Dipartimento di Scienze Politiche

Cattedra di Metodi quantitativi per la valutazione delle politiche pubbliche

ANALISI STATISTICA DEGLI INFLUENCER ELETTORALI SUI SOCIAL NETWORK

RELATORE
Professoressa
Livia de Giovanni

CANDIDATO
Gabriele Donnini
Matr. 626982

CORRELATORE
Professor
Lorenzo De Sio

ANNO ACCADEMICO: 2016/2017

Sommario

Ringraziamenti

Desidero ringraziare la Prof.ssa Livia De Giovanni, per avermi seguito durante la stesura di questo elaborato, ma soprattutto per avermi fatto innamorare della sua materia.

Un ringraziamento è rivolto anche alla Prof.ssa Emiliana De Blasio per il supporto offerto.

Introduzione

Il mondo sta cambiando rapidamente. Negli ultimi anni abbiamo assistito a come internet e i social network siano diventati una componente sempre più importante all'interno della nostra esistenza. Dispositivi sempre più semplici da usare, aumento della copertura della rete e costi sempre più accessibili hanno aumentato esponenzialmente il numero di utenti nel giro di pochi anni.

Il presente lavoro ha lo scopo di analizzare le caratteristiche e il ruolo degli utenti dei social network all'interno di una campagna elettorale. Utilizzeremo quindi strumenti presi dal marketing, dalla sociologia e dalla statistica.

I politici moderni hanno molto in comune con i marketer: per sopravvivere in un ambiente così ostile e competitivo entrambi devono concentrarsi sullo storytelling e sulla costruzione di un brand. La politica è solamente una forma più elevata di marketing dove invece di vendere un prodotto si cerca di vendere la propria personalità.

Iniziamo quindi il primo capitolo analizzando la forma di marketing più antica e potente: il passaparola (Word of Mouth), considerando come il suo potenziale sia cresciuto esponenzialmente con la diffusione della rete e dei social network. Analizziamo poi come gli utenti scelgano a quali messaggi esporsi e a chi dare ascolto. Il passo successivo è una breve analisi della campagna elettorale statunitense del 2016 e il ruolo che i social network e i singoli individui hanno avuto. Infine analizziamo il rapporto che invece gli italiani hanno con i social network in generale: quali usano, quanto li usano , come li usano, chi li usa.

Nel secondo capitolo forniamo gli strumenti statistici che abbiamo utilizzato per condurre la nostra analisi. Abbiamo utilizzato la cluster analysis, utilizzando prima metodi gerarchici e poi non gerarchici. Questa metodologia applicata ad una popolazione di unità statistiche permette di suddividerle in gruppi secondo un criterio di similarità rispetto ad un insieme di variabili. Ottenuti i cluster possiamo tracciarne un profilo che esprima la posizione complessiva del gruppo rispetto alle variabili considerate.

Nel terzo capitolo mostriamo come abbiamo utilizzato la cluster analysis per analizzare gli utenti che su Twitter hanno preso parte alla campagna elettorale relativa al terzo referendum costituzionale nella storia della Repubblica Italiana, che ha avuto luogo il 4 dicembre 2016. La maggioranza dei votanti respinse il testo di legge costituzionale della cosiddetta riforma Renzi-Boschi, approvato in via definitiva dalla Camera il 12 aprile 2016 e recante modifiche alla parte seconda della Costituzione. Il campione è composto da 97 tweet riconducibili a 90 utenti. L'arco temporale della raccolta dati va dal 29 al 5 dicembre. Nel periodo considerato sono stati scaricati i tweet contenenti le due keyword "referendum" e "costituzionale". Dai dati ottenuti sono stati selezionati i tweet contenenti hashtag caratterizzati in senso "partisan" ("iovotosi, iovotono, bastaunsi, iodicono). Il nostro intento è suddividere questi potenziali influencer e brand advocate in gruppi il più possibile omogenei al loro interno. Le variabili considerate sono per ciascun utente sono: produttività (media), numero

(medio) di follower, numero di amici, se il tweet era un retweet o meno, il giorno di generazione del tweet, il fatto che fosse favorevole o contrario alla riforma. In base agli strumenti analizzati nei capitoli precedenti siamo stati in grado di identificare le caratteristiche del brand advocate politico a cui bisogna mirare su Twitter, ossia la persona “comune” il cui messaggio è in grado di mobilitare amici e parenti nel mondo reale. Tali caratteristiche sono state quindi usate per interpretare i risultati della cluster analysis effettuata ed è stato identificato il gruppo di utenti che potrebbe aver avuto un’influenza maggiore. Tale gruppo al suo interno ha sostenuto in larga maggioranza il NO, la scelta che ha poi effettivamente vinto.

Nell’appendice sarà descritto il software R, specificando le istruzioni adoperate e i pacchetti utilizzati.

Capitolo 1: Influenzare per governare: la politica e i social network

1.1 La Word of Mouth

Il passaparola (Word of mouth o WOM) "viene spesso riferito al consiglio disinteressato che viene offerto da un consumatore a un altro in merito a un certo prodotto o servizio. Nasce da uno scambio informale di opinioni ed informazioni tra interlocutori che, in linea di principio, non sono mossi da interessi di natura commerciale nel raccomandare un particolare prodotto, trattandosi per lo più di consumatori che, dopo averlo provato ed esserne rimasti soddisfatti, decidono di consigliarlo ai propri conoscenti"[1].

Non sempre però il passaparola scaturisce in modo spontaneo tra i consumatori. In molti casi esso va costruito, è necessario cioè uno stimolo da parte dell'impresa. Le aziende, infatti, per accrescere la propria notorietà e reputazione si avvalgono di apposite campagne di comunicazione che incoraggiano i consumatori a parlare di un particolare prodotto o servizio ed agevolano lo scambio di informazioni attorno ad esso. In questo modo si favorisce una rapida diffusione di informazioni commerciali tramite le reti sociali dei consumatori stessi. Si parla a tal proposito di marketing del passaparola (Word of Mouth Marketing o WOMM), che può essere definito come "uno sforzo compiuto da un'organizzazione per influenzare il modo in cui i consumatori creano e/o distribuiscono le informazioni rilevanti dal punto di vista del marketing ad altri consumatori" (fonte: WOMMA, Word of Mouth Marketing Association).

1.1.1 Il potenziale dei social network

Tutti ci siamo resi conto di quanto i social network abbiano rivoluzionato il nostro modo di vivere e di pensare. Quello che prima era un *mondo piccolo* ora è diventato un *mondo piccolissimo*[2]. Mi sto riferendo a quello che viene chiamato Small World Phenomenon. Nel 1929 lo scrittore Frigyes Karinthy descrisse nel suo racconto *Catene* un concetto del tutto nuovo, quello di intermediario. Secondo lui gli intermediari fra una persona e qualsiasi altra persona al mondo erano al massimo cinque: "L'operaio conosce il capo officina che conosce mister Ford in persona, il quale ha buoni rapporti con il direttore generale dell'impero editoriale Hearst che ha avuto modo di conoscere il signor Pasztor che è un mio ottimo amico". Nel 1967 il ricercatore di Harvard Stanley Milgram riprende questa riflessione e la sviluppa ulteriormente elaborando quella che verrà poi chiamata "teoria del mondo piccolo". Selezionò 160 persone risiedenti in Nebraska e chiese a ciascuno di loro di inviare un pacco a un estraneo risiedente in Massachusetts (2680 km di distanza). Ogni mittente conosceva nome, mestiere e zona di residenza del destinatario, senza però conoscerne l'indirizzo. L'esperimento consisteva quindi nel chiedere a ciascun partecipante di individuare una strategia per far recapitare il pacchetto attraverso una serie di passaggi, inviandolo dapprima a una persona conosciuta e

[1] *http://www.glossariomarketing.it/significato/word-of-mouth/*
[2] *http://www.unipd.it/ilbo/content/il-mondo-%E2%80%9Cpiccolissimo%E2%80%9D-dei-social-network*

facendolo arrivare a destinazione ricorrendo al minor numero possibile di intermediari. Questa catena mostrò risultati sorprendenti: il pacco giungeva a destinazione dopo soli cinque o, al massimo, sette passaggi. Nel 2001 Duncan Watts ripropose l'esperimento utilizzando le potenzialità offerte da internet. Utilizzò una email al posto del pacchetto e coinvolse un campione di 48.000 provenienti da 157 stati. Lo scopo era raggiungere 19 destinatari. Il risultato fu simile a quello di Milgram: la media degli intermediari risultò sei. Le cose cambiano con l'avvento dei social network: nel 2011 il Laboratorio di algoritmica per il web dell'università di Milano utilizzando un campione di 721 milioni di utenti Facebook è riuscito a ridurre ulteriormente il numero dei passaggi, arrivando a 3,74 , i quali arrivano a 3 se ci si trova nella stessa nazione. Uno studio del 2013 della National Chiao Tung university of Taiwan ha scoperto addirittura che basterebbero in media due intermediari e tre interazioni.[3]

1.1.2 L'e-WOM

Word of Mouth marketing e campaigning è un termine che copre un ampio spettro di canali e strategie. La word of Mouth Marketing Association ha pubblicato le seguenti definizioni (che torneranno molto utili in seguito).

- Buzz Marketing: consiste nell'utilizzo di personaggi dell'intrattenimento o del mondo delle notizie, personaggi in ogni caso di alto profilo. Questi servono a far parlare la gente del tuo brand
- Viral Marketing: creare messaggi divertenti o informativi che sono strutturati per essere condivisi in maniera esponenziale (oggi tramite i social network).
- Community marketing: formazione o supporto di comunità di nicchia che molto probabilmente condivideranno interesse per il tuo brand provvedendo strumenti, contenuti e informazioni a supporto di tali community.
- Grassroot Marketing: organizzare e motivare volontari in modo che questi si impegnino a divulgare il messaggio ad altre persone o all'intera comunità.
- Evangelist Marketing: coltivare evangelisti, advocates o volontari che sono incoraggiati a prendere un ruolo di leadership nel divulgare attivamente il tuo messaggio per te.
- Product Seeding: piazzare il giusto prodotto nelle giuste mani al momento giusto, provvedendo informazioni o campioni ad individui influenti (influencer).
- Influencer Marketing: Identificare comunità chiave e opinion leader che probabilmente parleranno dei tuoi prodotti e hanno l'abilità di influenzare le opinioni altrui.

---

[3] **CAIAZZO, D., COLAIANNI, A., FEBBRAIO, A., MASI, D.** *Buzz marketing nei social media. Come scatenare il passaparola on-line, Fausto Lupetti Editore 2009, pos. 564 di 1766*

- Cause Marketing: supportare cause sociali per guadagnare il rispetto e il supporto delle persone che si sentono fortemente legate a quella causa.
- Conversation Creation: consiste in pubblicità interessante o divertente, email, catch phrases, intrattenimento o promozioni designate a far iniziare una attività di word of mouth.
- Brand Blogging: Creare blog e partecipare nella blogosfera con lo spirito di una comunicazione aperta e trasparente. Condividere informazioni importanti con la community, che potrebbe parlare di queste in seguito.
- Referral Programs: creare strumenti pe permettono ai clienti soddisfatti di suggerire ai loro amici contenuti.

È ovvio quindi che vi sia grande attenzione da parte del mondo del business verso il fenomeno del word of mouth online. Viene comunemente chiamato word of mouse o e-WOM ed ha assunto dimensioni mai raggiunte prima grazie alla diffusione delle nuove tecnologie di comunicazione che ne hanno amplificato e accelerato l'efficacia. I media digitali hanno infatti profondamente cambiato il modo in cui le informazioni vengono prodotte e distribuite. L'online Word-of-mouth presenta numerosi vantaggi:

- la rapida e ampia circolazione delle informazioni attraverso blog
- discussioni fra gente comune su forum e social network
- il fatto che esse rimangano disponibili in eterno e accessibili tramite una semplice ricerca tramite un motore di ricerca una volta indicizzate
- la possibilità per le aziende di monitorarne gli effetti delle azioni di WOM marketing

1.2 Le teorie dell'influenza selettiva e la brand advocacy

La sociologia ha studiato ampiamente l'importanza della WOM, in particolare dalle teorie dell'influenza selettiva sviluppatesi fra gli anni quaranta e cinquanta del ventesimo secolo. Esse raccolgono un vasto ed eterogeneo insieme di teorie fondate sul paradigma cognitivo generale della psicologia, ossia che l'influenza di un soggetto su un organismo determina risposte che sono proporzionate alle differenze esistenti fra gli individui. Sono tutte accomunate da una forte attenzione all'analisi del rapporto fra comportamento individuale e comportamenti di gruppo attivati dai mezzi di comunicazione di massa. La tabella sottostante sintetizza le varie teorie. Ci soffermeremo ad analizzare la teoria delle relazioni sociali.

Le teorie dell'influenza selettiva	
Teoria delle differenze individuali	• Teoria dell'apprendimento • Analisi degli istinti e degli atteggiamenti • Psicografie e segmentazione
Teoria della differenziazione sociale	• Ricerca empirica e analisi delle subculture • Teoria degli uses and gratifications • Studi di Lasswel e Lazarsfeld
Teoria delle relazioni sociali	• Two-step flow of communication

[4]

Nel 1955 Paul Lazarsfeld ed Elihu Katz pubblicarono Personal Influence: the Part Played by People in the Flow of Mass Communication. È qui che elaborarono la ormai famosa teoria del two step flow of communication. I due studiosi affermavano che non esiste un flusso unitario di informazioni che si muove dai media ai destinatari finali. Il flusso comunicativo segue un percorso composto da due fasi: la prima dai media agli opinion leader, la seconda dagli opinion leader al gruppo sociale di riferimento. L'opinion leader attua una mediazione, egli a sua volta influenza attraverso canali interpersonali gli individui meno esposti ai media.

La teoria introduce due concetti molto interessanti: il concetto di gruppo sociale e la nozione di opinion leader. Ma cosa è un opinion leader? È un "individuo con più o meno ampio seguito di pubblico che ha la capacità di influenzare le opinioni e gli atteggiamenti degli altri e che, dunque, può avere un ruolo determinante nella diffusione di un certo modello di comportamento o di un particolare bene di consumo"[5]. È un membro del gruppo sociale più disponibile all'esposizione dei media e più competente nell'uso degli stessi. Oggi il termine viene molto usato nel marketing e in ambito pubblicitario. Indica "quelle persone che, in virtù della loro capacità di esercitare una determinata influenza nei confronti dell'opinione pubblica, costituiscono per le imprese un target prioritario cui indirizzare messaggi pubblicitari, al fine di accelerarne l'accettazione presso

[4] *Sorice, M. (2009). Sociologia dei mass media, Carocci editore, p. 72*
[5] *http://www.glossariomarketing.it/significato/opinion-leader/*

un pubblico più vasto".[6] La teoria del two-step flow of communication considera quindi i contatti personali come più in grado di influenzare efficacemente il gruppo sociale di riferimento rispetto ai soli media. Detto in altre parole: il passaparola è più potente di qualsiasi messaggio mediale.

