	repubblicana	421	-	0.722	0.644
Infomap	1	D	democratica	60	0.216	-	0.761
Louvain	1	E	democratica	70	0.211	-	0.713
Infomap	1	F	neutrale	322	0.241	-	0.895
Louvain	1	G	neutrale	354	0.234	-	0.901



(a) Louvain, Angel, Infomap - semestre 1

(b) Louvain, Angel, Infomap - semestre 3

(c) Louvain, Angel, Infomap - semestre 5

(d) EVA - semestre 1

(e) EVA - semestre 3

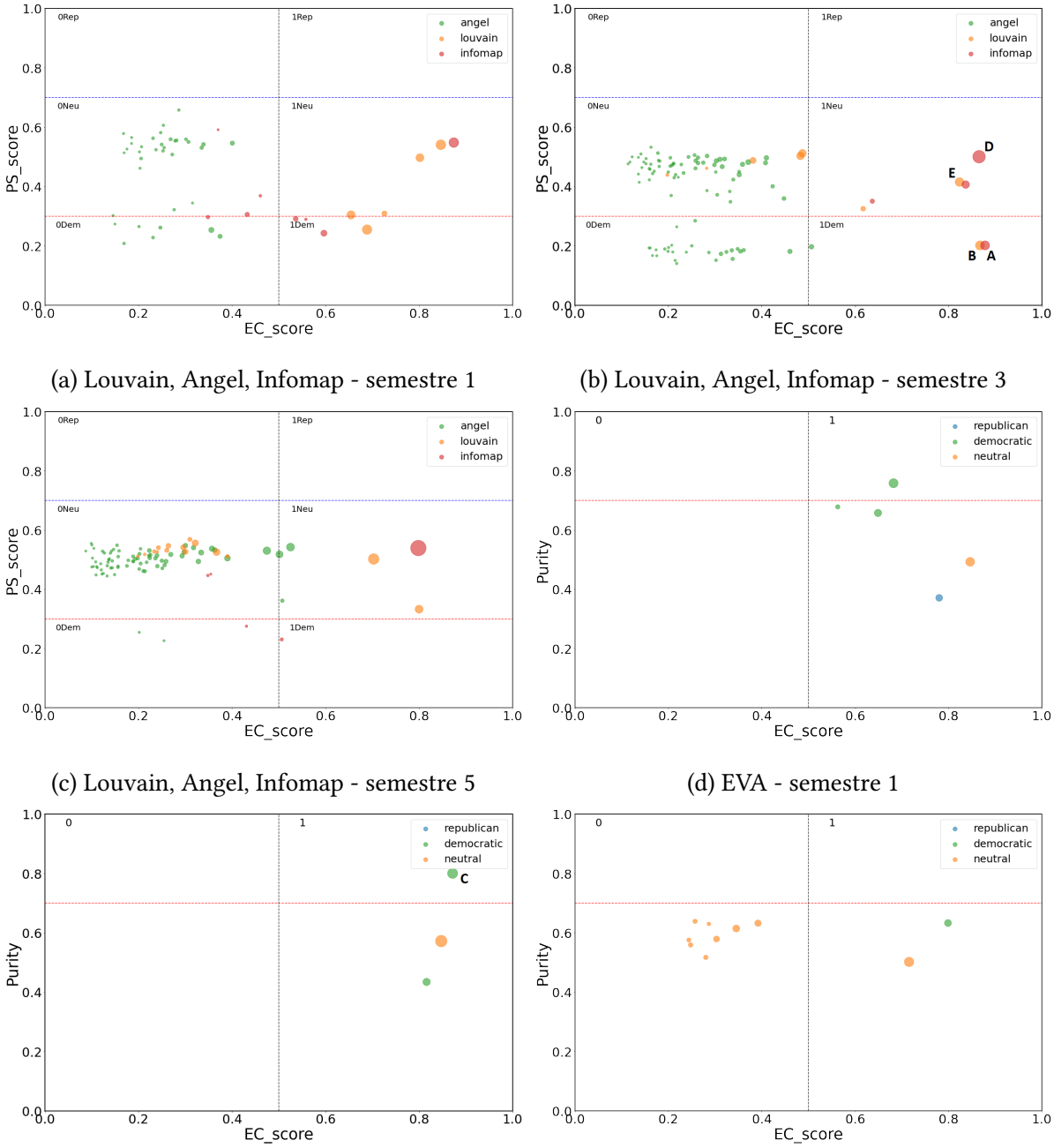
(f) EVA - semestre 5

Figura 5.7: *Minority Discrimination* - simple_community e EC_community estratte in base all'algoritmo di CD utilizzato.

