

**Università di Roma Tor Vergata**  
**Corso di Laurea magistrale in Ingegneria Informatica**  
**Dipartimento di Ingegneria Informazione**



**Analisi della polarizzazione di Endorsement**  
**Graph, attraverso sentiment analysis**

**Relatore:**

**Giuseppe F. Italiano**

**Correlatore:**

**Ing. Nikos Parotsidis**

**Candidato:**

**Alessandro Valenti**

**matricola 0228709**

**Anno Accademico 2016-2017**



*A qualcuno...*



# Sommario

Il sommario deve contenere 3 o 4 frasi tratte dall'introduzione di cui la prima inquadra l'area dove si svolge il lavoro (eventualmente la seconda inquadra la sottoarea più specifica del lavoro), la seconda o la terza frase dovrebbe iniziare con le parole "Lo scopo della tesi è ..." e infine la terza o quarta frase riassume brevemente l'attività svolta, i risultati ottenuti ed eventuali valutazioni di questi.

NB: se il relatore effettivo è interno al Politecnico di Milano nel frontesizio si scrive Relatore, se vi è la collaborazione di un altro studioso lo si riporta come Correlatore come sopra. Nel caso il relatore effettivo sia esterno si scrive Relatore esterno e poi bisogna inserire anche il Relatore interno. Nel caso il relatore sia un ricercatore allora il suo Nome COGNOME dovrà essere preceduto da Ing. oppure Dott., a seconda dei casi.



# Ringraziamenti

Ringrazio .....





# Capitolo 1

## Introduzione

*“Terence: Rotta a nord con circospezione*

*Bud: Ehi, gli ordini li do io qui!*

*Terence: Ok, comante*

*Bud: Rotta a nord*

*Terence: Soltanto?*

*Bud: Con circospezione!”*

Chi Trova un Amico Trova un Tesoro

La polarizzazione é un utilissimo strumento per lo studio e l’analisi delle rete sociali sulle opinioni all’interno di differenti aree di ricerca. Generalmente la polarizzazione può essere tranquillamente applicata all’interno di contesti politici, sociali e culturali permettendo di comprendere al meglio quali siano le vere opinioni delle persone riguardo tali argomenti. Una generica definizione della polarizzazione é la seguente:

Divisione in due gruppi fortemente contrastanti per una serie di opinioni o credenze.

Questo processo di analisi può assumere diversi significati a seconda dello scenario studiato.

- *Polarizzazione Politica*: divergenza di opinione su estremi ideologici.
- *Polarizzazione Sociale*: differenza di opinione all'interno delle società che possono essere scaturite da disuguaglianze sociali ed economiche.

La polarizzazione può comportare diversi cambiamenti sullo scenario in questione, in quanto mette in luce come la formazione di due grandi gruppi non consenta una diffusione democratica delle opinioni. A tal proposito è interessante notare come la divisione in queste due grandi partizioni generi alcune problematiche quali:

- La frammentazione della rete stessa.
- L'isolamento delle opinioni.

In conclusione potremmo definire la polarizzazione come un processo sociale per cui gli utenti che vi partecipano vengono divisi in due grandi sottogruppi aventi visioni, punti di vista ed opinioni differenti del problema in gestione, con alcuni individui che rimangono neutrali tra i due grandi gruppi.

Il problema che si può facilmente evincere risulta essere la formazione di due comunità isolate che non comunicano tra loro, ciò comporta un problema di isolamento delle opinioni, cioè un utente che appartiene a quel gruppo difficilmente potrà ricevere informazioni o aderire alle idee del gruppo opposto. Otteniamo un problema che genera la formazione degli *Echo-Chambers*. Si definiscono *Echo-Chambers* come:

Una situazione in cui le informazioni, le idee e le credenze vengono rinforzate e amplificate perché espresse all'interno dello stesso ambiente, rimanendo isolato.

Un'altro problema che può formare una forte polarizzazione delle opinioni e delle informazioni sono i Filter Bubble ovvero:

Uno stato di un isolamento intellettuale che può essere ottenuto a partire da risultati di ricerche su siti che registrano la storia del comportamento dell'utente. Questi siti sono in grado di utilizzare informazioni sull'utente per scegliere selettivamente tra tutte le risposte quelle che vorrà vedere l'utente stesso. L'effetto è di isolare l'utente da informazioni che sono in contrasto con il suo punto di vista, effettivamente isolandolo nella sua bolla culturale o ideologica.