Questo filone di studi si poneva in netta contrapposizione alla teoria dell'ago ipodermico o magic bullet theory che, invece, descriveva i media come onnipotenti. Non era una vera e propria teoria scientifica, va più interpretata come una modalità di lettura dei media intuitiva e immediata, vicina al sentire della gente comune (e alle sue paure). Le strategie di propaganda bellica incontrate durante il primo conflitto mondiale e soprattutto l'uso massiccio dei media da parte dei regimi totalitari potevano effettivamente suggerire una visione dei mezzi di comunicazione di massa decisamente pessimista: secondo la prospettiva ipodermica la radio, la stampa e il cinema altro non erano che potentissimi strumenti in grado di inoculare sotto la pelle delle persone qualsiasi tipo di messaggio (da qui la metafora dell'ago ipodermico). La massa era un bersaglio unico e informe, facile da colpire e controllare con i proiettili sparati dai mass media. L'intera teoria può essere, a dire il vero, riassunta con la semplice frase: i media manipolano le persone. I sociologi del tempo stavano analizzando il passaggio dalla Gemeinschaft (la comunità tradizionale basata sulla comunanza di sangue e di luogo) alla Gesellschaft (la società moderna asettica e impersonale). L'idea dominante era quindi quella di una società che stava diventando sempre più atomizzata, costituita da una moltitudine di individui alienati, privi di legami significativi tra loro e quindi soli di fronte ai messaggi dei media. In altre parole una società di massa. "Su questo modello sociologico della società di massa, la teoria ipodermica innesta un modello comunicativo altrettanto semplice, mutuato dalla psicologia comportamentista: il modello stimolo-risposta (S-R). Applicato al mondo della comunicazione il comportamentismo riconosceva in ogni messaggio mediale uno stimolo in grado di produrre una risposta identica nei comportamenti del pubblico. Nel modello S-R, stimolo e risposta rappresentano un'unità indissolubile, non esistono stimoli che non producono risposte, così come non esistono risposte che non siano state provocate da stimoli ben precisi. Il rapporto tra i due elementi è caratterizzato dalla causalità, dall'immediatezza e dalla necessitò: nel caso della comunicazione di massa, ogni messaggio è destinato a provocare senz'altro un preciso comportamento nelle persone colpite. Una simile prospettiva concede naturalmente ben poca autonomia al pubblico, che viene visto come un esercito di automi in balia dei media"[7]. Questo ci fa capire ancora di più quanto la teoria del two step flow of communication fu rivoluzionaria. Ricapitolando: l'influenza dei contatti personali è più importante rispetto a quella esercitata solamente dai "media onnipotenti". I mezzi di comunicazione, quindi, non fanno altro che partecipare alla efficacia comunicativa. Non risultano gli unici responsabili del cambiamento di opinione, non esercitano un controllo mentale sull'individuo. Katz e Lazarsfeld nei loro studi effettuati negli anni cinquanta evidenziano benissimo tutto questo. Una alta esposizione a contatti personali influenzerà molto (se non moltissimo) il

---

[6] *idem*
[7] *PACCAGNELLA, L. (2004). Sociologia della Comunicazione, il Mulino, p.98*

consumatore nella scelte personali su cosa consumare (cinema, radio, televisione, siti internet, giornali), molto di più di quanto lo influenzerebbe l'esposizione ai quotidiani o a qualunque altro mass media.

Le ricerche di Lazarsfeld, Berelson, Gaudet e in seguito lo studio congiunto di Katz e Lazarsfeld considerano il ruolo dei gruppi sociali e delle relazioni interpersonali nella fruizione mediale fondamentali, tanto da portare ad una influenza selettiva nella fruizione dei mass media: l'audience appare dotata di una capacità selettiva che le permette di selezionare i materiali informativi che riceve in maniera netta, molto di più rispetto a quanto ipotizzato dai comportamentisti. "Se la gente tende a esporsi soprattutto alle comunicazioni di massa secondo i propri atteggiamenti e i propri interessi e a evitare altri contenuti e se, per di più, tende a dimenticare questi altri contenuti appena se li trova davanti agli occhi e se, infine, tende a travisarli anche quando li ricorda, allora è chiaro che la comunicazione di massa molto probabilmente non ne cambierà il punto di vista. È di gran lunga molto più probabile anzi che essa rafforzerà le opinioni preesistenti".[8] La teoria del two step flow continuò ad influenzare i sociologi per anni. Ecco un altro estratto molto interessante: "nacque una ricca letteratura da cui risultava che le relazioni sociali informali erano importantissimi fattori intervenienti che determinavano il modo in cui le persone selezionavano il contenuto dei media, lo interpretavano e agivano di conseguenza. Così, la teoria delle relazioni sociali andò ad arricchire ulteriormente le conoscenze delle dinamiche e dei fattori alla base della selettività esercitata dai pubblici nella loro risposta alle comunicazioni di massa".[9]

[8] *KLAPPER, J. T. (1960). Effects of Mass Communication, trad. it. p. 246*
[9] *DE FLEUR, M. L., BALE-ROKEACH, S. (1989). Theories of Mass Communication, trad. it. p. 211*

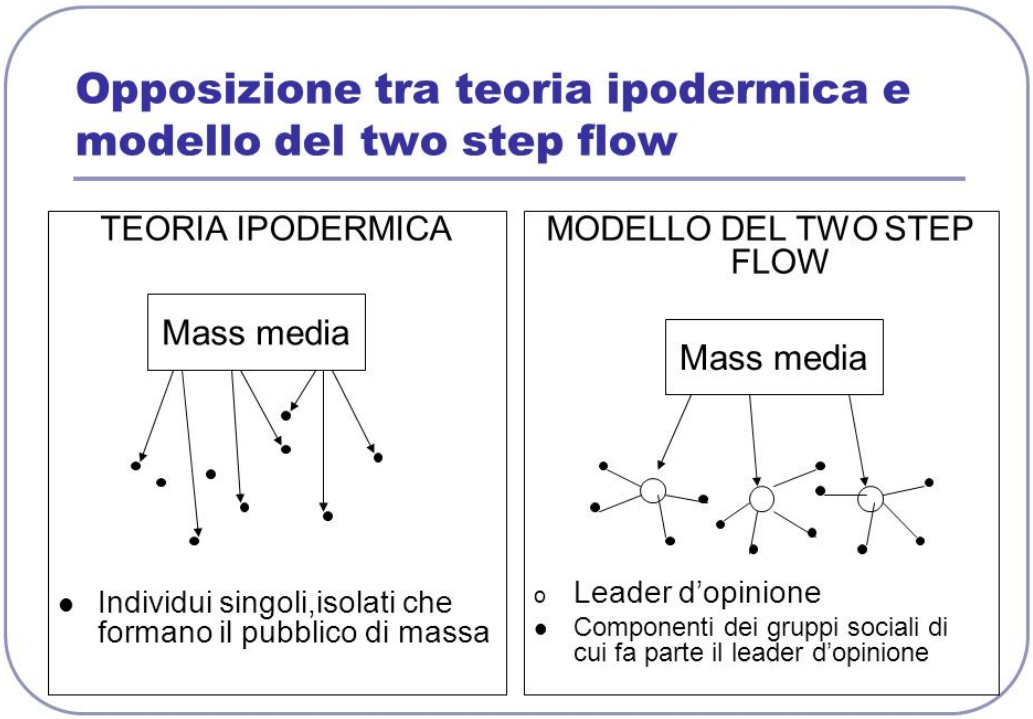


Figura 1 schema opposizione teoria ipodermica e modello del two step flow

10

Recentemente il modello del two step flow è stato ulteriormente rafforzato dagli studi sulla "brand advocacy". Una ricerca del 2008 effettuata dall'agenzia Weber Shandwick ha infatti scoperto che la brand advocacy, che potremmo definire come la forza del passaparola generato dagli advocate del brand, è cinque volte più efficiente nel generare commitment rispetto a qualsiasi investimento pubblicitario[11]. Andiamo però per ordine.

Quelle di cui parleremo ora sono diverse forme di influencer marketing. L'influencer marketing è una forma di marketing che consiste nell'identificare individui chiave che possono trasmettere contenuti importanti alla nostra audience di riferimento. Per riuscire a contattare e lavorare con gli influencer è necessario costruire delle relazioni. Queste vengono chiamate Public Relations (o PR) con le quali quindi si intende "l'insieme di attività che, attraverso lo sviluppo di relazioni con soggetti di una qualche influenza per il business dell'impresa, sono intese a promuovere o a proteggere l'immagine dell'impresa stessa e dei suoi prodotti"[12].

---

[10] http://slideplayer.it/slide/949961/3/images/50/Opposizione+tra+teoria+ipodermica+e+modello+del+two+step+flow.jpg
[11] CAIAZZO, D., COLAIANNI, A., FEBBRAIO, A., LISIERO, U., (2009). Buzz marketing nei social media, p. 62
[12] http://www.glossariomarketing.it/significato/public-relations/

Una buona relazione porterà beneficio ad entrambe le parti: farà crescere in modo organico la reach dell'azienda dando nel frattempo qualcosa in cambio all'influencer. Una classica forma di Public Relations è il celebrity endorsement, ossia promuovere un prodotto associandolo all'immagine e alle caratteristiche di un personaggio molto noto. Di recente abbiamo assistito ad una rapida evoluzione delle PR: è aumentata esponenzialmente l'importanza dell'Influencer Outreach Strategy (essa può essere quindi considerata come la forma post-moderna del celebrity endorsement). Abbiamo però visto anche come le aziende abbiano iniziato a dare sempre maggiore importanza ai brand advocate. "Il brand advocate è un cliente talmente soddisfatto dei prodotti o servizi offerti da un'impresa da consigliarli ai propri conoscenti attraverso il passaparola (Word-Of-Mouth o WOM)."[13] Facciamo quindi ora un breve confronto fra gli influencer e i brand advocate.

Un recente studio di Forrester[14] ha analizzato quanto i consumatori abbiano fiducia negli influencer (ha preso in considerazione blogger, opinionisti e celebrità) ed è risultato che solamente il 18% ha fiducia in loro. Uno studio condotto dalla Nielsen[15] ha invece dimostrato che la fiducia dei consumatori nei brand advocate ha un tasso del 92%, che è lo stesso livello di fiducia che avrebbero in un amico o in un parente. Un influencer è definito tramite dimensione della sua audience (numero di follower su Twitter, numero di persone iscritte al suo blog, follower sul suo canale youtube). Un brand advocate è invece definito tramite la probabilità che raccomandi un prodotto. Passando alle motivazioni che guidano i due: l'influencer è interessato solamente a far aumentare la sua audience, il brand advocate è interessato ad aiutare i suoi amici. Gli influencer rimarranno fedeli per poco tempo, i brand advocate rimarranno fedeli a lungo. Un influencer non è necessariamente guidato da una passione sincera, un brand advocate sì. Un influencer solitamente ha bisogno di incentivi economici, un brand advocate no.

[13] *http://www.glossariomarketing.it/significato/brand-advocate/*
[14] *http://www.zuberance.com/downloads/brandAdvocateInsights.pdf*
[15] *http://www.nielsen.com/us/en/newswire/2012/consumer-trust-in-online-social-and-mobile-advertising-grows.html*

	Influencer	Brand advocate
Fiducia del consumatore	18%	92%
Profilo tipico	Opinionista, celebrità, blogger	Cliente soddisfatto
Definito da	Dimensione della sua audience	Quanto sia probabile che raccomandino il tuo prodotto ad altri
motivazione	Far aumentare la propria audience	Aiutare amici e persone care
Sostegno e fedeltà	Breve termine	Lungo termine
Sincerità del sostegno	forse	sì
Necessità di incentivi economici	Tipicamente sì	Tipicamente no

Molto spesso si tende a confondere audience con influence. Avere un ampio numero di persone che ci segue non implica che noi siamo influenti, significa che abbiamo una audience ampia (ben pochi influencer sono in grado di guidare i comportamenti di masse di persone). Un altro problema è che molto spesso gli influencer hanno una propria agenda: maggiore è la loro fama maggiore è la difficoltà nell'attirare la loro attenzione per far promuovere il tuo prodotto (ciò spesso implica incentivi economici sostanziosi). Il brand advocate è invece una marketing force sostenibile. Desiderano engagement nei confronti del tuo marchio e quindi, al contrario degli influencer, non aspettano altro che supportarti, promuoverti, difenderti anche nel lungo periodo.

1.2.1 Il caso americano: l'armata di Trump su Twitter e Reddit

I politici moderni hanno molto in comune con i marketer: per sopravvivere in un ambiente così ostile e competitivo entrambi devono concentrarsi sullo storytelling e sulla costruzione di un brand. La politica è solamente una forma più elevata di marketing dove invece di vendere un prodotto si cerca di vendere la propria personalità.

Intendo qui analizzare rapidamente le elezioni americane su cui molto si è discusso, anche perché reputo che gli Stati Uniti abbiano semplicemente anticipato quello che succederà anche in Italia.

A luglio il vincitore sui social è decisamente Donald Trump. Analizziamo alcuni dati[16]:

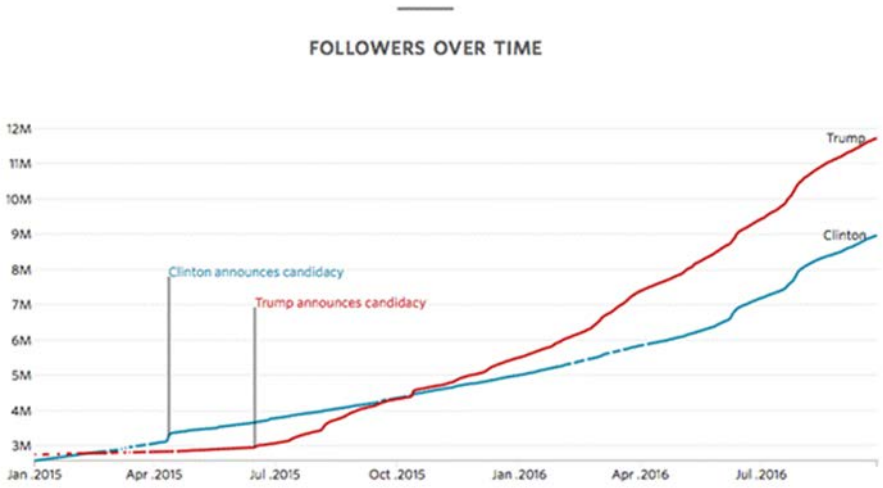


Figura 2 I follower durante il tempo. Image: Wall Street Journal

17

	Donald Trump	Hillary Clinton
Follower	10.267.655	7.765.519
Nuovi follower al giorno	30.574	22.086
Retweet totali	12 milioni	5,5 milioni
Like totali tweet	33 milioni	12 milioni

[16] *https://www.weforum.org/agenda/2016/08/hillary-clinton-or-donald-trump-winning-on-twitter/*
[17] *Image: Wall Street Journal*

Retweet medi per tweet	5639	2154
Tweet medi al giorno ultimi 6 mesi	12	18

Durante la campagna elettorale quello con una audience più ampia sui social network era indubbiamente Donald Trump: a ottobre era arrivato a 12 milioni 127mila follower mentre Hillary Clinton ne aveva 9 milioni 407mila. Stando a queste statistiche risalenti al 2 agosto 2016[18] Trump guadagna in media 30.574 nuovi follower al giorno mentre la Clinton ne guadagna 22.086. Per calcolare però chi dei due è più effettivo dobbiamo concentrarci sul numero di retweet: Trump ha ottenuto 12 milioni di retweet mentre la Clinton ne ha ottenuti solo 5,5 milioni. Questo ci aiuta a capire il livello di engagement. Facendo una semplice proporzione la Clinton data la sua base di follower ne avrebbe dovuti ricevere 9.076.060 per avere la stessa effettività di Trump. Trump ha un numero di like totali decisamente superiore, 33 milioni contro 12 milioni. Anche qui, facendo una semplice proporzione possiamo vedere che data la sua base di follower la Clinton avrebbe dovuto avere 24.959.166 like totali per stare ai livelli di Trump. Come retweet medi Trump ne ha 5639 mentre la Clinton ne ha 2154. Anche qui, facendo un semplice calcolo notiamo che avrebbe dovuti averne almeno 4264 data la sua base di follower. L'unico punto in cui la Clinton supera Trump è il numero di tweet al giorno. Devo dire che mi ha sorpreso, visto che da come è stato dipinto il presidente in campagna elettorale mi sarei aspettato il contrario. Possiamo quindi dire che la candidata democratica non solo ha una base numericamente inferiore, ma anche meno affezionata.

Passiamo ora ad analizzare più da vicino l'armata di Trump su Twitter. Uno studio condotto dalla San Antonio database marketing agency Stirista[19] citato da POLITICO e dal Wall Street Journal aveva scoperto dei dati a dir poco sorprendenti. Lo studio cercava di capire quanto fossero effettivamente di supporto nella vita reale i follower dell'account @realdonaldtrump. Lo studio non ha fatto altro che associare i Twitter Handles dei follower di Trump ai dati presenti nei voter database americani. Il risultato è stato il seguente: 7 follower su 10 erano sostenitori anche nella vita reale, il 90% di essi sarebbe andato quasi sicuramente a votare e che solamente per l'11% di essi era la prima volta (ricordiamo che gli Stati Uniti hanno una affluenza elettorale molto bassa).

Andiamo ad analizzare da vicino questi follower.