Political Sphere. Ad eccezione di Angel, tutti gli algoritmi riescono a identificare `EC_community` polarizzate (Figura 5.8). In generale, il topic risulta essere polarizzato rispetto alle ideologie democratiche: le EC democratiche soddisfano pienamente sia i requisiti ideologici (i.e., std $\leq$ 0.173 e purezza $\geq$ 0.79) sia struttuali (i.e.,EC_score $\geq$ 0.86) (Tabella 5.4). Infomap e Louvain inoltre individuano in tutti i semestri `EC_community` neutre tendenti sia all'ala democratica che a quella repubblicana.

Tabella 5.4: *Political Sphere* - Descrizione EC_community più rilevanti in termini di algoritmo utilizzato, semestre, label, dimensione, numero di nodi, std dei PS_score e EC_Score.

Algoritmo	semestre	label	Pol. Ideologica	# nodi	std	purezza	EC_score
Infomap	3	A	democratica	472	0.173	-	0.878
Louvain	3	B	democratica	479	0.170	-	0.867
Eva	3	C	democratica	589	-	0.799	0.870
Infomap	3	D	neutrale	898	0.235	-	0.865
Louvain	3	E	neutrale	506	0.238	-	0.815



(a) Louvain, Angel, Infomap - semestre 1

(b) Louvain, Angel, Infomap - semestre 3

(c) Louvain, Angel, Infomap - semestre 5

(d) EVA - semestre 1

(e) EVA - semestre 3

(f) EVA - semestre 5

Figura 5.8: *Political Sphere* - simple_community e EC_community estratte in base all'algoritmo di CD utilizzato.

Nonostante le differenze fra i topic, notiamo dei trend generali. In primo luogo, possiamo asserire che l'approccio utilizzato in questo lavoro permette di individuare Echo Chamber polarizzate in tutti i topic considerati. Per quanto riguarda gli approcci di CD, in questo contesto specifico, Infomap risulta essere quello più performante. Questo algoritmo riesce ad individuare `EC_community` in tutti i topic con risultati più che soddisfacenti in termini di coerenza ideologica e strutturale: la maggioranza delle comunità estratte hanno un EC_score ≥ 0.5.

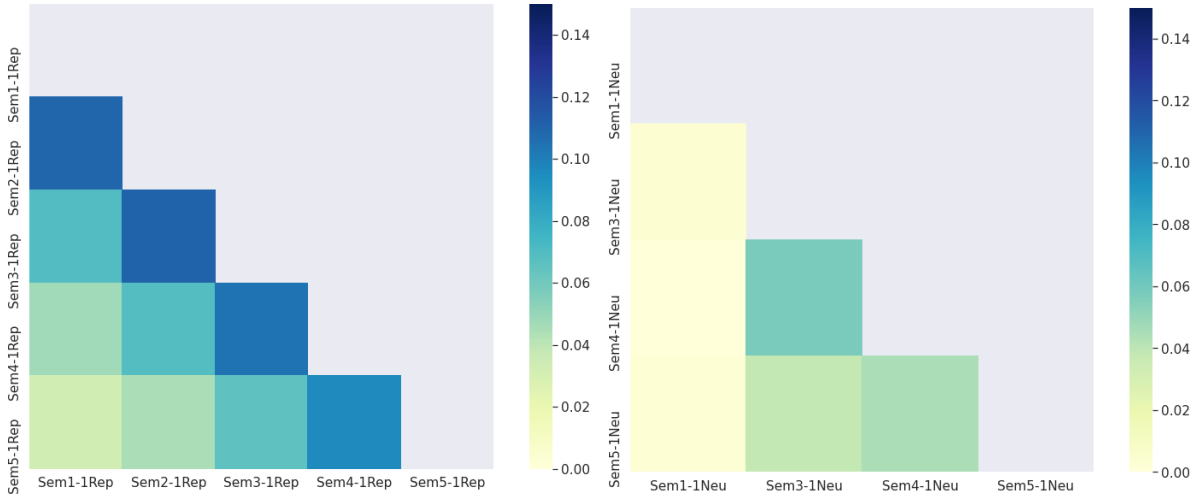
Al contrario, la maggior parte delle EC individuate da Louvain non sono politicamente polarizzate (i.e., neutre). Poichè Louvain tende a massimizzare la modularità di una comunità, possiamo dedurre che le `EC_community` neutre hanno una densità interna più elevata rispetto a quelle polarizzate. A livello locale, sia EVA che Angel non riescono ad ottenere risultati particolarmente interessanti. Nello specifico, notiamo che Angel individua tendenzialmente `simple_community` piccole e con un EC_score piuttosto basso.