Come precedentemente anticipato la polarizzazione è uno strumento che può essere facilmente utilizzato per individuare tutte queste problematiche all'interno dei moderni Social Network come Facebook e Twitter e molti altri. Questo perché sempre più si stanno facendo largo nella vita di tutti i giorni e le problematiche relativi a contesti sociali, culturali e politici vengono sempre più affrontati all'interno di queste piattaforme, in cui gli utenti si sentono sempre più liberi di poter esprimere le proprie opinioni. Il problema è che non è sempre possibile uscire dalle Filter Bubble perché gli stessi social network tendono a indirizzare l'utente a visualizzare informazioni che potrebbero interessarli senza fargli confrontare con opinioni divergenti. Alla luce di questo grande problema il calcolo di una polarizzazione può consentire agli amministratori dei social network di individuare i topic più polarizzati e consentire una diffusione democratica delle informazioni.

L'obiettivo della mia tesi consiste nell'utilizzare la polarizzazione per poter individuare quegli argomenti fortemente polarizzati e comprendere come tali informazioni vengono prodotte all'interno della rete sociale. Lo sviluppo di questo strumento è stato effettuato sfruttando due algoritmi, presentati nei seguenti paper:

- *Measuring Political Polarization: Twitter shows the two sides of Venezuela*: Studia la diffusione delle informazioni all'interno di un *endorsement graph* collezionando i dati relativi alle elezioni in Venezuela all'interno del *social network Twitter*. Viene effettuato uno studio della polarizzazione all'interno di un contesto politico sfruttando la diffusione delle informazioni, l'*endorsement graph* viene costruito partendo da un nodo che scrive un Tweet esprimendo la propria opinioni, formando un nodo, mentre eventuali follower di quell'utente che retwettano tale notizia sono nuovi nodi all'interno del grafo con archi uscenti verso il nodo che hanno retwettato. In questo modo viene generato un grafo basato sul retweet. Una volta generato il grafo vengono catalogati i nodi in due categorie:

- *Elite*: l'utente che ha tweekato un'opinione.
- *Listener*: l'utente che ha retwettato il tweet di uno o più nodi *Elite*

Partendo da queste categorie viene calcolata la polarizzazione sfruttando il grado di ogni nodo. (Per una più dettagliata spiegazione si rimanda al Capitolo??)

- *Reducing Controversy by Connecting Opposing Views*: Identifica la polarizzazione sfruttando la struttura del grafo. Il grafo viene generato utilizzando la medesima tecnica precedentemente illustrata, così come il *social network* di riferimento. La differenza principale è che non vengono catalogati i nodi in due gruppi in base al loro comportamento nel grafo. Adotta la tecnica dei *Random Walk* sfruttando la probabilità di retweet per ottenere il valore assoluto della polarizzazione per ogni nodo appartenente all' *endorsment graph*.

## 1.1 Inquadramento generale

La prima parte contiene una frase che spiega l'area generale dove si svolge il lavoro; una che spiega la sottoarea più specifica dove si svolge il lavoro e la terza, che dovrebbe cominciare con le seguenti parole “lo scopo della tesi è ...”, illustra l'obiettivo del lavoro. Poi vi devono essere una o due frasi che contengano una breve spiegazione di cosa e come è stato fatto, delle attività sperimentali, dei risultati ottenuti con una valutazione e degli sviluppi futuri. La prima parte deve essere circa una facciata e mezza o due

## 1.2 Breve descrizione del lavoro

La seconda parte deve essere una esplosione della prima e deve quindi mostrare in maniera più esplicita l'area dove si svolge il lavoro, le fonti bibliografiche più importanti su cui si fonda il lavoro in maniera sintetica (una pagina) evidenziando i lavori in letteratura che presentano attinenza con il lavoro affrontato in modo da mostrare da dove e perché è sorta la tematica di studio. Poi si mostrano esplicitamente le realizzazioni, le direttive future di ricerca, quali sono i problemi aperti e quali quelli affrontati e si ripete lo scopo della tesi. Questa parte deve essere piena (ma non grondante come la sezione due) di citazioni bibliografiche e deve essere lunga circa 4 facciate.