20

[18] *idem*
[19] *http://www.stirista.com/wpcontent/uploads/2016/06/WhosFollowingTrumpAndClinton-1.pdf*
[20] *Wall Street Journal*

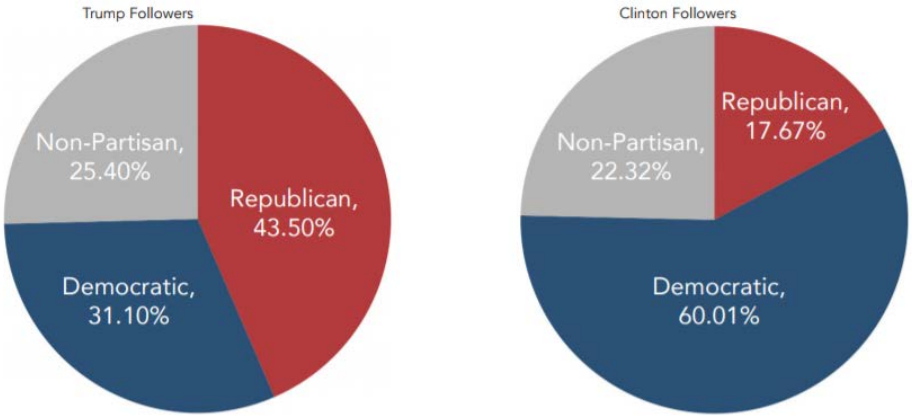


Figura 3 la provenienza dei follower

Ben il 31% dei follower di Trump aveva votato Democratico alle elezioni precedenti stando a quanto ha scoperto Stirista. Su questo punto intendo tornare in seguito.

Dal punto di vista della razza nulla di interessante, le minoranze favoriscono i democratici mentre i bianchi favoriscono i repubblicani. Questo studio conferma quanto detto più volte da tutti i media.

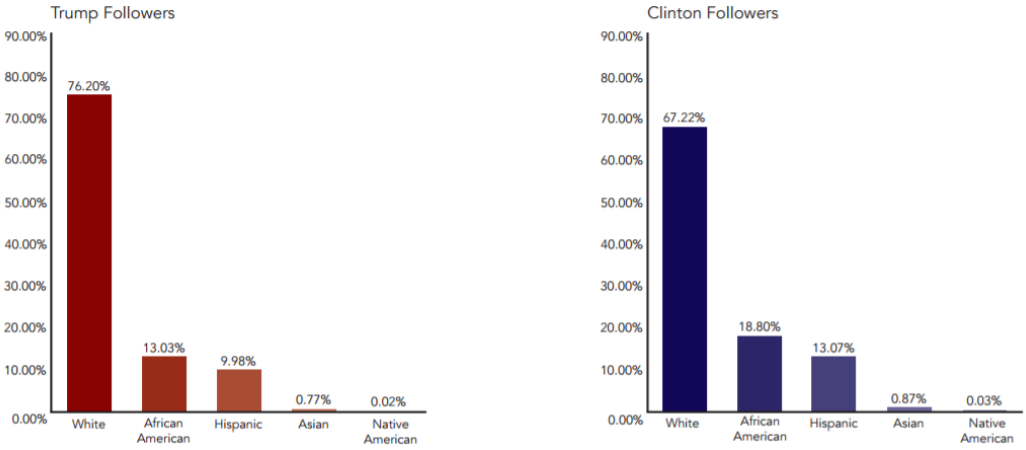


Figura 4 la razza dei follower

Lo stesso vale per il genere, con i follower di Trump composti al 56% da uomini e al 44% da donne, mentre i follower della Clinton sono al 43% uomini e al 57% donne.

Le cose diventano più interessanti quando si va invece ad analizzare l'età, il reddito e il titolo di studio dove non troviamo alcuna differenza statisticamente significativa.

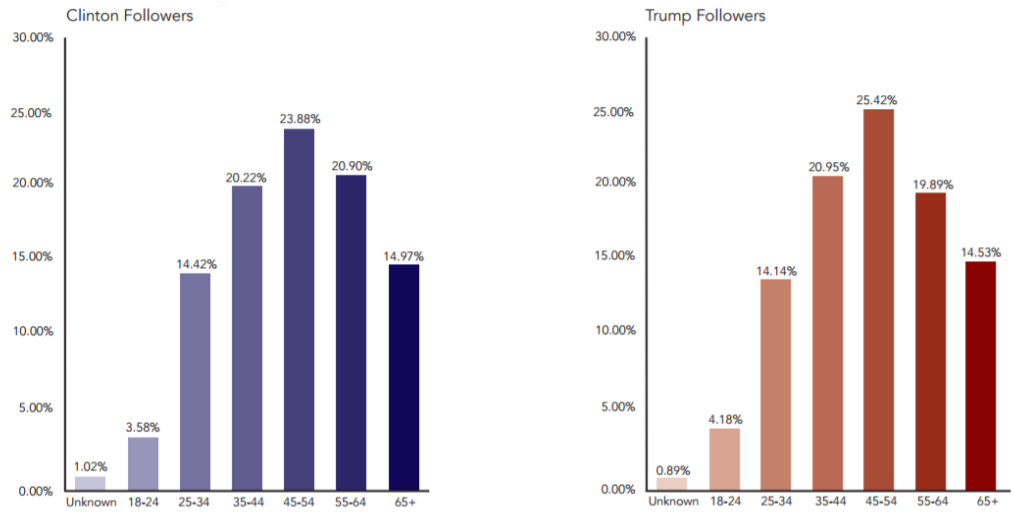


Figura 5 l'età dei follower

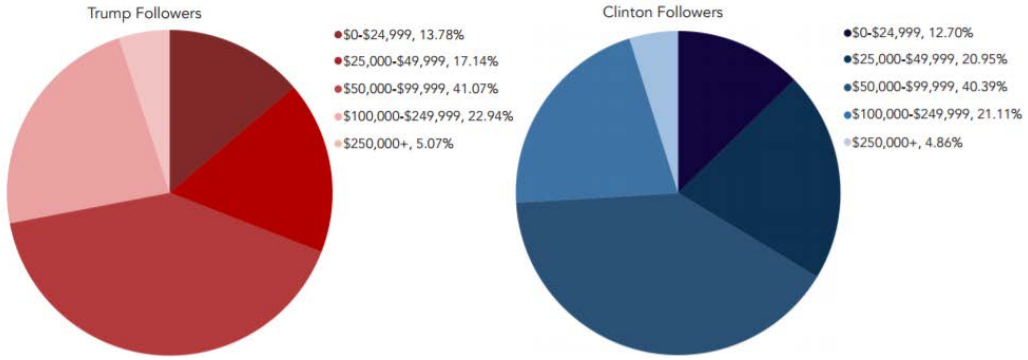


Figura 6 il reddito dei follower

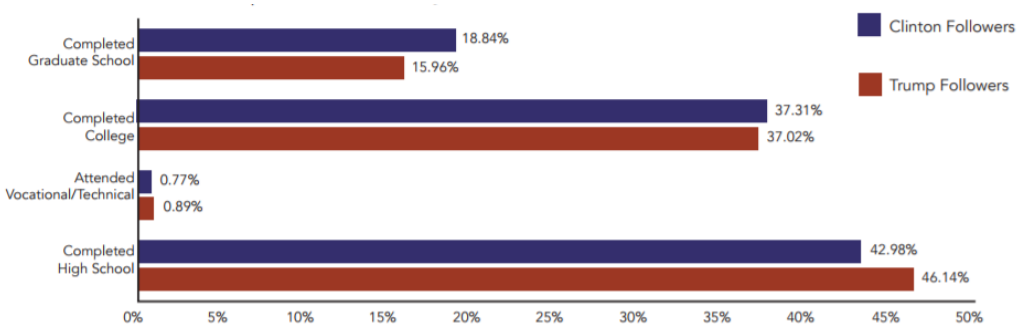


Figura 7 il grado di istruzione dei follower

Un ruolo ancora più importante è stato però svolto da Reddit, potremmo addirittura dire che Twitter ha avuto un ruolo ancillare. Trovo utile però prima introdurre questo sito che in Italia non è molto conosciuto. Per

popolarità è l'ottavo sito al mondo, il quinto negli Stati Uniti[21]. Si autodefinisce "the front page of the internet", la copertina di internet. È un sito contenente notizie, intrattenimento e social media content. Con social media content si intende una parte specifica "del Social Media Management (SMM), un'estensione del Digital marketing che si occupa di dare visibilità alle aziende o brand attraverso i social media, le comunità digitali e le diverse piattaforme web"[22]. Tali contenuti sono creati dagli utenti registrati della community e comprendono testi, link, video e immagini. In poche parole è un Bulletin Board System. Questi siti esistono da molto tempo, sin dai primi giorni in cui furono inventati i modem dial-up. La loro origine è così antica che in passato non esisteva nemmeno una interfaccia grafica (GUI). Questi primi siti BBS erano molto grandi in passato, nel 1994 c'erano 17 milioni di utenti solamente negli Stati Uniti. I primi BBS erano gestiti a livello amatoriale da persone appassionate, il primo tentativo di creare una rete professionale destinata al pubblico generale fu Usenet, di cui molti considerano Reddit il diretto erede.

- Sia Reddit che Usenet permettono agli utenti di postare contenuti: entrambi permettono di postare testo, immagini, link e interazione sociale
- Il contenuto è categorizzato per interessi, su Reddit questi si chiamano "subreddit" mentre su Usenet queste categorie si chiamano "Newsgroup"
- Entrambi generano un senso di "community" nei membri che ci partecipano i quali guardano e postano contenuti nei gruppi specializzati.
- Entrambi hanno una elevata quantità di contenuti volgari
- Chiunque può avere accesso a Usenet e Reddit

Senza scendere ulteriormente in dettagli tecnici, dal punto di vista culturale Usenet pur non essendo conosciuto dalla maggioranza degli utenti italiani ha svolto un ruolo fondamentale: elementi come le emoticon, il flaming, i troll e la maggior parte degli acronimi slang come "LOL" sono nati su Usenet. Ci tengo a dire tutto questo perché la maggioranza dei giornalisti e dell'opinione pubblica crede che questi siano fenomeni moderni, quando invece abbiamo avuto da sempre comunità con un elevato senso di appartenenza, composte da individui con un'alta conoscenza informatica e un elevato quoziente intellettivo (anche se estremamente triviali).

Quello che ci interessa è il subreddit r/The_Donald che durante la campagna elettorale era arrivato intorno a 500.000 utenti e ha svolto un ruolo importantissimo. Lo stesso Trump ha scritto in questo subreddit aprendo una sezione dove si proponeva di rispondere alle domande dei suoi sostenitori. Purtroppo il progetto non è durato molto e non ha dato molte risposte.

---

[21] *https://www.alexa.com/siteinfo/reddit.com*
[22] *https://www.gruppodigitouch.it/servizi/amplification/social-media-content/*

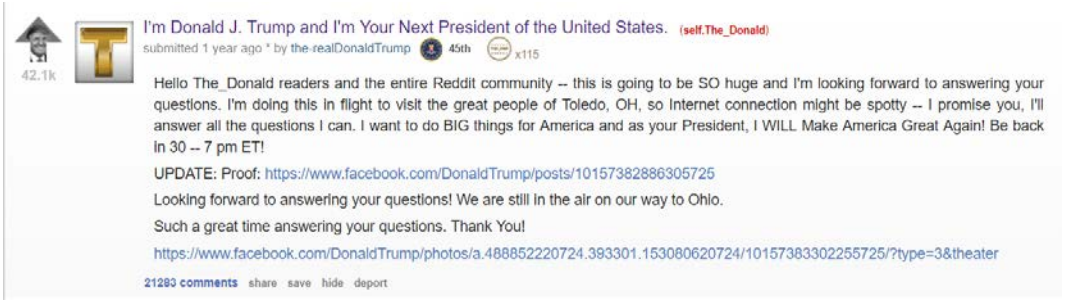


Figura 8 Trump che risponde alle domande su Reddit

È una comunità molto coesa con una propria identità e un proprio linguaggio: i loro membri si definiscono ad esempio "centipede" (centopiedi). Desidero soffermarmi un su questo aspetto che reputo particolarmente significativo. La scelta di un termine ci fa capire lo spirito di questa intera campagna elettorale post-moderna e l'importanza che i brand advocate, di cui parlavamo in precedenza, hanno assunto. Prima di tutto: ci troviamo di fronte ad un meme. Si sente molto spesso parlare di meme ultimamente, sono molto utili ad un politico perché riescono a trasmettere messaggi complessi in pochissimo tempo (sopperendo all'enorme problema della bassissima soglia di attenzione degli utenti dei vari social network) e sono utili a superare i meccanismi difensivi che utenti con una determinata ideologia politica potrebbero mettere in atto per non ascoltare messaggi che mettano in discussione ciò in cui credono. Iniziamo però dall'inizio. "Un meme è un'unità di trasmissione culturale (uno slogan, un pensiero, una melodia, un concetto di moda, filosofia, politica) che si trasmette di cervello in cervello. I memi lottano per riprodursi e si diffondono fra la popolazione in maniera molto simile al modo in cui i geni vanno a caratterizzare una specie biologica. I memi più potenti sono in grado di cambiare le menti, di alterare i comportamenti, di catalizzare cambiamenti collettivi di opinione e di trasformare intere culture. Ecco perché la guerra dei memi è diventata la principale battaglia geopolitica dell'era dell'informazione. Chiunque sia in grado di controllare i memi ha, di fatto, il potere fra le mani.[23] Trump ha condiviso sul suo account Twitter molti memi, che hanno fatto discutere i suoi oppositori regalandogli pubblicità gratuita e compattando i suoi sostenitori. Tornando all'analisi del nostro meme da cui deriva il termine "centopiedi": per analizzare il meme bisogna analizzare ogni singolo "strato sovrapposto" di cui è composto. Il "contenitore", il "formato" è quello di un video. Il video ha delle immagini (prese dai vari discorsi di Trump, in questo caso le primarie repubblicane) e una melodia (come diceva la stessa definizione di meme citata precedentemente). Il contenitore è un video YouTube da uno dei tanti fan di Trump, "You Can't Stump the Trump (Volume 4)"[24] retwittato dallo stesso Trump. Le immagini rappresentano i vari attacchi nei confronti degli altri concorrenti repubblicani. Un altro stato è invece costituito dalla melodia. La melodia è a sua volta è un contenitore: essa utilizza pezzi remixati di un documentario sui centopiedi che

[23] *LASN, K., (1999). Culture Jam: The Uncooling of America, Eagle Brook, p. 187*
[24] *https://www.youtube.com/watch?v=MKH6PAoUuD0*

recita: “Despite it's impressive length, it's a nimble navigator, and some can be highly venomous. As quick as lightning, just like the tarantula it's killing, the centipede has two curved hollow fangs which inject paralyzing venom. Even tarantulas aren't immune from an ambush. This centipede is a predator...”. Tradotto: nonostante la sua impressionante lunghezza è un agile nuotatore e alcuni possono essere altamente velenosi. Veloce come il lampo, proprio come la tarantola che sta uccidendo, è dotato di due zanne cave che iniettano veleno paralizzante. Questo centopiedi è un predatore.

Trump è il “centopiedi”, il “predatore” dotato di “zanne cave in grado di iniettare veleno paralizzante”. Si stanno riferendo ai suoi modi da “duro” che non chiede mai scusa e dice quello che vuole dire senza filtri (quindi ci riferiamo alle azioni di Trump in prima persona, alla sua figura di leader carismatico, al suo culto della personalità). Tuttavia anche un predatore così temibile per muoversi ha bisogno delle sue innumerevoli zampe: queste rappresentano i suoi sostenitori più fedeli. Anche loro si chiamano fra di loro “centopiedi” su Reddit. È complicato capire il concetto di “centopiedi”: unità distinte ma uniche e unite, il leader che viene identificato come “padre” e guida, senza tuttavia dimenticare la sua natura umana, il fatto che senza la sua base coesa di sostenitori il suo potere è nullo.

È proprio questo il suo punto di forza, non è possibile immobilizzare il centopiedi calpestandolo (come dice appunto il titolo del video) poiché non appena viene calpestata una zampa ha tutte le altre che lo sostengono (la sua comunità compatta). A calpestare il centopiedi sono ovviamente i “globalisti” (parola ricorrente all’interno del subreddit). Il predatore sarà quindi sempre in grado di avanzare e combattere fino alla vittoria finale.

Addirittura Ben Garrison, vignettista politico molto popolare nell’ambiente conservatore, è arrivato a dedicare una vignetta al “centipede”.