Avendo condotto un'analisi su base semestrale, riteniamo interessante studiare la persistenza nel tempo delle `EC_community` estratte dall'algoritmo più performante, Infomap. Per ognuno dei tre livelli di polarizzazione considerati, abbiamo osservato se le EC risultano, nei vari semestri, correlate fra loro in termini di utenti che le compongono. A tal fine distinguiamo le `EC_community` per livello di polarizzazione (i.e., democratico, repubblicano, neutrale). Per aiutare il lettore in questa procedura, tale suddivisione è indicata negli Scatter plot dalle label presenti in ogni quadrante (i.e., *1Rep*, *1Neu*, *1Dem*).

Per misurare la similarità tra ogni EC abbiamo calcolato l'indice di Jaccard tra gli insiemi di utenti che le compongono. Tale indice, dati due insiemi, calcola il rapporto tra la dimensione della loro intersezione e la dimensione della loro unione. Per ogni topic e label presente, mostriamo una *Heatmap* in cui ogni cella rappresenta la similarità, ovvero l'indice di Jaccard, tra le `EC_community` estratte in due diversi semestri.

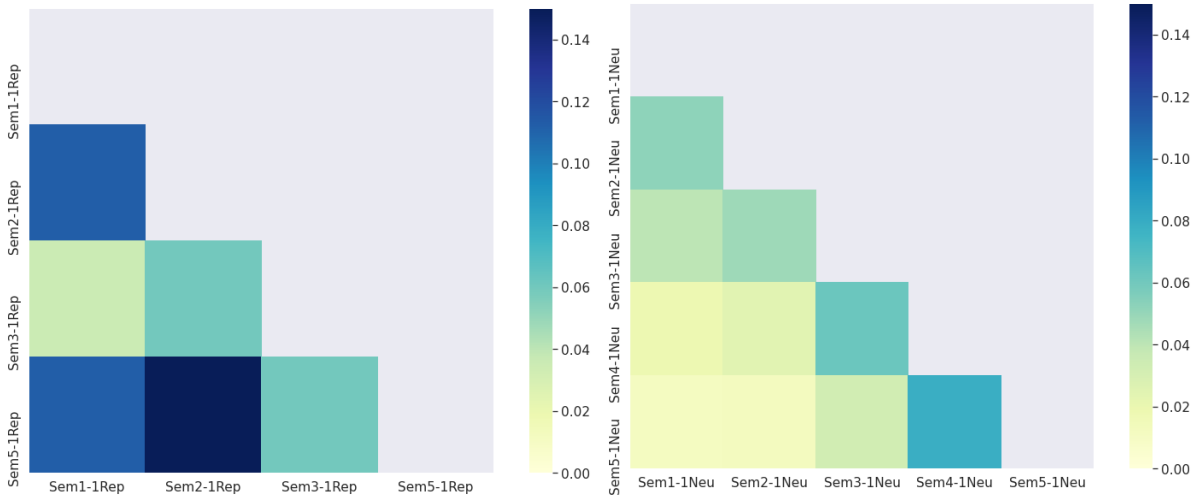
Anche se l'indice di Jaccard raggiunge il valore massimo di 0.15, notiamo che le `EC_community` polarizzate, estratte nei vari semestri, sono piuttosto correlate. La correlazione tende ad accentuarsi tra semestri contigui (Figure 5.9a, 5.11a). Non osserviamo invece una correlazione significativa per le `EC_community` poco polarizzate (Figure 5.9b, 5.10b). Tali risultati sono in linea con il concetto di polarizzazione: è comprensibile infatti che utenti fortemente polarizzati tendano a rimanere ancorati alle proprie ideologie con il passare del tempo; viceversa è più

probabile che utenti poco polarizzati siano più predisposti ad avere un'appartenenza politica meno stabile.