## 1.3 Struttura della tesi

La terza parte contiene la descrizione della struttura della tesi ed è organizzata nel modo seguente. “La tesi è strutturata nel modo seguente.

Nella sezione due si mostra ...

Nella sez. tre si illustra ...

Nella sez. quattro si descrive ...

Nelle conclusioni si riassumono gli scopi, le valutazioni di questi e le prospettive future ...

Nell'appendice A si riporta ... (Dopo ogni sezione o appendice ci vuole un punto).”

I titoli delle sezioni da 2 a M-1 sono indicativi, ma bisogna cercare di mantenere un significato equipollente nel caso si vogliano cambiare. Queste sezioni possono contenere eventuali sottosezioni.

# Capitolo 2

## Stato dell'arte

### 2.1 Stato dell'arte

All'interno delle reti sociali sta sempre più prendendo piede il problema della polarizzazione delle opinioni. Nel linguaggio comune il confronto tra individui ha sempre generato una forte controversia nelle opinioni oppure una situazione di neutralità nelle opinioni oppure una visione comune nelle opinioni. I social network hanno permesso all'utente di poter diffondere attraverso post, messaggi o espressioni audio video le proprie opinioni e pensieri all'interno di una comunità sociale. A tal proposito per favorire la diffusione delle diverse correnti di pensiero i social network stanno sempre più sviluppando algoritmi per permettere di identificare le comunità isolate che condividono un unico punto di vista di un problema. La polarizzazione è un algoritmo matematico che applicato all'interno delle reti sociali permette di capire quanto un utente che accede per la prima volta all'interno di una rete sociale venga influenzato dagli altri utenti e quanto una news o un giudizio si propaga all'interno di una rete sociale. Prima di poter illustrare questo algoritmo con le relative problematiche verrà illustrata una definizione di rete sociale.

**Rete Sociale** Una rete sociale consiste in un qualsiasi gruppo di individui connessi tra loro da diversi legami sociali. Per gli esseri umani i legami vanno dalla conoscenza casuale, ai rapporti di lavoro, ai vincoli familiari. Le reti sociali sono spesso usate come base di studi interculturali in sociologia, in antropologia, in etologia.

L'analisi delle reti sociali, ovvero la mappatura e la misurazione delle reti sociali, può essere condotta con un formalismo matematico usando la teoria dei grafi. In generale, il corpus teorico ed i modelli usati per lo studio delle reti sociali sono compresi nella cosiddetta social network analysis.

La ricerca condotta nell'ambito di diversi approcci disciplinari ha evidenziato come le reti sociali operino a più livelli e svolgano un ruolo cruciale nel determinare le modalità di risoluzione di problemi e i sistemi di gestione delle organizzazioni, nonché le possibilità dei singoli individui di raggiungere i propri obiettivi.

**Le reti** La diffusione del web e del termine social network ha creato negli ultimi anni alcune ambiguità di significato. La rete sociale è infatti storicamente, in primo luogo, una rete fisica.

Rete sociale è, ad esempio, una comunità di lavoratori, che si incontra nei relativi circoli dopolavoristici e che costituisce una delle associazioni di promozione sociale. Esempi di reti sociali sono inoltre le comunità di sportivi, attivi o sostenitori di eventi, le comunità unite da problematiche strettamente lavorative e di tutela sindacale del diritto nel lavoro, le confraternite e in generale le comunità basate sulla pratica comune di una religione e il ritrovo in chiese, templi, moschee, sinagoghe e altri luoghi di culto.

Una rete sociale si può inoltre basare su di un comune approccio educativo come nello scautismo, o nel pionierismo, di visione sociale, come nelle reti



segrete della carboneria e della massoneria.



## Capitolo 3

# Impostazione del problema di ricerca

*“Bud: Apri!*

*Cattivo: Perch   $\frac{1}{2}$ , altrimenti vi arrabbiate?*

*Bud e Terence: Siamo gi   $\frac{1}{2}$  arrabbiati!”*

Altrimenti ci arrabbiamo

In questa sezione si deve descrivere l’obiettivo della ricerca, le problematiche affrontate ed eventuali definizioni preliminari nel caso la tesi sia di carattere teorico.