Figura 9 Ben Garrison dedica una vignetta al centipede

Ho trovato importante soffermarmi a descrivere questo fenomeno perché questa community non è incentrata solo sulla creazione di meme o altri fenomeni legati alla rete. I cinquanta moderatori del subreddit hanno abilmente canalizzato e focalizzato il potenziale della community insegnando loro come supportare Trump nel mondo reale. Hanno fornito un host per un Wiki contenente tutte le policy proposte da Trump e hanno istruito i supporter del subreddit su come aiutare con la campagna. I membri suggerivano spesso strategie su come argomentare in modo persuasivo per convincere altre persone a supportare

Trump.[25] È stato inoltre utile per raccogliere volontari per la campagna di Trump. Qui ad esempio nella fig. 10 si cercavano volontari per la campagna in Ohio (leggendo i commenti successivi si trova addirittura gente proveniente da stati limitrofi che si propone per aiutare).

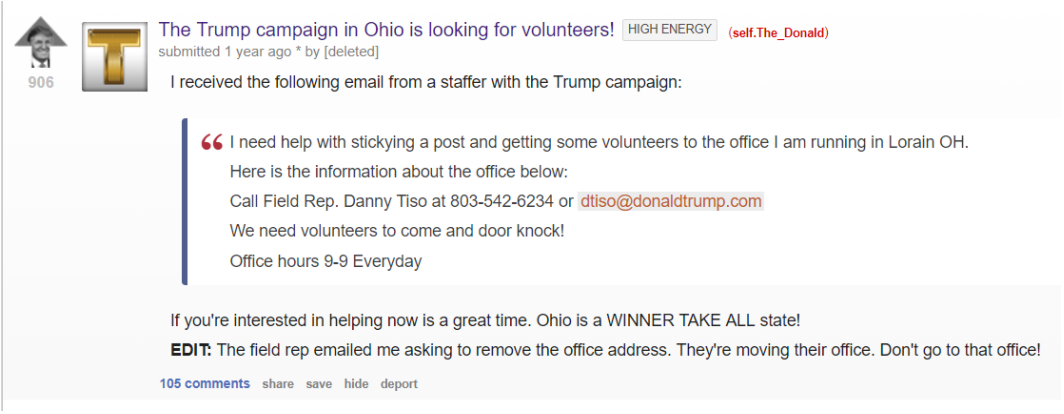


Figura 10 reddit come mezzo per reclutare volontari

Numerose ricerche hanno analizzato perché i social media rafforzano l'attivismo.[26] [27] [28] Le spiegazioni date sono fondamentalmente due:

- L'uso dei social media aiuta i movimenti sociali a pubblicizzare cause locali ad audience distanti, il tutto a basso costo.
- Attraverso questi strumenti gli attivisti sono in grado di migliorare le loro comunicazioni logistiche ed organizzare meglio proteste ed eventi.

Un altro ruolo importantissimo lo ha svolto nell'analizzare tutte le email di Hillary Clinton non appena venivano postate su WikiLeaks. Ecco un esempio (fig. 11).

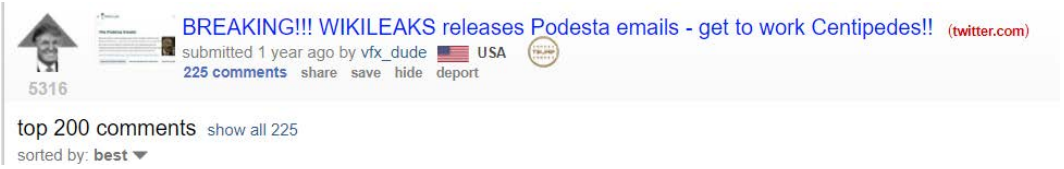


Figura 11 i sostenitori di Trump analizzano le email di Hillary Clinton

[25] *https://www.nytimes.com/2016/11/20/opinion/sunday/reddit-and-the-god-emperor-of-the-internet.html?mcubz=3*
[26] *HARLOW, S. HARP, D. (2012). Collective action on the Web: A cross-cultural study of social networking sites and online and offline activism in the United States and Latin America. Information, Communication & Society, 15(2), 196-216.*
[27] *KARPF, D. (2010). Online political mobilization from the advocacy group's perspective: Looking beyond clicktivism. Policy & Internet, 2(4), 7-41*
[28] *REBER, B. H., KIM, J. K. (2006). How activist groups use websites in media relations: evaluating online press rooms. Journal of Public Relations Research, 18(4), 313-333*

Le loro scoperte venivano poi raccolte e utilizzate dai reporter dei media mainstream, dai media digitali conservatori[29] ed eventualmente postati su Twitter dallo stesso Trump.

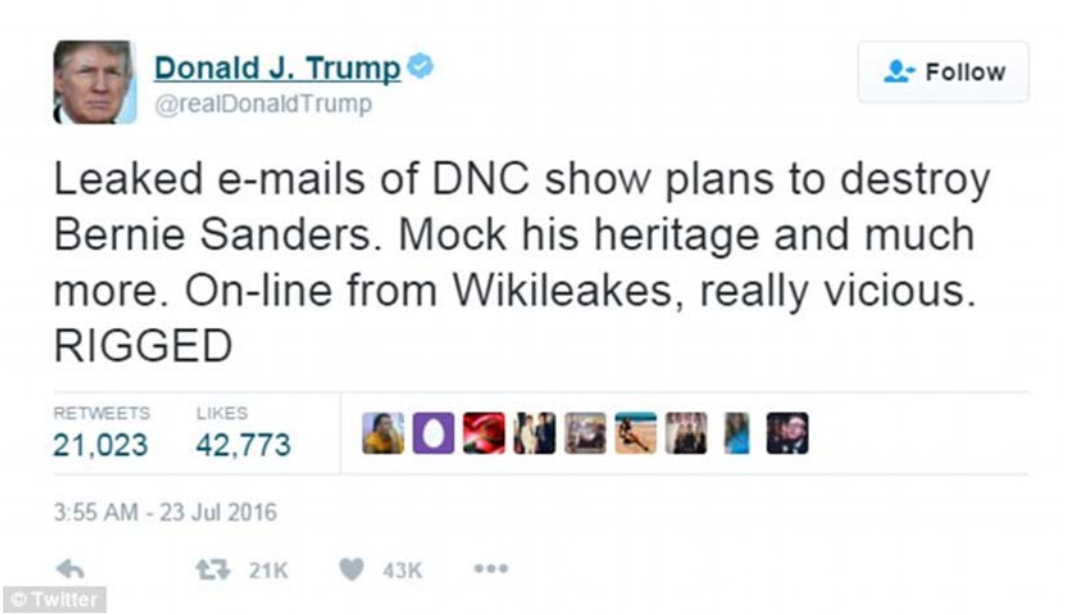


Figura 12 Trump condivide quanto scoperto nelle email

Le mie conclusioni a riguardo sono le seguenti: nelle campagne elettorali post-moderne il segreto per vincere è nel saper scalare la catena mediatica. Scalare la catena mediatica significa "partire da un media inferiore (per costi e portata) e finire gradualmente su media superiori"[30]. Per poter essere in grado di fare tutto questo in maniera così efficace i politici devono avere un nucleo di fedelissimi che li aiutino in tutte le fasi del processo (Reddit con The_Donald) e una base più ampia che dia risonanza e dia la spinta propulsiva al loro messaggio, sia su internet che nel mondo reale (in questo caso Twitter). Abbiamo visto dall'analisi precedentemente citata dei follower degli account Twitter di Trump e della Clinton che Twitter non è più limitato solamente ai giovani, viene usato trasversalmente da tutte le fasce di età. Nello specifico ecco uno dei tanti scenari che si prospettano:

- Reddit (molte volte aiutato da /pol/ di 4chan[31]) genera un contenuto o scopre una email della Clinton particolarmente compromettente, qualunque cosa che sia in grado di attirare l'attenzione.
- Trump lo pubblica sul suo profilo Twitter. Il contenuto raggiunge quindi i suoi follower che lo condividono con i propri amici e parenti. Qui abbiamo la spinta propulsiva iniziale e siamo nella fase ascendente.

[29] *https://www.dailydot.com/layer8/donald-trump-inauguration-donations-crowdsourced-journalism-reddit-twitter/*
[30] *https://www.dariovignali.net/marketing-politico-ed-elettorale/*
[31] *http://www.repubblica.it/speciali/esteri/presidenziali-usa2016/2016/11/12/news/trump_internet_meme_virali_social_4chan-151826943/*

- I media mainstream notano il contenuto che inizia ad essere condiviso e ne colgono il potenziale virale. Reagiscono e lo pubblicano sulle loro piattaforme con l'intento di criticare Trump. La tempesta perfetta è stata generata dal fatto che quasi tutti i media hanno criticato Trump.
- Il contenuto ha ora raggiunto il pubblico di massa che non utilizza Twitter o addirittura nemmeno utilizza internet abitualmente, siamo quindi nella fase discendente. Qui scatta l'imprevisto, ossia l'elevata sfiducia nei confronti dei media mainstream da parte dei conservatori (ben l'80% crede che i media mainstream siano "fake news"[32] e troppo orientati a sinistra). Per questa parte di pubblico i media mainstream non sono più dei validi gatekeeper. Tutto questo fa scattare un effetto underdog potentissimo. Con effetto underdog intendiamo la tendenza di alcuni elettori a votare il candidato che viene percepito come sfavorito. Chi potrebbe diventare a questo punto un valido gatekeeper? Questo nuovo pubblico potrebbe avvicinarsi a Trump e al suo profilo, prendendo le notizie direttamente da lui, potrebbe avvicinarsi a media su internet percepiti come "alternativi" ai media mainstream, potrebbe iniziare ad avvicinarsi ai vari social network, scoprire gente fra amici e parenti lontani che la pensa come loro. Questo non fa altro che rafforzare la loro simpatia nei confronti del candidato. Dopo essersi informato su media "alternativi" potrebbe essere spinto addirittura a diventare lui stesso un gatekeeper, tentando di convincere altri indecisi a supportare Trump. Potrebbe, in poche parole, diventare un brand advocate.

1.3 Internet e i social network in Italia

Questa è una analisi presentata a luglio 2017 da Audiweb[33].

"La total digital audience rappresenta il consumo totale del mezzo, offrendo informazioni sulla reach totale (utenti unici al netto delle sovrapposizioni tra i device rilevati), le pagine viste (per quanto riguarda la fruizione via browser) e il tempo speso online. La total digital audience è la dimensione più completa del sistema di misurazione messo a punto da Audiweb e disponibile a partire dai dati di gennaio 2014."[34]

[32] *http://thehill.com/homenews/campaign/334897-poll-majority-says-mainstream-media-publishes-fake-news*
[33] *https://www.tvdigitaldivide.it/2017/09/15/audiweb-32-mln-gli-italiani-online-a-luglio-2017/*
[34] *http://www.audiweb.it/dati_it/total-digital-audience_it/*

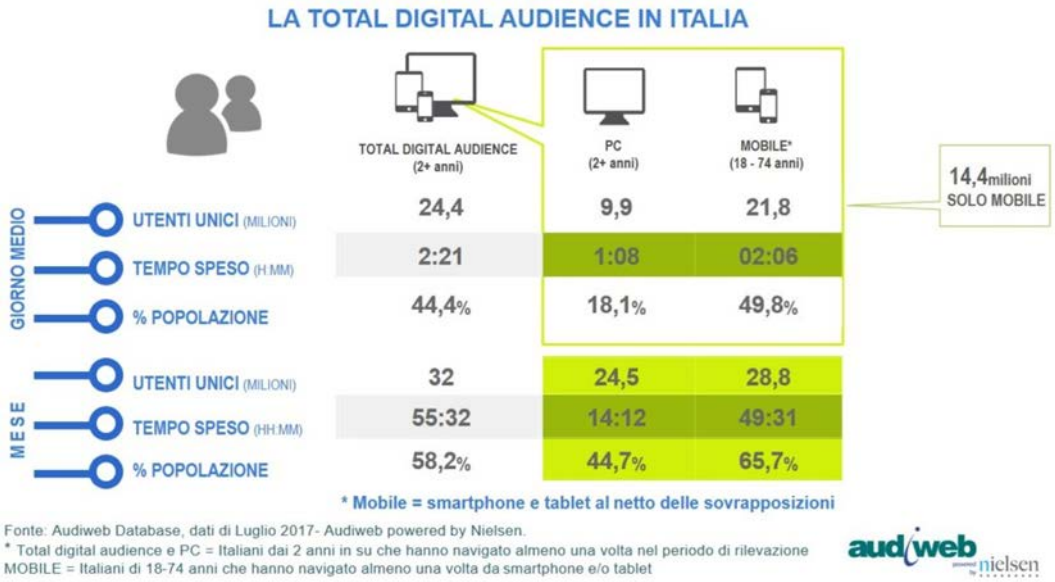


Figura 13 La total digital audience in Italia

Nel mese di luglio 2017, stando alle statistiche di Audiweb, sono stati circa 32 milioni gli italiani dai 2 anni in su che hanno navigato sia da mobile (smartphone e/o tablet) che da PC, collegandosi complessivamente per 55 ore e 32 minuti. I dati mostrano che il 65,7% degli italiani maggiorenni, ossia 28,8 milioni di abitanti, ha navigato da mobile (smartphone e/o tablet), dedicando alla navigazione in mobilità circa 49 ore e mezza. Gli italiani che hanno navigato anche o solo da computer hanno invece trascorso solo 14 ore totali. Nel giorno medio la total digital audience ha raggiunto 24,4 milioni di italiani, online per una durata di 2 ore e 20 minuti tramite i device rilevati.

La fruizione quotidiana dell'online è quindi ormai principalmente spostata sul mobile (smartphone e/o tablet), con 21,8 milioni di utenti fra i 18 e i 74 anni online da questi device. Una quota significativa, 14,4 milioni, ha addirittura navigato esclusivamente in mobilità. La fruizione di internet da PC raggiunge valori inferiori nel giorno medio, con 9,9 milioni di italiani di età superiore ai 2 anni (che diventano 9,5 milioni quando si considerano quelli di età compresa fra i 18 e i 74 anni) che accedono dai device "fissi" per poco più di un'ora.

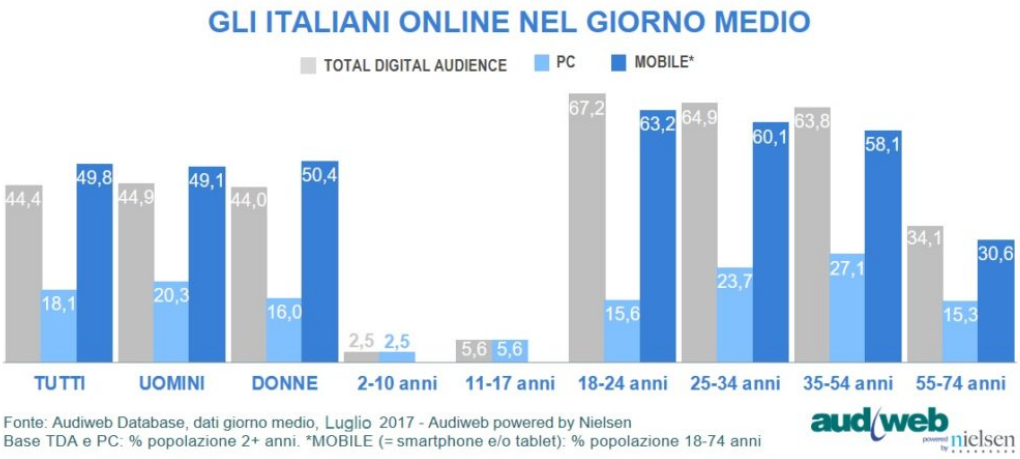


Figura 14 gli italiani online nel giorno medio

Analisi più dettagliate sul tempo speso online attraverso i device rilevati, mostrano che nel mese di luglio 2017 gli utenti maggiorenni hanno dedicato ben l'81% del tempo totale online alla navigazione tramite mobile (smartphone e/o tablet) e solamente il 19% alla navigazione da computer. Device diversi portano a stili di fruizione diversi. Stili di fruizione portano a dover generare tipi di contenuti diversi per cogliere l'attenzione dell'utente. Le donne fanno un uso maggiore di internet, privilegiando i dispositivi mobili. Dedicano all'online da mobile 2 ore e 19 minuti nel giorno medio, mentre gli uomini gli dedicano 1 ora e 54 minuti. I 18-24enni raggiungono invece la soglia delle 2 ore e 43 minuti online da mobile, seguiti dai 25-34enni con 2 ore e 20 minuti.