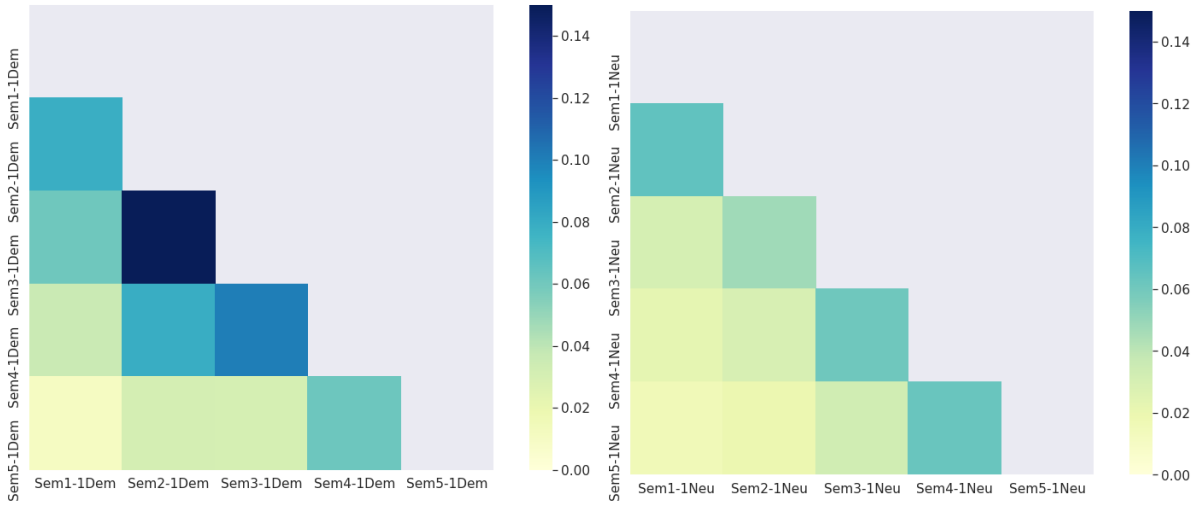
(a) EC_community repubblicane (b) EC_community neutrali

Figura 5.9: *Gun Control* - Similarità tra le EC_community estratte nei diversi semestri in base al livello di polarizzazione.



(a) EC_community repubblicane (b) EC_community neutrali

Figura 5.10: *Minority Discrimination* - Similarità tra le EC_community estratte nei diversi semestri in base al livello di polarizzazione.



(a) EC_community democratiche

(b) EC_community neutrali

Figura 5.11: *Political Sphere* - Similarità tra le EC_community estratte nei diversi semestri in base al livello di polarizzazione.

Capitolo 6

Conclusioni

In questa tesi abbiamo proposto una metodologia per l'identificazione di fenomeni quali polarizzazione politica e Echo chamber, considerati nella loro accezione digitale. Come *chamber* d'analisi abbiamo scelto la piattaforma Reddit con la volontà di contribuire agli studi già esistenti in questo settore. Come *echo* abbiamo deciso di studiare la scena politica Americana in un momento storico altamente polarizzante come i primi due anni e mezzo di presidenza di Donald Trump. In particolare, con il presente progetto di tesi, abbiamo voluto verificare l'esistenza su Reddit di EC di stampo democratico e repubblicano.

La prima fase di questo lavoro è stata dedicata alla definizione di una metodologia per predire la polarizzazione politica di un post condiviso sulla piattaforma. Non avendo a disposizione utenti/post annotati, abbiamo costruito la nostra ground truth estraendo post da Subreddits noti per essere particolarmente polarizzati rispetto alle ideologie democratiche e repubblicane. Per risolvere il task, abbiamo testato parallelamente due diversi approcci:

- *Approccio "Corpus as a proxy":* il problema è stato modellato come un task di Topic Modeling. Dalla ground truth abbiamo estratto, tramite Latent Dirichlet Allocation, keywords (i.e., bigrammi) caratterizzanti il linguaggio dei due partiti e successivamente classificato i post in base alla presenza/assenza di tali features, tramite Support Vector Machine. Il classificatore migliore ha ottenuto un Accuracy finale del 62.3%. Ci aspettavamo risultati non ottimali in quanto i vettori di occorrenze dati in input a SVM sono molto sparsi. Possiamo dedurre che i bigrammi estratti, data la loro bassa frequenza all'interno dei

post, non sono sufficientemente discriminanti e rappresentativi.

- *Approccio Neurale:* il problema è stato modellato come un task di Text Classification. Abbiamo implementato una Rete Neurale composta da a) Word Embedding layer; b) LSTM layer; c) Ouput layer composto da un solo neurone al fine di eseguire predizioni binarie. Il modello, allenato sulla ground truth creata, quantifica il livello di polarizzazione di un post rispetto alle ideologie democratiche e repubblicane. Dopo aver testato varie configurazioni degli iper-parametri, il modello con le migliori performances è stato ottenuto utilizzando 128 unità di LSTM e GloVe Word Embeddings aventi 100 dimensioni. Esso ottiene un Accuracy del 82.9% sul Validation Set e 84.3% sul Test set. Approcci di questo tipo su SNS differenti raggiungono Accuracy comparabili (i.e., 85-87%) nonostante la presenza di una ground truth di utenti annotati. Infine, abbiamo verificato la capacità del modello di generalizzare, validando il modello su post provenienti da Subreddits differenti.