## Capitolo 4

# Progetto logico della soluzione del problema

*“Bud: No, calma, calma, stiamo calmi, noi siamo su un’isola deserta, e per il momento non t’ammazzo perché mi potresti servire come cibo ...”*

Chi trova un amico trova un tesoro

In questa sezione si spiega come è stato affrontato il problema concettualmente, la soluzione logica che ne è seguita senza la documentazione.



# Capitolo 5

## Architettura del sistema

*“Terence: Ma scusa di che ti preoccupi, i piedipiatti hanno altro a cui pensare, in questo momento stanno cercando due cadaveri scomparsi*

*Bud: Se non spegni quella sirena uno di quei due cadaveri scomparsi lo trovano di sicuro!”*

Nati con la camicia

Si mostra il progetto dell’architettura del sistema con i vari moduli.





## Capitolo 6

# Realizzazioni sperimentali e valutazione

*“Bambino: Questo  $\dot{\varphi}_{\frac{1}{2}}$  l’ultimo avviso per voi e i vostri rubagalline*

*Il pistolero si alza: Che avete detto?*

*Bambino: RUBAGALLINE*

*Il pistolero si risiede: Aaah.”*

Lo chiamavano Trinità ...

Si mostra il progetto dal punto di vista sperimentale, le cose materialmente realizzate. In questa sezione si mostrano le attività sperimentali svolte, si illustra il funzionamento del sistema (a grandi linee) e si spiegano i risultati ottenuti con la loro valutazione critica. Bisogna introdurre dati sulla complessità degli algoritmi e valutare l’efficienza del sistema.



## Capitolo 7

# Direzioni future di ricerca e conclusioni

*“Terence: Mi fai un gelato anche a me? Lo vorrei di pistacchio.*

*Bud: Non ce l’ho il pistacchio. C’ho la vaniglia, cioccolato, fragola, limone e caffè.*

*Terence: Ah bene. Allora fammi un cono di vaniglia e di pistacchio.*

*Bud: No, non ce l’ho il pistacchio. C’ho la vaniglia, cioccolato, fragola, limone e caffè.*

*Terence: Ah, va bene. Allora vediamo un po’, fammelo al cioccolato, tutto coperto di pistacchio.*

*Bud: Ehi, macchiò  $\frac{1}{2}$  sei sordo? Ti ho detto che il pistacchio non ce l’ho!*

*Terence: Ok ok, non c’è  $\frac{1}{2}$  bisogno che t’arrabbi, no? Insomma, di che ce l’hai?*

*Bud: Ce l’ho di vaniglia, cioccolato, fragola, limone e caffè!*

*Terence: Ah, ho capito. Allora fammene uno misto: mettici la fragola, il cioccolato, la vaniglia, il limone e il caffè. Charlie, mi raccomando il pistacchio, eh.”*

Pari e dispari

Si mostrano le prospettive future di ricerca nell’area dove si è svolto il lavo-

ro. Talvolta questa sezione può essere l'ultima sottosezione della precedente. Nelle conclusioni si deve richiamare l'area, lo scopo della tesi, cosa è stato fatto, come si valuta quello che si è fatto e si enfatizzano le prospettive future per mostrare come andare avanti nell'area di studio.

# Appendice A

## Documentazione del progetto logico

Documentazione del progetto logico dove si documenta il progetto logico del sistema e se è il caso si mostra la progettazione in grande del SW e dell'HW. Quest'appendice mostra l'architettura logica implementativa (nella Sezione 4 c'era la descrizione, qui ci vanno gli schemi a blocchi e i diagrammi).



# Appendice B

## Documentazione della programmazione

Documentazione della programmazione in piccolo dove si mostra la struttura ed eventualmente l'albero di Jackson.





# Appendice C

## Listato

Il listato (o solo parti rilevanti di questo, se risulta particolarmente esteso)  
con l'autodocumentazione relativa.



# Appendice D

## Il manuale utente

Manuale utente per l'utilizzo del sistema



# Appendice E

## Esempio di impiego

Un esempio di impiego del sistema realizzato.



# Appendice F

## Datasheet

Eventuali Datasheet di riferimento.