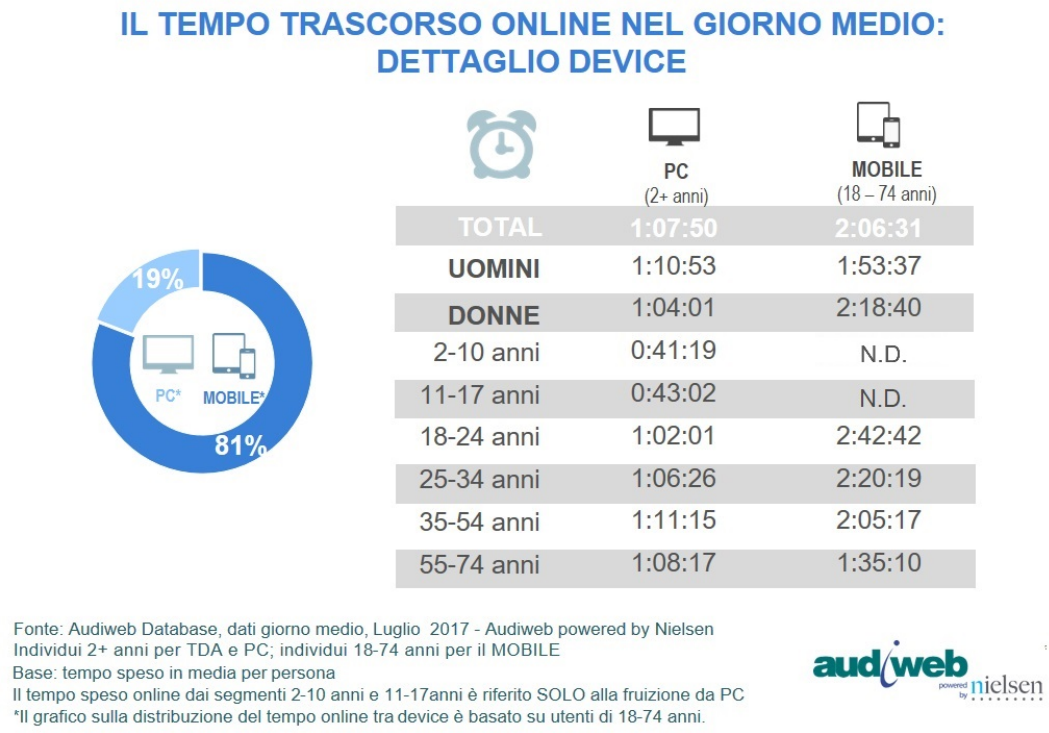
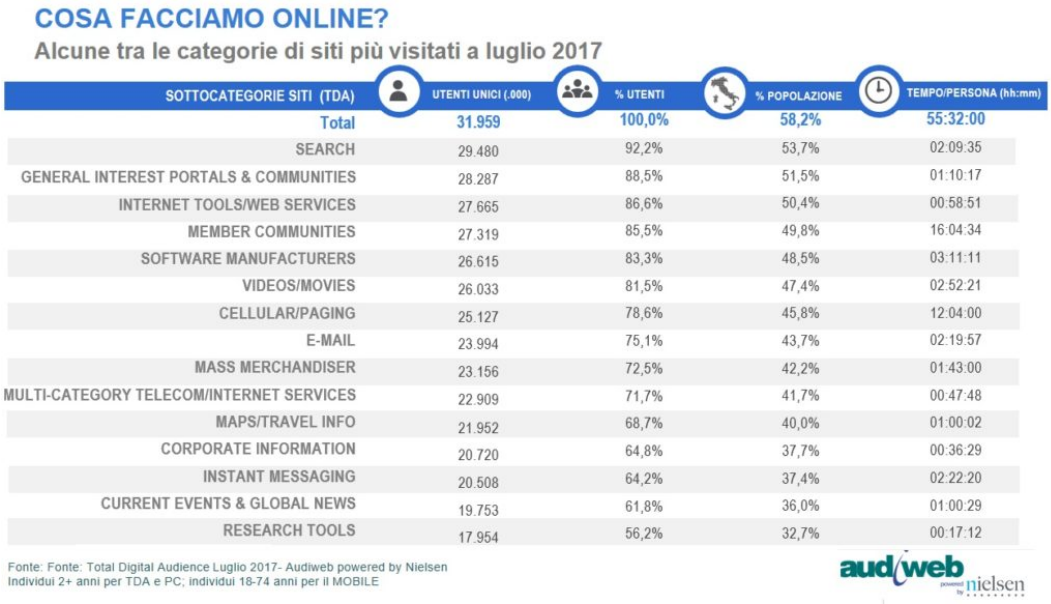


Figura 15 il tempo trascorso online nel giorno medio

In base ai dati il 92,2% degli utenti online nel mese di luglio 2017 ha navigato tra le applicazioni e servizi dedicati alla ricerca di contenuti e servizi online. L'88,5% degli utenti ha consultato portali generalisti. L'86,6% ha utilizzato servizi e strumenti online, l'85,5% degli utenti ha utilizzato Social Network e l'81,5% ha guardato contenuti video.

Per quanto riguarda le news solamente il 61,8% degli utenti ha navigato per cercarle! È al penultimo posto nella tabella. Se il 100% degli utenti corrisponde al 58,2% della popolazione questo significa che solamente il 35,96% della popolazione si è esposto alle news su internet! Numericamente sono 22.520.988 .



COSA FACCIAMO ONLINE?

Alcune tra le categorie di siti più visitati a luglio 2017

SOTTOCATEGORIE SITI (TDA)	UTENTI UNICI (.000)	% UTENTI	% POPOLAZIONE	TEMPO/PERSONA (hh:mm)
Total	31.959	100,0%	58,2%	55:32:00
SEARCH	29.480	92,2%	53,7%	02:09:35
GENERAL INTEREST PORTALS & COMMUNITIES	28.287	88,5%	51,5%	01:10:17
INTERNET TOOLS/WEB SERVICES	27.665	86,6%	50,4%	00:58:51
MEMBER COMMUNITIES	27.319	85,5%	49,8%	16:04:34
SOFTWARE MANUFACTURERS	26.615	83,3%	48,5%	03:11:11
VIDEOS/MOVIES	26.033	81,5%	47,4%	02:52:21
CELLULAR/PAGING	25.127	78,6%	45,8%	12:04:00
E-MAIL	23.994	75,1%	43,7%	02:19:57
MASS MERCHANDISER	23.156	72,5%	42,2%	01:43:00
MULTI-CATEGORY TELECOM/INTERNET SERVICES	22.909	71,7%	41,7%	00:47:48
MAPS/TRAVEL INFO	21.952	68,7%	40,0%	01:00:02
CORPORATE INFORMATION	20.720	64,8%	37,7%	00:36:29
INSTANT MESSAGING	20.508	64,2%	37,4%	02:22:20
CURRENT EVENTS & GLOBAL NEWS	19.753	61,8%	36,0%	01:00:29
RESEARCH TOOLS	17.954	56,2%	32,7%	00:17:12

Fonte: Fonte: Total Digital Audience Luglio 2017- Audiweb powered by Nielsen
Individui 2+ anni per TDA e PC; individui 18-74 anni per il MOBILE

audiweb powered by nielsen

Figura 16 cosa facciamo online

Tra gli altri contenuti di interesse emergono le categorie dedicate all'intrattenimento e al tempo libero, come ad esempio i servizi di messaggistica da mobile (sotto-categoria "Cellular/Paging"), con il 78,6 degli utenti online nel mese, i siti di e-commerce ("Mass merchandiser") con il 72,5% degli utenti, mappe e informazioni di viaggio con il 68,7% e le news ("Current event & global news) con il 61,8% degli utenti.

Passiamo ora all'analisi dei social network. Presenterò prima varie analisi quantitative e poi una analisi qualitativa. Un approccio quantitativo è sempre utile specialmente quando l'analisi riguarda il target potenziale da raggiungere sul canale scelto in una strategia di comunicazione (la quale può essere a fini commerciali o, come abbiamo già visto, a fini politici). Presenterò diverse analisi provenienti da fonti diverse.

Questo è il numero di utenti attivi secondo una analisi di juliusdesign.net[35] . Rispetto agli "utenti registrati", quelli "attivi" sono molto più utili e interessanti: sono infatti quelle persone che utilizzano in modo assiduo la piattaforma Social Media, sono dunque coloro che assiduamente si espongono ai media. Sono dei potenziali gatekeeper.

FACEBOOK	30 Milioni Utenti Attivi	via Vincos
YOUTUBE	24 Milioni Utenti Attivi	via YouTube
TWITTER	6.9 Milioni Utenti Attivi	via Wired Italia
TUMBLR	2.5 Milioni Utenti Attivi	via Yhaoo
SNAPCHAT	2 Milioni Utenti Attivi	via Wired
LINKEDIN	8 Milioni Utenti Attivi	via La Stampa
INSTAGRAM	14 Milioni Utenti Attivi	via Wired Italia
GOOGLE PLUS	7.3 Milioni Utenti Attivi	via GlobalWebIndex
PINTEREST	4.7 Milioni Utenti Attivi	via PinterestItaly

Figura 17 numero di utenti attivi

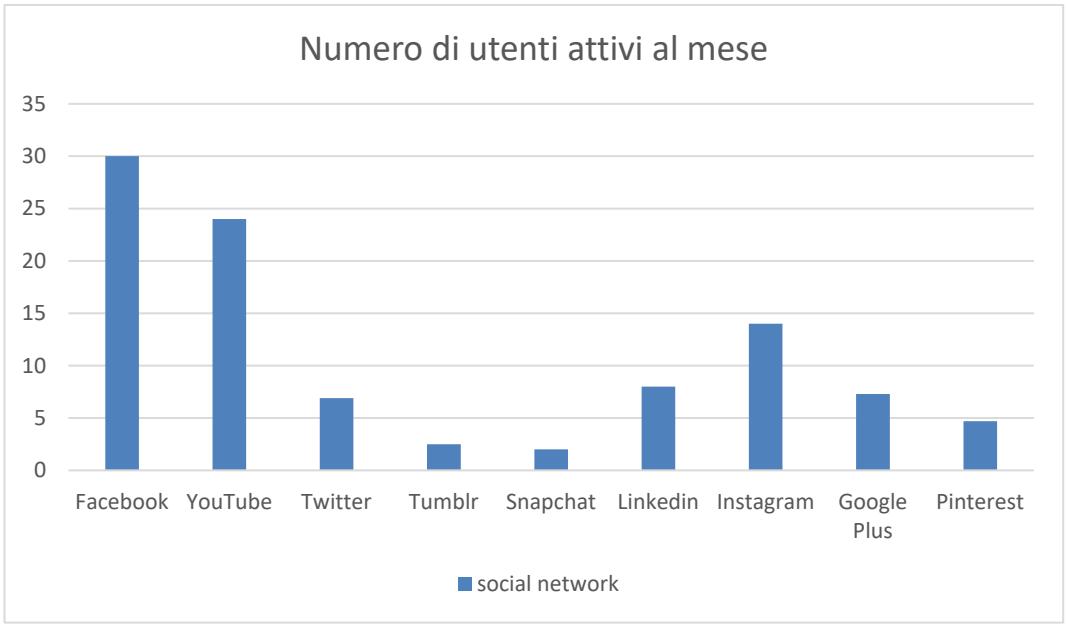


Figura 18 numero di utenti attivi al mese

Analizziamo ora il report Digital in 2017 nato dalla collaborazione tra We Are Social e Hootsuite.

[35] *http://www.juliusdesign.net/28700/lo-stato-degli-utenti-attivi-e-registrati-sui-social-media-in-italia-e-mondo-2015/*

Figura 19 il digitale in Italia

Il tasso di penetrazione per quanto riguarda il numero di utenti internet è più alto rispetto a quello fornito da audiweb, 66% contro 58,2%. Per quanto riguarda il numero di utenti attivi sui social media invece le percentuali sono simili. Anche le percentuali riguardanti i dispositivi mobili sono simili. Possiamo quindi dire con sicurezza che gli italiani si connettono sempre di più e sempre di più da dispositivi mobili.

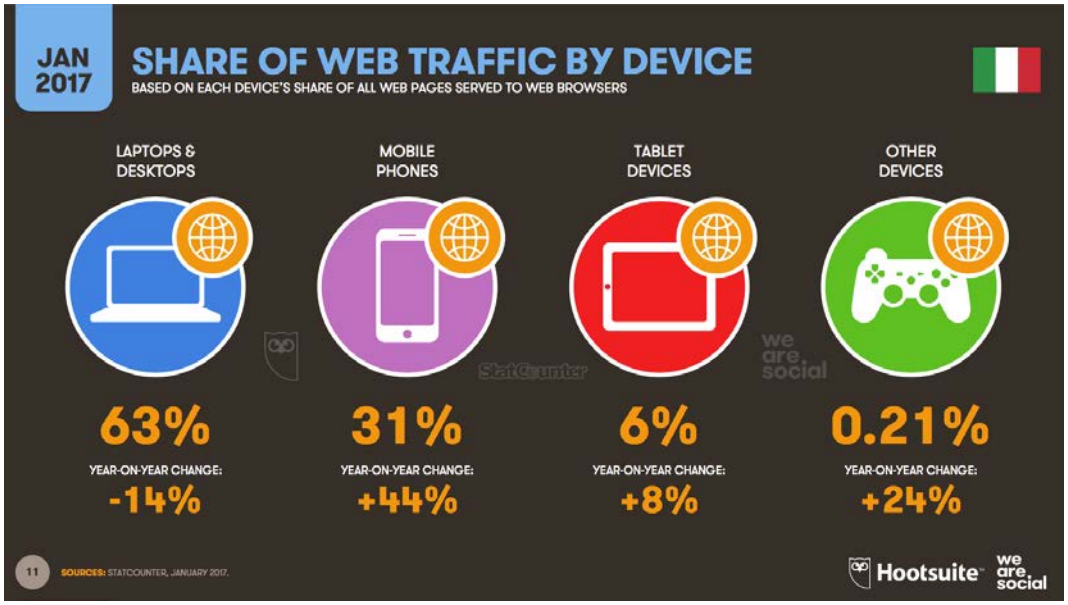


Figura 20 provenienza del traffico

Qui notiamo il calo significativo nel traffico generato da PC e nell'aumento vertiginoso del traffico generato da dispositivi mobili. Come detto in precedenza, dispositivi diversi portano a stili di fruizione diversi che portano a favorire tipi di contenuti e formati diversi.