Le performances ottenute dai due approcci risultano nettamente differenti. Possiamo dunque dedurre che, in questo scenario, è preferibile catturare la valenza semantica veicolata da un post, piuttosto che basarsi esclusivamente sulla presenza di determinate keywords. Quindi, dati i risultanti incoraggianti raggiunti dalla Rete Neurale, abbiamo scelto questa metodologia per svolgere gli step successivi. Tuttavia, essendo le Echo Chamber sistemi fortemente polarizzati, abbiamo deciso di classificare i post in tre categorie di polarizzazione a seconda del Prediction Score: 1) polarizzati rispetto alle ideologie democratiche se Prediction Score ≤ 0.3; 2) polarizzati rispetto alle ideologie repubblicane se Prediction Score ≥ 0.7; 3) altrimenti poco polarizzati/neutri.

La seconda fase di questo lavoro è stata dedicata all'effettiva identificazione di Echo Chamber in tre topic relativi alla sfera socio-politica. Inizialmente, abbiamo creato per ogni topic e per ogni semestre Network di interazioni tra utenti annotati in base alla polarizzazione dei loro post. Successivamente, abbiamo testato differenti approcci di Community Discovery (i.e., Infomap, Angel, Louvain, EVA) per estrarre comunità omogenee dai Network creati.
Poiché in letteratura non esiste una definizione formale di Echo Chamber abbiamo definito dei requisiti ideologici e strutturali per stabilire se le comunità estratte possono essere considera-

te EC. Per quanto riguarda i requisiti ideologici, data la definizione di EC, riteniamo che una comunità possa essere considerata tale solo se gli utenti che la compongono hanno ideologie simili. Con questo fine, abbiamo definito una threshold per valutare la coesione ideologica di una comunità a seconda dell'algoritmo di CD utilizzato (i.e., Standard Deviation ≤ 0.25, Purezza ≥ 0.7). Dal punto di vista strutturale, essendo un EC un sistema in cui gli utenti rafforzano le proprie ideologie a vicenda, valutiamo le comunità in termini di densità interna e di link che escono dalla comunità. È stata definita dunque una metrica (i.e., *EC_score*) che stabilisce, in termini di densità e conduttanza, se una comunità è coesa strutturalmente.
L'algoritmo di Community Discovery più performante in questo contesto è risultato essere Infomap. Quest'ultimo riesce infatti a individuare sistemi polarizzati coesi ideologicamente e strutturalmente in ogni topic preso in considerazione. Tuttavia, riteniamo che questo risultato sia strettamente dipendente dal contesto di studio. Risulta quindi utile, in ricerche di questo tipo, testare diversi approcci di CD in modo da selezionare quello che meglio si adatta ai Network definiti.
Riassumendo, i principali contributi apportati in questo lavoro di tesi sono descritti di seguito:

1. Definizione di una metodologia mirata all'identificazione di Echo Chamber su Reddit, facilmente estendibile ad altre piattaforme di Social Networking e ad altri topic.

2. Utilizzo di Reddit, piattaforma mai utilizzata in ricerche relative a fenomeni quali polarizzazione e EC.

3. Implementazione di un modello per predire il grado di polarizzazione politica di un post.

4. Definizione di requisiti ideologici e strutturali che un insieme di utenti connessi deve poter soddisfare per essere considerato un Echo Chamber.

Per quanto concerne i possibili sviluppi futuri, il presente lavoro può essere ampliato e modificato sotto diversi aspetti. Innanzitutto, in questa tesi abbiamo caratterizzato le Echo Chamber estratte da un punto di vista prettamente quantitativo. Sarebbe interessante studiare tali sistemi polarizzati adottando approcci qualitativi, cioè studiando le caratteristiche e le dinamiche degli utenti che la compongono. Ad esempio, potremmo utilizzare modelli di diffusione per

osservare come i membri di un EC reagiscono ad una nuova informazione immessa nel Network.

Un altro aspetto riguarda la metodologia proposta per predire la polarizzazione politica di un post. Dal momento che numerosi post condivisi su Reddit contengono immagini in allegato (e.g., meme), potremmo testare un classificatore che integri al contenuto testuale anche il messaggio veicolato dall'immagine.

Infine, potremmo adattare l'intera metodologia proposta ad altre piattaforme di Social Networking, comparando i risultati ottenuti.