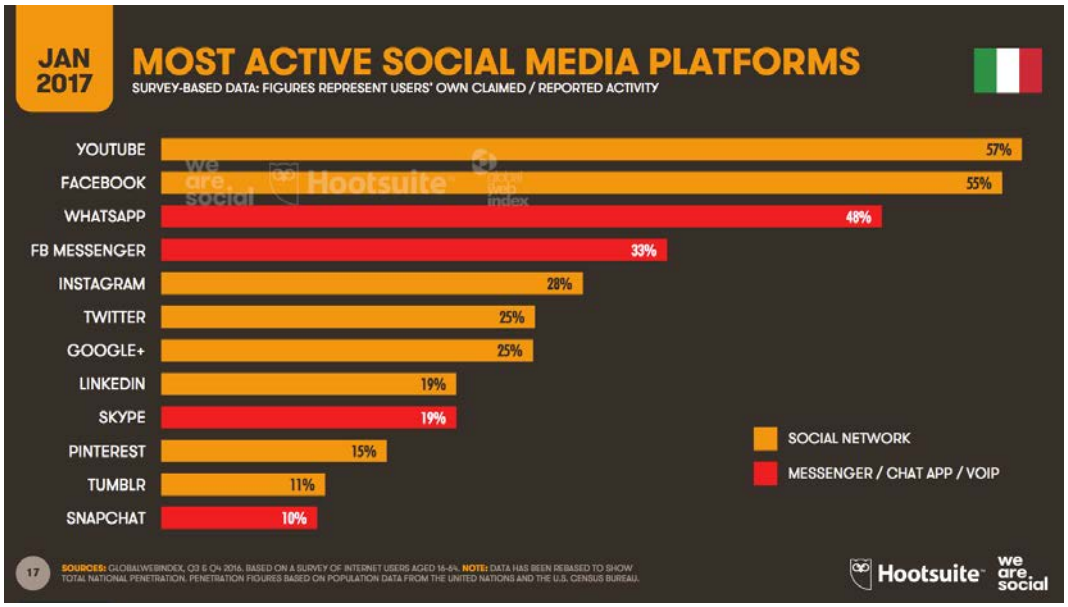


Figura 21 i social media più attivi

Il 13° Rapporto Censis-Ucsi sulla comunicazione pubblicato nel 2016 ci fornisce altri dati importantissimi[36]. Secondo il rapporto bel il 73,7% degli italiani sul web, il livello di penetrazione è quindi superiore rispetto a quello stimato da Hootsuite. "Social network e piattaforme online indispensabili nella nostra vita quotidiana. Facebook è il social network più popolare: è usato dal 56,2% degli italiani (il 44,3% nel 2013), raggiunge l'89,4% di utenza tra i giovani under 30 e il 72,8% tra le persone più istruite, diplomate e laureate. L'utenza di YouTube è passata dal 38,7% del 2013 al 46,8% del 2016 (fino al 73,9% tra i giovani). Instagram è salito dal 4,3% di utenti del 2013 al 16,8% del 2016 (e il 39,6% dei giovani). E WhatsApp ha conosciuto un vero e proprio boom: nel 2016 è usato dal 61,3% degli italiani (l'89,4% dei giovani)."[37] Utilissima è l'analisi fatta riguardante il rapporto tra nuovi media e sfiducia nei confronti della classe dirigente: "I media digitali tra élite e popolo. Le ultime tendenze indicano che gli strumenti della disintermediazione digitale si stanno infilando come cunei nel solco di divaricazione scavato tra élite e popolo, prestandosi all'opera di decostruzione delle diverse forme di autorità costituite, fino a sfociare nelle mutevoli forme del populismo che si stanno diffondendo rapidamente in Italia e in Occidente. Si tratta di una sfiducia nelle classi dirigenti al potere e in istituzioni di lunga durata che oggi si salda alla fede nel potenziale di emancipazione delle comunità attribuito ai processi di disintermediazione resi possibili dalla rete. Si sta così radicando un nuovo mito fondativo della cultura web: la convinzione che il lifelogging, i dispositivi di self-tracking e i servizi di social networking potranno fornire risposte ai bisogni della collettività più efficaci, veloci, trasparenti ed economiche di quanto finora sia stato fatto." Importante è anche l'analisi riguardante il rapporto fra anziani e social media: "La frattura generazionale: giovani e anziani sempre più lontani. Le distanze tra i consumi mediatici giovanili e

[36] *http://www.censis.it/7?shadow_comunicato_stampa=121073*
[37] *idem*

quelli degli anziani continuano ad essere rilevantissime. Tra i giovani under 30 la quota di utenti della rete arriva al 95,9%, mentre è ferma al 31,3% tra gli over 65 anni. L'89,4% dei primi usa telefoni smartphone, ma lo fa solo il 16,2% dei secondi. L'89,3% dei giovani è iscritto a Facebook, contro appena il 16,3% degli anziani. Il 73,9% dei giovani usa YouTube, come fa solo l'11,2% degli ultrasessantacinquenni. Oltre la metà dei giovani (il 54,7%) consulta i siti web di informazione, contro appena un anziano su dieci (il 13,8%). Il 37,3% dei primi ascolta la radio attraverso il telefono cellulare, mentre lo fa solo l'1,2% dei secondi. E se un giovane su tre (il 36,3%) ha già un tablet, solo il 7,7% degli anziani lo usa. Su Twitter poi c'è un quarto dei giovani (il 24%) e un marginale 1,7% degli over 65."

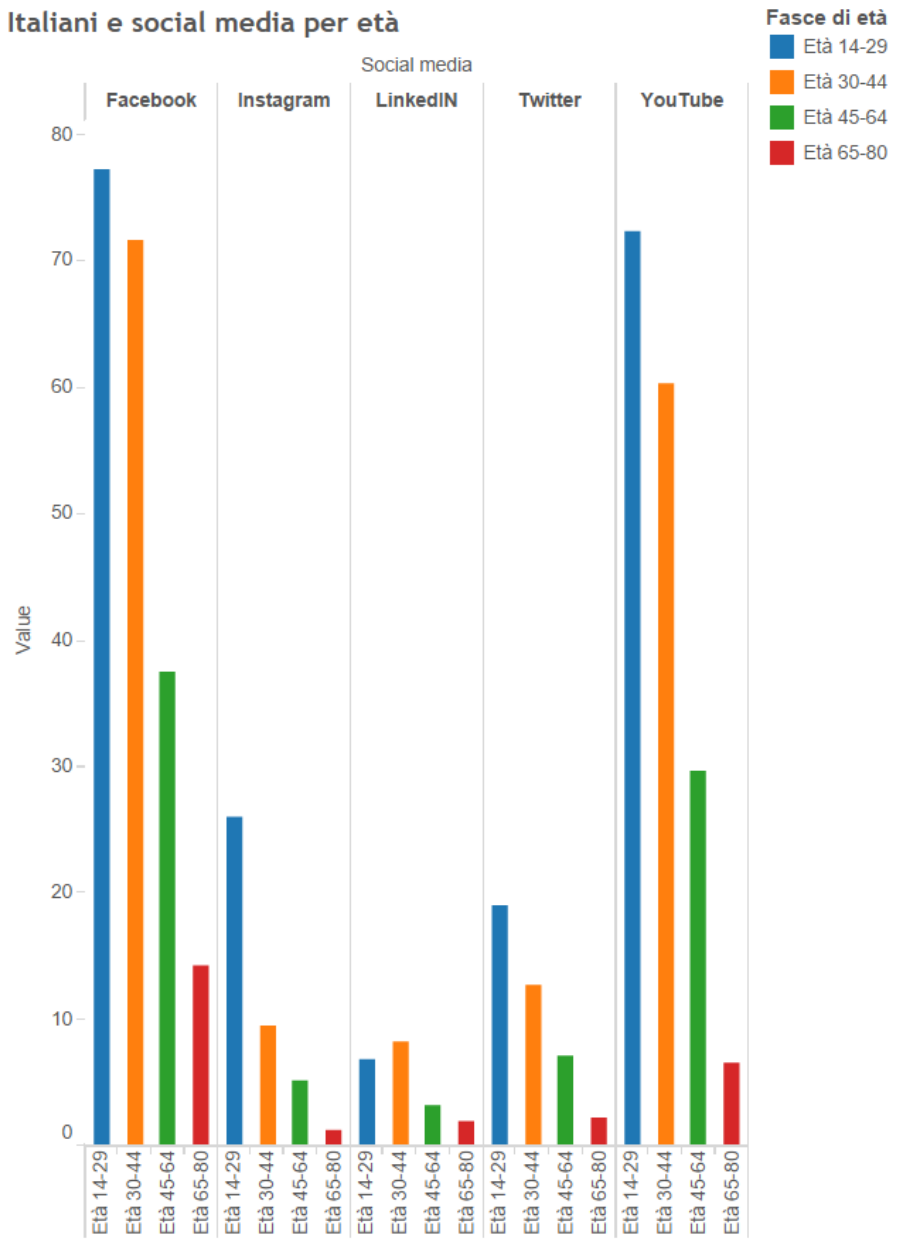


Figura 22 italiani e social media per fascia di età

Trovo utile analizzare il numero di like e follower dei politici su Facebook e Twitter[39].

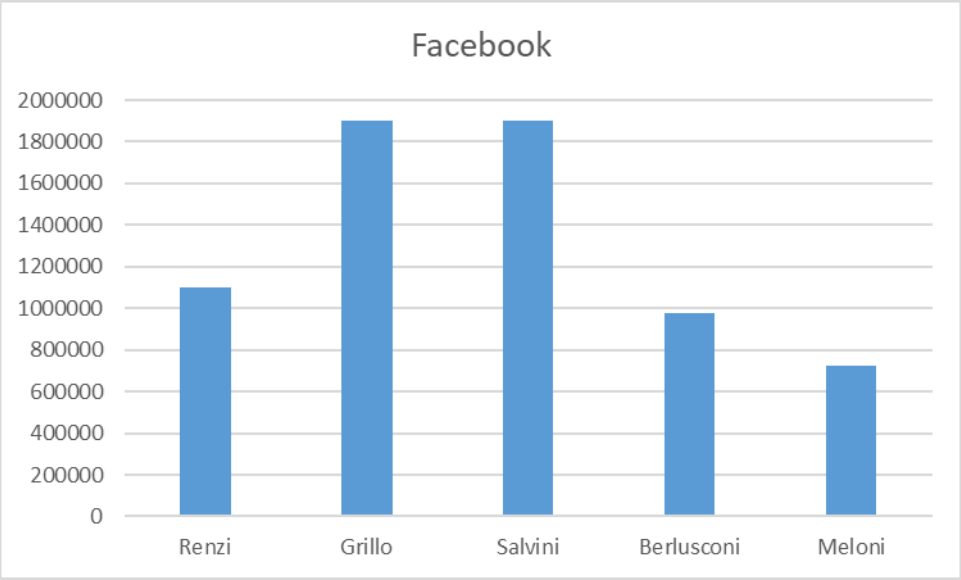


Figura 23 i like su Facebook

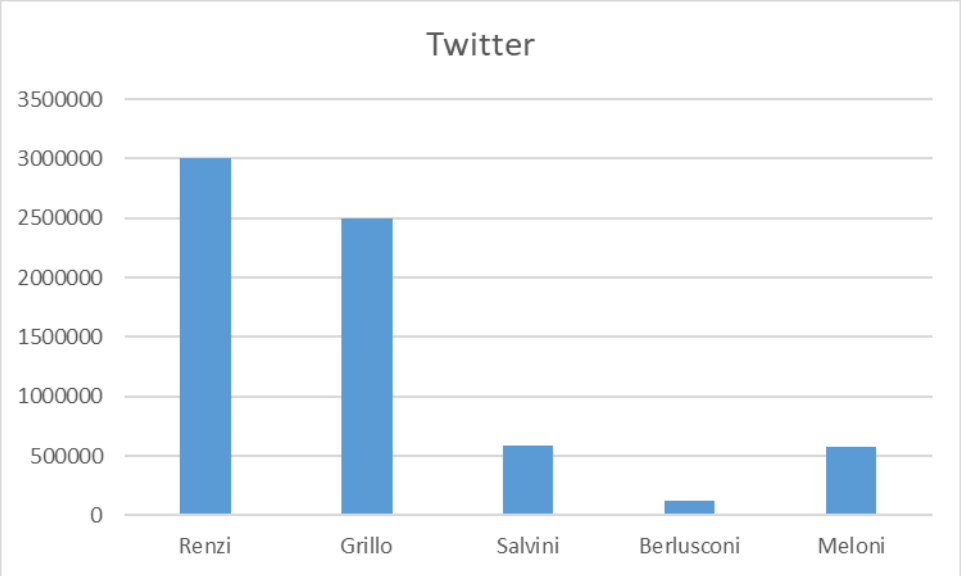


Figura 24 i follower su Twitter

Sembrerebbe che utenti con ideologie politiche diverse preferiscono piattaforme diverse, con la sinistra che favorisce decisamente Twitter e la destra che favorisce Facebook. Il movimento 5 stelle ha un elettorato estremamente eterogeneo, per questo in entrambi i casi ha un ampio numero di follower e di like.

---

[38] https://www.wired.it/internet/social-network/2016/03/08/italiani-social-media/
[39] http://www.ilsole24ore.com/art/notizie/2017-09-28/su-facebook-testa-testa-grillo-e-salvini-doppiato-renzi-che-si-rifa-twitter-091110.shtml

In generale Twitter sembra una piattaforma più orientata a sinistra. Guardiamo i 20 account italiani più seguiti su Twitter nel 2015 e poi nel 2016.

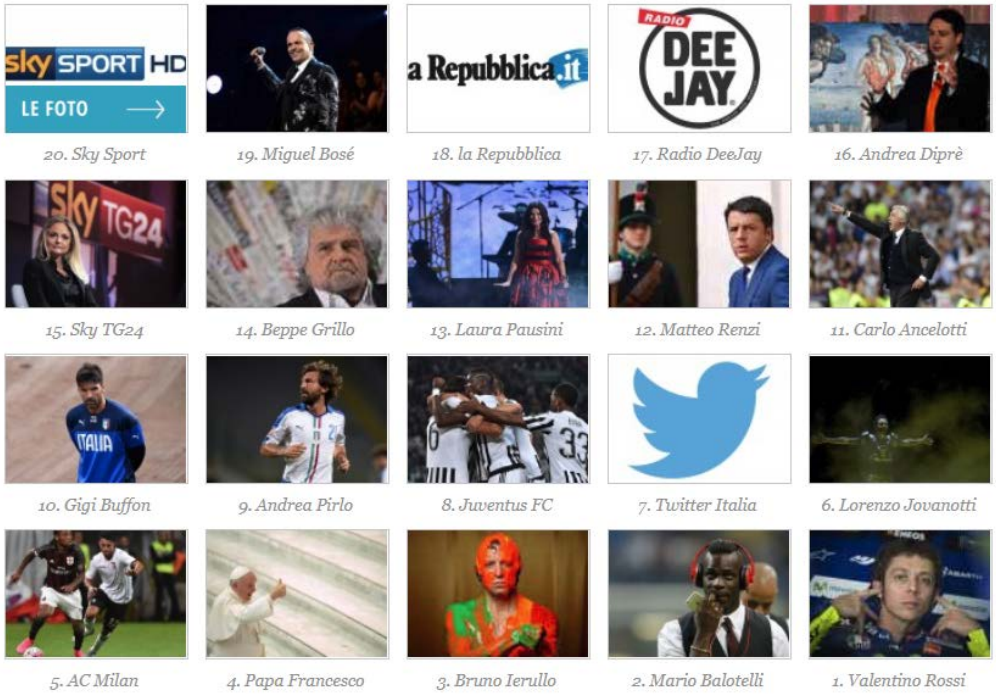


Figura 25 i 20 account più seguiti su Twitter in Italia nel 2015

40

Come politici abbiamo solamente Renzi e Grillo e come giornale solamente la Repubblica. Guardiamo cosa succede nel 2016.

[40] *http://www.ilpost.it/2015/10/09/account-italiani-piu-seguiti-su-twitter/*

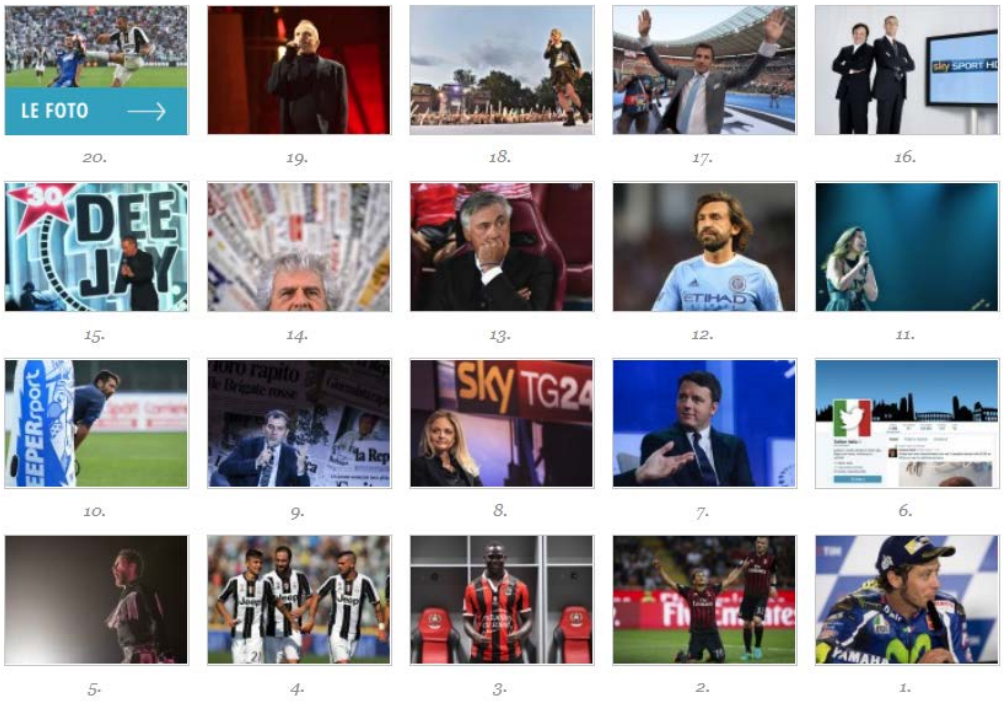


Figura 26 i 20 account più seguiti su Twitter in Italia nel 2016

41

Renzi è ora addirittura al settimo posto mentre Grillo è al quattordicesimo. È impressionante il numero di follower di Renzi su Twitter alla luce di quanto tale social sia meno popolare di Facebook in Italia.