Bibliografia

[1] A. I. Abramowitz and K. L. Saunders. Is polarization a myth? *The Journal of Politics*, 70(2):542–555, 2008.

[2] L. A. Adamic and N. Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43, 2005.

[3] S. E. Asch. Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychological monographs: General and applied*, 70(9):1, 1956.

[4] E. Bakshy, S. Messing, and L. A. Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.

[5] D. Baldassarri and A. Gelman. Partisans without constraint: Political polarization and trends in american public opinion. *American Journal of Sociology*, 114(2):408–446, 2008.

[6] A. Banerjee and R. N. Dave. Validating clusters using the hopkins statistic. In *2004 IEEE International conference on fuzzy systems (IEEE Cat. No. 04CH37542)*, volume 1, pages 149–153. IEEE, 2004.

[7] A.-L. Barabási et al. *Network science*. Cambridge university press, 2016.

[8] L. M. Bartels. Partisanship in the trump era. *The Journal of Politics*, 80(4):1483–1494, 2018.

[9] R. F. Baumeister and M. R. Leary. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin*, 117(3):497, 1995.

[10] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn. The pushshift reddit dataset. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 830–839, 2020.

[11] F. Bélanger and R. E. Crossler. Privacy in the digital age: a review of information privacy research in information systems. *MIS quarterly*, 35(4):1017–1042, 2011.

[12] Y. Bengio, N. Boulanger-Lewandowski, and R. Pascanu. Advances in optimizing recurrent networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8624–8628. IEEE, 2013.

[13] E. Berscheid and H. T. Reis. Attraction and close relationships. *The handbook of social psychology*, 1998.

[14] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[15] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[16] P. Borah, K. Thorson, and H. Hwang. Causes and consequences of selective exposure among political blog readers: The role of hostile media perception in motivated media use and expressive participation. *Journal of Information Technology & Politics*, 12(2):186–199, 2015.

[17] S. Boulianne. Social media use and participation: A meta-analysis of current research. *Information, communication & society*, 18(5):524–538, 2015.

[18] D. M. Boyd and N. B. Ellison. Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, 13(1):210–230, 2007.

[19] J. Brundidge. Encountering "difference" in the contemporary public sphere: The contribution of the internet to the heterogeneity of political discussion networks. *Journal of Communication*, 60(4):680–700, 2010.

[20] B. ChandraShekar and G. Shoba. Classification of documents using kohonen's self-organizing map. *International Journal of Computer Theory and Engineering*, 1(5):610, 2009.

[21] C.-C. Chang, S.-I. Chiu, and K.-W. Hsu. Predicting political affiliation of posts on facebook. In *Proceedings of the 11th International Conference on Ubiquitous Information Management and Communication*, pages 1–8, 2017.

[22] S. Citraro and G. Rossetti. Eva: Attribute-aware network segmentation. In *International Conference on Complex Networks and Their Applications*, pages 141–151. Springer, 2019.

[23] R. L. Claassen and B. Highton. Policy polarization among party elites and the significance of political awareness in the mass public. *Political Research Quarterly*, 62(3):538–551, 2009.

[24] S. Coll. Power, knowledge, and the subjects of privacy: understanding privacy as the ally of surveillance. *Information, Communication & Society*, 17(10):1250–1263, 2014.

[25] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[26] M. Coscia, F. Giannotti, and D. Pedreschi. A classification for community discovery methods in complex networks. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 4(5):512–546, 2011.

[27] L. Dahlberg. Computer-mediated communication and the public sphere: A critical analysis. *Journal of Computer-mediated communication*, 7(1):JCMC714, 2001.

[28] S. Debortoli, O. Müller, I. Junglas, and J. vom Brocke. Text mining for information systems researchers: An annotated topic modeling tutorial. *Communications of the Association for Information Systems*, 39(1):7, 2016.

[29] P. DiMaggio, J. Evans, and B. Bryson. Have american's social attitudes become more polarized? *American journal of Sociology*, 102(3):690–755, 1996.

[30] J. Ding, Y. Liu, L. Zhang, and J. Wang. Modeling the process of event sequence data generated for working condition diagnosis. *Mathematical Problems in Engineering*, 2015, 2015.

[31] E. Dubois and G. Blank. The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, communication & society*, 21(5):729–745, 2018.

[32] L. Festinger. *A theory of cognitive dissonance*, volume 2. Stanford university press, 1957.

[33] M. Fiorina. The 2016 presiedential election–an abundance of controversies, 2017.

[34] M. P. Fiorina and S. J. Abrams. Political polarization in the american public. *Annu. Rev. Polit. Sci.*, 11:563–588, 2008.