Passiamo ora ad una analisi di tipo qualitativo. Nel 2017 Blogmeter, una società italiana che si occupa di social media intelligence, utilizzando un campione di 1501 residenti italiani di età compresa fra i 15 e i 64 anni, ha tentato di scoprire "perché gli italiani usano i social media e quali sono i loro impieghi nella vita di tutti i giorni". Che relazione hanno i social media con le relazioni personali, con gli acquisti, con l'informazione? A chi crediamo? A chi dedichiamo più tempo?

[41] *http://www.ilpost.it/2016/10/10/account-italiani-piu-seguiti-twitter-2/*

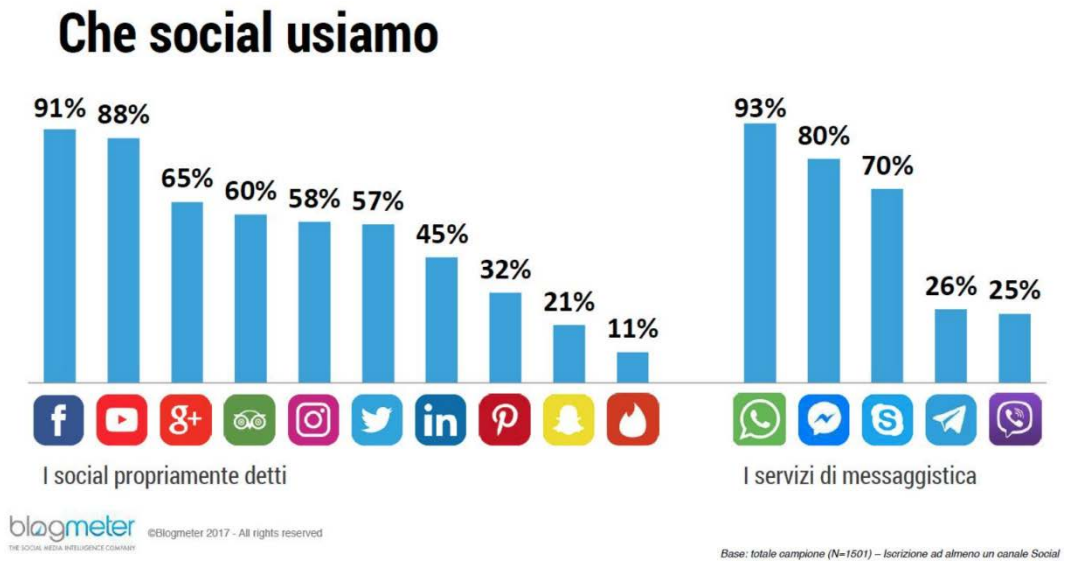


Figura 27 i social network usati dagli italiani

Analizzando le modalità con cui vengono utilizzati i vari canali lo studio fa una importante distinzione fra social di cittadinanza e social funzionali. "Della prima categoria fanno parte quei social che usiamo tutti i giorni, anche più volte al giorno, e più volte a settimana, che in un certo senso definiscono la nostra identità online" ha spiegato Alberto Stracuzzi, customer intelligence director di BlogMeter. "Facebook è il maggiore rappresentate: ben l'84% degli intervistati ha dichiarato di utilizzarlo più volte al giorno; gli altri sono YouTube, Instagram e Whatsapp"[42].

Figura 28 social di cittadinanza e social funzionali

[42] *https://www.youtube.com/watch?v=CTvzvyy3EIk*

Per social funzionali invece si intendono quei canali che vengono utilizzati per soddisfare un bisogno o un interesse specifico. I principali sono Google Plus, Twitter e Linkedin, che rispettivamente il 40%, il 35% e il 31% dei 1501 intervistati afferma di usare saltuariamente. C'è anche TripAdvisor, consultato per scegliere ristoranti o locali. Questo diverso approccio influenza anche l'atteggiamento e il posizionamento delle aziende sui social. "Stare su un social di cittadinanza è faticoso, con investimenti, per avere una presenza continuativa, con il rischio anche di essere asfissiante. Al contrario su un social funzionale come TripAdvisor, l'importante è saper rispondere alle domande che un utente può porre connettendosi una volta a settimana".

Il 6-7% dice di non poter più fare a meno dei social e il 4% degli intervistati pensa che sia inevitabile iscriversi. Tuttavia stando alla ricerca gli italiani si fanno problemi a cancellarsi da quelli che non apprezzano. Il social più abbandonato in assoluto è Tinder, con ben 3,5 italiani su 10 che hanno dichiarato di essersi iscritti e poi cancellati. Seguono Snapchat, con il 25%, Pinterest e Twitter, con il 10%.

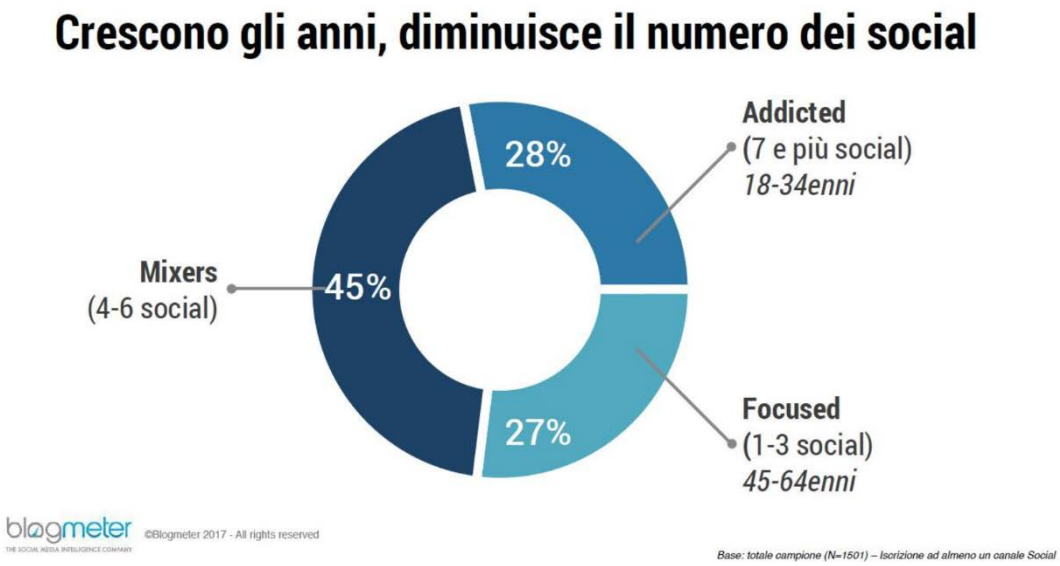


Figura 29 il numero dei social network usati in base all'età

Con l'aumentare dell'età diminuisce il numero di social a cui si è iscritti: nella fascia di età compresa tra i 18 e i 34 anni, la media di social e servizi di messaggistica posseduti è superiore a sette. Dopo i 45 anni, tuttavia, scende a tre canali.

Instagram e YouTube sono i canali su cui gli utenti più giovane, quelli nella fascia di età compresa tra i 15 e i 17 anni, dichiarano di passare più tempo. All'aumentare dell'età subentrano poi Facebook (18-24) e, dagli over 35 anni in su, anche tv e giornali.

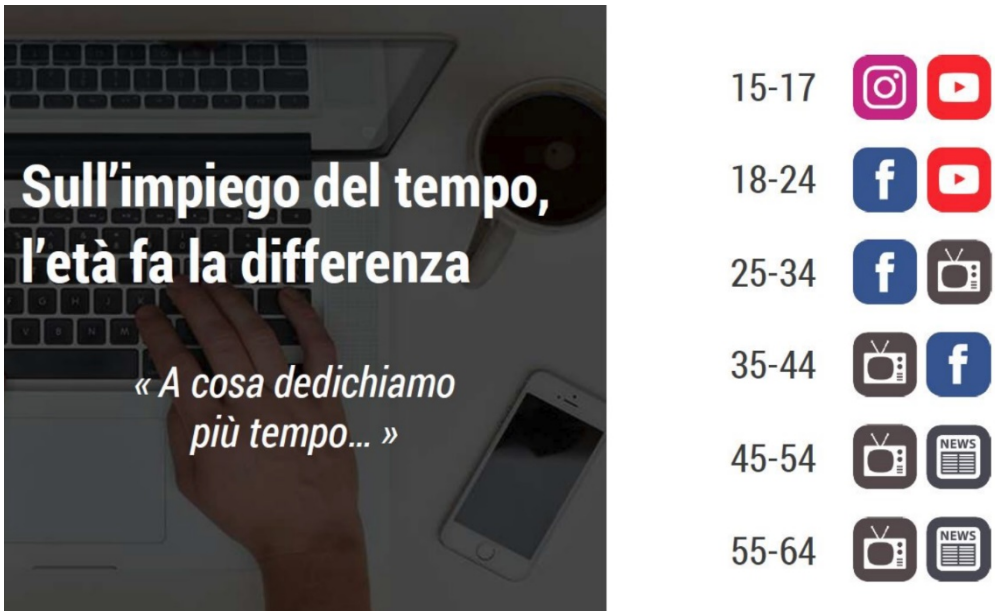


Figura 30 a cosa gli italiani dedicano più tempo

Ma cosa spinge ad utilizzare i social? Tra le motivazioni la più gettonata è la curiosità e l'interesse (21%), seguita poi dal desiderio di creazione di relazioni nuove e personali (17%), mentre il 14% afferma di utilizzarli per svago o piacere. Quali sono le ragioni che spingono ad usare un social piuttosto di un altro? Facebook è il più versatile, il più adatto a rispondere a quasi tutte le esigenze (fatta eccezione forse per le ricerche di lavoro). TripAdvisor è utile per leggere recensioni, YouTube per informarsi, mentre per seguire brand e personaggi celebri gli intervistati preferiscono Instragram.

Figura 31 social network diversi per attività diverse

Canali di comunicazione più tradizionali come la televisione e i magazine continuano a mantenere una forte credibilità anche tra gli utenti del web che ritengono poco affidabili Facebook, YouTube e i blog. "Un dato questo chemesso anche in relazione al tema delle fake news, dimostra come gli utenti se hanno bisogno di

credibilità si rivolgono ad altre fonti". È quindi un errore considerare gli utenti dei social dei "creduloni. Il problema non sorge quando una news circola sui social, ma quando a rilanciarla sono le testate ritenute credibili".

Figura 32 di quali media si fidano gli italiani

Quando invece si tratta di fare compere online i canali digitali – tra i siti di ecommerce e quelli di recensioni – tornano ad essere ritenuti attendibili.

Figura 33 dove si informano gli italiani prima di comprare

Nell'ultima parte della ricerca viene dato anche spazio a celebrities e influencer. Cantanti, giornalisti e scrittori sono i personaggi di cui ci si fida di più, anche se i più seguiti restano musicisti e personaggi televisivi (33%). Tra i giornalisti popolari sui social abbiamo: Beppe Severgnini, Alberto Angela, Giordano Bruno Guerri e Selvaggia Lucarelli

Dall’analisi, emerge anche che il rapporto con gli influencer è però complesso e sfaccettato: se fan-base e credibilità sono aspetti non sempre correlati, età e numero di influencer seguiti sì. I giovani sembrano seguire infatti un numero maggiore di personaggi appartenenti a categorie diverse, mentre invecchiando si diventa più selettivi.

Capitolo 2: la cluster analysis

In italiano la parola "cluster" viene tradotta come "grappolo". Essa fu utilizzata per la prima volta dallo psicologo e statistico statunitense Robert Choate Tryon[43] nell'ambito dei suoi lavori in psicometria. La scelta del termine deriva dal fatto che lo scopo della cluster è quello di "raggruppare le unità di classificazione (in questo caso le forme grafiche) in classi tali che la variabilità interna, cioè fra i soggetti dello stesso gruppo, sia la minima possibile, mentre quella esterna tra i gruppi sia la massima possibile"[44].

Gli oggetti raggruppati (campioni, misurazioni, eventi, pattern) sono solitamente rappresentati come punti (vettori) in uno spazio multidimensionale dove ogni dimensione rappresenta un distinto attributo (variabile, misurazione) descrivente tale oggetto. Per semplicità normalmente si presume che i valori siano presenti per tutti gli attributi.

2.1 la matrice dei dati

terIl primo passo di un'analisi di aggregazione è costruire una matrice dati nella quale raccogliamo le misurazioni di *p* caratteri effettuate su *n* unità statistiche.

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x_{11}} & \boldsymbol{x_{12}} & \cdots & \boldsymbol{x_{1k}} & \cdots & \boldsymbol{x_{1p}} \\ \boldsymbol{x_{21}} & \boldsymbol{x_{22}} & \cdots & \boldsymbol{x_{2k}} & \cdots & \boldsymbol{x_{2p}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{x_{i1}} & \boldsymbol{x_{i2}} & \cdots & \boldsymbol{x_{ik}} & \cdots & \boldsymbol{x_{ip}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{x_{n1}} & \boldsymbol{x_{n2}} & \cdots & \boldsymbol{x_{nk}} & \cdots & \boldsymbol{x_{np}} \end{pmatrix}$$

Il termine generico x_{ik} indica la k-esima variabile misurata sull'unità *i*. Di solito l'indice i contrassegna un individuo o un prodotto, mentre l'indice *j* contrassegna un attributo di i. In alternativa si può rappresentare la matrice dati come una matrice a blocchi il cui blocco generico è rappresentato da un vettore riga relativo all'unità *i* di dimensione 1 x *p*, x'_i

2.2 Le misure di distanza

Una volta costruita la matrice è finalmente possibile calcolare la distanza tra i vettori che rappresentano le *n* unità statistiche. Ogni unità viene confrontata con le altre per valutare e quantificare il grado di similarità\dissimilarità rispetto alle *p* variabili di rilevazione. Il modo in cui viene calcolata la distanza è fondamentale, è ciò che rende diversi i vari metodi di clustering. Qualora le variabili non abbiano la stessa

---

[43] Tryon, Robert C. (1939). Cluster Analysis: Correlation Profile and Orthometric (factor) Analysis for the Isolation of Unities in Mind and Personality. Edwards Brothers.
[44] AMATURO E., PUNZIANO G., Content Analysis: tra comunicazione e politica, Ledizioni, Milano, 2013 p. 183.

unità di misura è opportuno standardizzarle, ossia fare in modo che tutte abbiano la stessa media e la stessa varianza (rispettivamente 0 e 1). In questo modo le variabili sono confrontabili.

Il procedimento è il seguente: a ciascuna osservazione viene sottratta la media delle osservazioni; il risultato viene poi diviso per la deviazione standard. In formula:

$$\boldsymbol{z_{ik} = \frac{x_{ik} - \bar{x}_k}{\sigma_k} \; per \; i = 1, 2, \dots, n}$$

Dove con $\bar{x}_k$ indichiamo la media e con σ_k la varianza delle osservazioni relative alla variabile *k*.

Una volta standardizzate le variabili si può procedere al calcolo della distanza tra i vettori. Dati due vettori, i e j, entrambi di dimensioni 1xp, una misura di distanza deve godere delle seguenti proprietà:

1) $\boldsymbol{d_{ij} \geq 0}$ (non negatività)
2) $\boldsymbol{d_{ii} = 0}$ e analogamente $\boldsymbol{d_{jj} = 0}$ (identità)
3) $\boldsymbol{d_{ij} = d_{ji}}$ (simmetria)
4) $\boldsymbol{d_{ij} \leq d_{ir} + d_{rj}}$ (disuguaglianza triangolare)

Come accennato in precedenza esistono diverse misure di distanza. Per la misurazione di caratteri quantitativi abbiamo: distanza euclidea, distanza della città a blocchi (Manhattan), distanza di Lagrange, distanza di Canberra. Per i caratteri qualitativi abbiamo la distanza di Jaccard.

- Distanza euclidea: Un metodo di misurazione che può essere immaginato in termini geometrici come la distanza in linea retta fra due punti. Quindi dati due vettori, *i* e *j*, essa può essere definita come la norma della loro differenza:

$$\boldsymbol{d_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2}}$$

Altro non è che un'applicazione del teorema di Pitagora.