[35] P. Fischer, E. Jonas, D. Frey, and S. Schulz-Hardt. Selective exposure to information: The impact of information limits. *European Journal of Social Psychology*, 35(4):469–492, 2005.

[36] S. T. Fiske. *Social beings: Core motives in social psychology*. John Wiley & Sons, 2018.

[37] G. W. Flake, S. Lawrence, C. L. Giles, and F. M. Coetzee. Self-organization and identification of web communities. *Computer*, 35(3):66–70, 2002.

[38] G. Forman and E. Kirshenbaum. Extremely fast text feature extraction for classification and indexing. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 1221–1230, 2008.

[39] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of the 2018 World Wide Web Conference*, pages 913–922, 2018.

[40] K. Garimella et al. *Polarization on social media.* PhD thesis, Aalto University, 2018.

[41] K. Garimella, G. D. F. Morales, A. Gionis, and M. Mathioudakis. Quantifying controversy on social media. *ACM Transactions on Social Computing*, 1(1):1–27, 2018.

[42] M. Gentzkow, J. M. Shapiro, and M. Taddy. Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica*, 87(4):1307–1340, 2019.

[43] F. A. Gers and J. Schmidhuber. Recurrent nets that time and count. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, volume 3, pages 189–194. IEEE, 2000.

[44] E. Gilbert, T. Bergstrom, and K. Karahalios. Blogs are echo chambers: Blogs are echo chambers. In *2009 42nd Hawaii International Conference on System Sciences*, pages 1–10. IEEE, 2009.

[45] M. Girvan and M. E. Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[46] E. W. Groenendyk and A. J. Banks. Emotional rescue: How affect helps partisans overcome collective action problems. *Political Psychology*, 35(3):359–378, 2014.

[47] R. Guimera and L. A. N. Amaral. Functional cartography of complex metabolic networks. *nature*, 433(7028):895–900, 2005.

[48] V. Gupta, G. S. Lehal, et al. A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence*, 1(1):60–76, 2009.

[49] M. A. Halliday. Jr firth: Selected papers of jr firth, 1952–59. edited by fr palmer.(longmans' linguistics library.) x, 209 pp. london: Longmans, green and co. ltd., 1968. 30s. *Bulletin of the School of Oriental and African Studies*, 34(3):664–667, 1971.

[50] M. Hearst. What is text mining. *SIMS, UC Berkeley*, 5, 2003.

[51] J. Hedman, N. Srinivasan, and R. Lindgren. Digital traces of information systems: Sociomateriality made researchable. *Proceedings of the 2013 International Conference on Information System (ICIS 2013)*, 2013.

[52] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[53] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[54] R. Irfan, C. K. King, D. Grages, S. Ewen, S. U. Khan, S. A. Madani, J. Kolodziej, L. Wang, D. Chen, A. Rayes, et al. A survey on text mining in social networks. *The Knowledge Engineering Review*, 30(2):157–170, 2015.

[55] D. J. Isenberg. Group polarization: A critical review and meta-analysis. *Journal of personality and social psychology*, 50(6):1141, 1986.

[56] K. H. Jamieson and J. N. Cappella. *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press, 2008.

[57] S. Jean Tsang. Cognitive discrepancy, dissonance, and selective exposure. *Media Psychology*, 22(3):394–417, 2019.

[58] M. Jeong, H. Zo, C. H. Lee, and Y. Ceran. Feeling displeasure from online social media postings: A study using cognitive dissonance theory. *Computers in Human Behavior*, 97:231–240, 2019.

[59] T. Joachims. Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning*, pages 137–142. Springer, 1998.

[60] T. Joachims. Transductive inference for text classification using support vector machines. In *Icml*, volume 99, pages 200–209, 1999.

[61] E. Jonas, S. Schulz-Hardt, D. Frey, and N. Thelen. Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure to information. *Journal of personality and social psychology*, 80(4):557, 2001.

[62] M. I. Jordan. Serial order: A parallel distributed processing approach. In *Advances in psychology*, volume 121, pages 471–495. Elsevier, 1997.

[63] Y. Kim, S.-H. Hsu, and H. G. de Zúñiga. Influence of social media use on discussion network heterogeneity and civic engagement: The moderating role of personality traits. *Journal of communication*, 63(3):498–516, 2013.

[64] J. T. Klapper. *The effects of mass communication.* Free press, 1960.

[65] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown. Text classification algorithms: A survey. *Information*, 10(4):150, 2019.

[66] S. Kull, C. Ramsay, and E. Lewis. Misperceptions, the media, and the iraq war. *Political science quarterly*, 118(4):569–598, 2003.

[67] B. S. Kumar and V. Ravi. A survey of the applications of text mining in financial domain. *Knowledge-Based Systems*, 114:128–147, 2016.

[68] Z. Kunda. The case for motivated reasoning. *Psychological bulletin*, 108(3):480, 1990.

[69] B. Liu and L. Zhang. A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463. Springer, 2012.

[70] T. E. Mann and N. J. Ornstein. *It's Even Worse Than It Looks: How the American Constitutional System Collided with the New Politics of Extremism.* Hachette UK, 2016.

[71] A. H. Maslow. A theory of human motivation. *Psychological review*, 50(4):370, 1943.

[72] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *preprint arXiv:1301.3781*, 2013.

[73] S. M. Monnat. Deaths of despair and support for trump in the 2016 presidential election. *Pennsylvania State University Department of Agricultural Economics Research Brief*, 5, 2016.

[74] Morini and al. Capturing political polarization of reddit submissions in the trump era. In *SEBD*, 2020.

[75] D. G. Myers. Discussion-induced attitude polarization. *Human Relations*, 28(8):699–714, 1975.

[76] D. G. Myers and H. Lamm. The polarizing effect of group discussion: the discovery that discussion tends to enhance the average prediscussion tendency has stimulated new insights about the nature of group influence. *American Scientist*, 63(3):297–303, 1975.

[77] P. S. Negi, M. Rauthan, and H. Dhami. Language model for information retrieval. *International Journal of Computer Applications*, 12(7):13–17, 2010.

[78] M. E. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[79] R. E. Pahl. *Divisions of labour*. Blackwell Oxford, 1984.

[80] E. Pariser. *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.

[81] S. Parthasarathy, Y. Ruan, and V. Satuluri. Community discovery in social networks: Applications, methods and emerging trends. In *Social network data analytics*, pages 79–113. Springer, 2011.

[82] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *Empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[83] M. Prior. *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press, 2007.

[84] A. Rao and N. Spasojevic. Actionable and political text classification using word embeddings and lstm. *arXiv preprint arXiv:1607.02501*, 2016.

[85] J. M. Reyes. Social network, polarizzazione e democrazia: dall'entusiasmo al disincanto., 2018.

[86] M. Röder, A. Both, and A. Hinneburg. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pages 399–408, 2015.

[87] G. Rossetti. Angel: efficient, and effective, node-centric community discovery in static and dynamic networks. *Applied Network Science*, 5(1):1–23, 2020.

[88] M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.

[89] F. Sebastiani. Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1):1–47, 2002.

[90] H. J. Smith. Ethics and information systems: Resolving the quandaries. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 33(3):8–22, 2002.

[91] L. Sorensen. User managed trust in social networking-comparing facebook, myspace and linkedin. In *2009 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology*, pages 427–431. IEEE, 2009.

[92] X. Sun, X. Liu, B. Li, Y. Duan, H. Yang, and J. Hu. Exploring topic models in software engineering data analysis: A survey. In *2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pages 357–362. IEEE, 2016.

[93] C. R. Sunstein. *Republic. com.* Princeton university press, 2001.

[94] C. R. Sunstein. The law of group polarization. *Journal of Political Philosophy*, 2002.

[95] C. R. Sunstein. *Republic.Com 2.0.* Princeton University Press, USA, 2007.

[96] K. Y. Tam and S. Y. Ho. Understanding the impact of web personalization on user information processing and decision outcomes. *MIS quarterly*, pages 865–890, 2006.

[97] J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher, and M. S. Wetherell. *Rediscovering the social group: A self-categorization theory.* Basil Blackwell, 1987.

[98] C. Vaccari, A. Valeriani, P. Barberá, R. Bonneau, J. T. Jost, J. Nagler, and J. A. Tucker. Political expression and action on social media: Exploring the relationship between lower-and higher-threshold political activities among twitter users in italy. *Journal of Computer-Mediated Communication*, 20(2):221–239, 2015.

[99] P. Van Aelst, J. Strömbäck, T. Aalberg, F. Esser, C. De Vreese, J. Matthes, D. Hopmann, S. Salgado, N. Hubé, A. Stępińska, et al. Political communication in a high-choice media environment: a challenge for democracy? *Annals of the International Communication Association*, 41(1):3–27, 2017.

[100] S. Yardi and D. Boyd. Dynamic debates: An analysis of group polarization over time on twitter. *Bulletin of science, technology & society*, 30(5):316–327, 2010.