- Distanza della città a blocchi o distanza di Manhattan.
  In questo caso invece dell'ipotenusa calcoliamo la lunghezza dei due cateti. Il nome deriva infatti dal fatto che questa è la distanza che bisogna percorrere per andare da un punto *i* a un punto *j* quando è consentito muoversi solo in direzioni parallele agli assi (come avviene in una città divisa in blocchi con strade che si intersecano ad angolo retto).

$$\boldsymbol{d_{ij} = \sum_{k=1}^{p} |x_{ik} - x_{jk}|}$$

- Distanza di Minkowski

$$\boldsymbol{d_{ij} = \left[\sum_{k=1}^{p} |x_{ik} - x_{jk}|^{\lambda}\right]^{1/\lambda} \quad con\ \lambda > 0}$$

  È una generalizzazione delle varie distanze. Se $\boldsymbol{\lambda}$ = 1 avremo la distanza di Manhattan; se $\boldsymbol{\lambda}$ = 2 avremo la distanza euclidea; se $\boldsymbol{\lambda}$ si avvicina ad ∞ avremo la distanza di Chebychev (Lagrange).
  In formula la distanza di Lagrange è

$$\boldsymbol{d_{ij} = max_k\{|x_{ik} - x_{jk}|\}}$$

- Distanza di Canberra:
  è una versione ponderata della distanza di Manhattan

$$\boldsymbol{d_{ij} = \sum_{k=1}^{p} \frac{|x_{ik} - x_{jk}|}{|x_{ik}| + |x_{jk}|}}$$

- Coefficiente di similarità di Jaccard:
  misura la similarità tra unità su cui siano osservate *p* variabili qualitative binarie, ed è definito mediante le concordanze e le discordanze degli attributi nelle unità.

		Unità *i*		
		Presente (1)	Assente (0)	Totale
Unità *j*	Presente (1)	$\boldsymbol{M_{00}}$	$\boldsymbol{M_{10}}$	$\boldsymbol{M_{00}+M_{10}}$
	Assente (0)	$\boldsymbol{M_{01}}$	$\boldsymbol{M_{11}}$	$\boldsymbol{M_{01}+M_{11}}$
	Totale	$\boldsymbol{M_{00}+M_{01}}$	$\boldsymbol{M_{10}+M_{11}}$	*P*

  Il coefficiente di similarità di Jaccard è dato da:

$$J = \frac{\boldsymbol{M_{11}}}{\boldsymbol{M_{01}+M_{10}+M_{11}}}$$

La distanza di Jaccard è quindi data da:

$\boldsymbol{d_j} = 1 - J$

Tali distanze vanno a formare la matrice delle distanze D, una matrice di dimensioni n x n (in quanto per ogni unità viene calcolata la distanza rispetto alle altre). Essa è simmetrica in quanto la distanza dell'unità 1 dall'unità 2 è uguale alla distanza dell'unità 2 dall'unità 1, ha valori nulli lungo la diagonale principale (in quanto la distanza di una unità da sé stessa è sempre pari a zero).

$$\boldsymbol{D} = \begin{pmatrix} \mathbf{0} & \boldsymbol{d_{12}} & \cdots & \cdots & \boldsymbol{d_{1n}} \\ & \mathbf{0} & & & \boldsymbol{d_{2n}} \\ & & \ddots & & \vdots \\ & & & \ddots & \boldsymbol{d_{n-1,n}} \\ & & & & \mathbf{0} \end{pmatrix}$$

Riportiamo un esempio dei concetti appena esposti. Utilizzeremo il dataset USarrests, ossia *Violent Crime Rates by US State*[45]. Per comodità in questo esempio mi riferirò solamente ai primi sei Stati in ordine alfabetico.

	Murder	*Assault*	*UrbanPop*	*Rape*
Alabama	*13.2*	*236*	*58*	*21.2*
Alaska	*10*	*263*	*48*	*44.5*
Arizona	*8.1*	*294*	*80*	*31*
Arkansas	*8.8*	*190*	*50*	*19.5*
California	*9*	*276*	*91*	*40.6*
Colorado	*7.9*	*204*	*78*	*38.7*

	Murder	*Assault*	*UrbanPop*	*Rape*
Alabama	*13.2*	*236*	*58*	*21.2*
Alaska	*10*	*263*	*48*	*44.5*
Arizona	*8.1*	*294*	*80*	*31*
Arkansas	*8.8*	*190*	*50*	*19.5*
California	*9*	*276*	*91*	*40.6*

[45] *World Almanac and Book of facts 1975. (Crime rates).*

Statistical Abstracts of the United States 1975, p.20, (Urban rates)
McNeil, D. R. (1977) Interactive Data Analysis. New York: Wiley.

Colorado	*7.9*	*204*	*78*	*38.7*
Connecticut	*3.3*	*110*	*77*	*11.1*
Delaware	*5.9*	*238*	*72*	*15.8*
Florida	*15.4*	*335*	*80*	*31.9*
Georgia	*17.4*	*211*	*60*	*25.8*
Hawaii	*5.3*	*46*	*83*	*20.2*
Idaho	*2.6*	*120*	*54*	*14.2*
Illinois	*10.4*	*249*	*83*	*24*
Indiana	*7.2*	*113*	*65*	*21*
Iowa	*2.2*	*56*	*57*	*11.3*
Kansas	*6*	*115*	*66*	*18*
Kentucky	*9.7*	*109*	*52*	*16.3*
Louisiana	*15.4*	*249*	*66*	*22.2*
Maine	*2.1*	*83*	*51*	*7.8*
Maryland	*11.3*	*300*	*67*	*27.8*
Massachusetts	*4.4*	*149*	*85*	*16.3*
Michigan	*12.1*	*255*	*74*	*35.1*
Minnesota	*2.7*	*72*	*66*	*14.9*
Mississippi	*16.1*	*259*	*44*	*17.1*
Missouri	*9*	*178*	*70*	*28.2*
Montana	*6*	*109*	*53*	*16.4*
Nebraska	*4.3*	*102*	*62*	*16.5*
Nevada	*12.2*	*252*	*81*	*46*
New Hampshire	*2.1*	*57*	*56*	*9.5*
New Jersey	*7.4*	*159*	*89*	*18.8*
New Mexico	*11.4*	*285*	*70*	*32.1*
New York	*11.1*	*254*	*86*	*26.1*
North Carolina	*13*	*337*	*45*	*16.1*
North Dakota	*0.8*	*45*	*44*	*7.3*
Ohio	*7.3*	*120*	*75*	*21.4*
Oklahoma	*6.6*	*151*	*68*	*20*
Oregon	*4.9*	*159*	*67*	*29.3*
Pennsylvania	*6.3*	*106*	*72*	*14.9*
Rhode Island	*3.4*	*174*	*87*	*8.3*

South Carolina	*14.4*	*279*	*48*	*22.5*
South Dakota	*3.8*	*86*	*45*	*12.8*
Tennessee	*13.2*	*188*	*59*	*26.9*
Texas	*12.7*	*201*	*80*	*25.5*
Utah	*3.2*	*120*	*80*	*22.9*
Vermont	*2.2*	*48*	*32*	*11.2*
Virginia	*8.5*	*156*	*63*	*20.7*
Washington	*4*	*145*	*73*	*26.2*
West Virginia	*5.7*	*81*	*39*	*9.3*
Wisconsin	*2.6*	*53*	*66*	*10.8*
Wyoming	*6.8*	*161*	*60*	*15.6*

Standardizziamo le variabili.

	Murder	*Assault*	*UrbanPop*	*Rape*
Alabama	*1.46792469*	*1.231923*	*0.8558452*	*0.9058448*
Alaska	*1.11206416*	*1.3728633*	*0.7082857*	*1.9014196*
Arizona	*0.90077197*	*1.5346837*	*1.1804761*	*1.3245844*
Arkansas	*0.97861646*	*0.9918024*	*0.7377976*	*0.8332063*
California	*1.00085774*	*1.4407235*	*1.3427916*	*1.7347783*
Colorado	*0.87853068*	*1.0648826*	*1.1509642*	*1.6535941*

	Murder	Assault	UrbanPop	Rape
Alabama	1.46792469	1.231923	0.8558452	0.9058448
Alaska	1.11206416	1.3728633	0.7082857	1.9014196
Arizona	0.90077197	1.5346837	1.1804761	1.3245844
Arkansas	0.97861646	0.9918024	0.7377976	0.8332063
California	1.00085774	1.4407235	1.3427916	1.7347783
Colorado	0.87853068	1.0648826	1.1509642	1.6535941
Connecticut	0.36698117	0.5742014	1.1362083	0.4742867
Delaware	0.65611785	1.242363	1.0624285	0.6751108
Florida	1.7125788	1.7487042	1.1804761	1.3630401
Georgia	1.93499163	1.1014227	0.8853571	1.1023961
Hawaii	0.589394	0.2401206	1.224744	0.8631163
Idaho	0.28913668	0.6264015	0.7968214	0.6067451
Illinois	1.15654672	1.2997832	1.224744	1.0254847

Indiana	0.80068619	0.5898614	0.9591368	0.8972991
Iowa	0.24465411	0.2923207	0.8410892	0.4828324
Kansas	0.66723849	0.6003015	0.9738928	0.7691135
Kentucky	1.07870223	0.5689814	0.7673095	0.696475
Louisiana	1.7125788	1.2997832	0.9738928	0.9485734
Maine	0.23353347	0.4332611	0.7525535	0.3332825
Maryland	1.2566325	1.5660038	0.9886487	1.1878531
Massachusetts	0.48930823	0.7777819	1.2542559	0.696475
Michigan	1.34559763	1.3311032	1.0919404	1.4997714
Minnesota	0.30025732	0.3758409	0.9738928	0.6366551
Mississippi	1.79042329	1.3519833	0.6492619	0.7306579
Missouri	1.00085774	0.9291623	1.0329166	1.2049445
Montana	0.66723849	0.5689814	0.7820654	0.7007479
Nebraska	0.47818759	0.5324413	0.914869	0.7050207
Nevada	1.35671827	1.3154432	1.1952321	1.9655124
New Hampshire	0.23353347	0.2975407	0.8263333	0.405921
New Jersey	0.82292748	0.829982	1.3132797	0.8032964
New Mexico	1.26775314	1.4877036	1.0329166	1.3715858
New York	1.23439121	1.3258832	1.2690118	1.1152146
North Carolina	1.4456834	1.7591443	0.6640178	0.6879293
North Dakota	0.08896513	0.2349006	0.6492619	0.3119183
Ohio	0.81180683	0.6264015	1.1066964	0.9143905
Oklahoma	0.73396234	0.7882219	1.0034047	0.8545706
Oregon	0.54491144	0.829982	0.9886487	1.2519459
Pennsylvania	0.70060042	0.5533213	1.0624285	0.6366551
Rhode Island	0.37810181	0.9082822	1.2837678	0.3546468
South Carolina	1.60137239	1.4563835	0.7082857	0.9613919
South Dakota	0.42258438	0.4489211	0.6640178	0.5469252
Tennessee	1.46792469	0.9813624	0.8706011	1.1493975
Texas	1.41232148	1.0492225	1.1804761	1.0895775
Utah	0.35586053	0.6264015	1.1804761	0.9784833
Vermont	0.24465411	0.2505606	0.4721904	0.4785595
Virginia	0.94525453	0.814322	0.9296249	0.8844806
Washington	0.44482566	0.7569018	1.0771845	1.1194875

West Virginia	0.63387657	0.422821	0.5754821	0.3973753
Wisconsin	0.28913668	0.2766607	0.9738928	0.4614681
Wyoming	0.75620363	0.840422	0.8853571	0.6665651

	Murder	***Assault***	***UrbanPop***	***Rape***
Alabama	***1.46792469***	***1.231923***	***0.8558452***	***0.9058448***
Alaska	***1.11206416***	***1.3728633***	***0.7082857***	***1.9014196***
Arizona	***0.90077197***	***1.5346837***	***1.1804761***	***1.3245844***
Arkansas	***0.97861646***	***0.9918024***	***0.7377976***	***0.8332063***
California	***1.00085774***	***1.4407235***	***1.3427916***	***1.7347783***
Colorado	***0.87853068***	***1.0648826***	***1.1509642***	***1.6535941***
Connecticut	***0.36698117***	***0.5742014***	***1.1362083***	***0.4742867***
Delaware	***0.65611785***	***1.242363***	***1.0624285***	***0.6751108***
Florida	***1.7125788***	***1.7487042***	***1.1804761***	***1.3630401***
Georgia	***1.93499163***	***1.1014227***	***0.8853571***	***1.1023961***
Hawaii	***0.589394***	***0.2401206***	***1.224744***	***0.8631163***
Idaho	***0.28913668***	***0.6264015***	***0.7968214***	***0.6067451***
Illinois	***1.15654672***	***1.2997832***	***1.224744***	***1.0254847***
Indiana	***0.80068619***	***0.5898614***	***0.9591368***	***0.8972991***
Iowa	***0.24465411***	***0.2923207***	***0.8410892***	***0.4828324***
Kansas	***0.66723849***	***0.6003015***	***0.9738928***	***0.7691135***
Kentucky	***1.07870223***	***0.5689814***	***0.7673095***	***0.696475***
Louisiana	***1.7125788***	***1.2997832***	***0.9738928***	***0.9485734***
Maine	***0.23353347***	***0.4332611***	***0.7525535***	***0.3332825***
Maryland	***1.2566325***	***1.5660038***	***0.9886487***	***1.1878531***
Massachusetts	***0.48930823***	***0.7777819***	***1.2542559***	***0.696475***
Michigan	***1.34559763***	***1.3311032***	***1.0919404***	***1.4997714***
Minnesota	***0.30025732***	***0.3758409***	***0.9738928***	***0.6366551***
Mississippi	***1.79042329***	***1.3519833***	***0.6492619***	***0.7306579***
Missouri	***1.00085774***	***0.9291623***	***1.0329166***	***1.2049445***
Montana	***0.66723849***	***0.5689814***	***0.7820654***	***0.7007479***
Nebraska	***0.47818759***	***0.5324413***	***0.914869***	***0.7050207***
Nevada	***1.35671827***	***1.3154432***	***1.1952321***	***1.9655124***
New Hampshire	***0.23353347***	***0.2975407***	***0.8263333***	***0.405921***
New Jersey	***0.82292748***	***0.829982***	***1.3132797***	***0.8032964***
New Mexico	***1.26775314***	***1.4877036***	***1.0329166***	***1.3715858***

New York	***1.23439121***	***1.3258832***	***1.2690118***	***1.1152146***
North Carolina	***1.4456834***	***1.7591443***	***0.6640178***	***0.6879293***
North Dakota	***0.08896513***	***0.2349006***	***0.6492619***	***0.3119183***
Ohio	***0.81180683***	***0.6264015***	***1.1066964***	***0.9143905***
Oklahoma	***0.73396234***	***0.7882219***	***1.0034047***	***0.8545706***
Oregon	***0.54491144***	***0.829982***	***0.9886487***	***1.2519459***
Pennsylvania	***0.70060042***	***0.5533213***	***1.0624285***	***0.6366551***
Rhode Island	***0.37810181***	***0.9082822***	***1.2837678***	***0.3546468***
South Carolina	***1.60137239***	***1.4563835***	***0.7082857***	***0.9613919***
South Dakota	***0.42258438***	***0.4489211***	***0.6640178***	***0.5469252***
Tennessee	***1.46792469***	***0.9813624***	***0.8706011***	***1.1493975***
Texas	***1.41232148***	***1.0492225***	***1.1804761***	***1.0895775***
Utah	***0.35586053***	***0.6264015***	***1.1804761***	***0.9784833***
Vermont	***0.24465411***	***0.2505606***	***0.4721904***	***0.4785595***
Virginia	***0.94525453***	***0.814322***	***0.9296249***	***0.8844806***
Washington	***0.44482566***	***0.7569018***	***1.0771845***	***1.1194875***
West Virginia	***0.63387657***	***0.422821***	***0.5754821***	***0.3973753***
Wisconsin	***0.28913668***	***0.2766607***	***0.9738928***	***0.4614681***
Wyoming	***0.75620363***	***0.840422***	***0.8853571***	***0.6665651***

Calcoliamo le distanze euclidee tra i 6 vettori unità

```
           [1,]       [2,]       [3,]       [4,]       [5,]
[2,]  1.07677476
[3,]  0.83309923 0.79153828
[4,]  0.56239847 1.14235122 0.85918366
[5,]  1.08903477 0.66883401 0.46200762 1.17510622
[6,]  1.01069879 0.63778659 0.57474001 0.92687683 0.44677614
```

2.3 I metodi di raggruppamento

I metodi di raggruppamento si dividono in due tipologie:

- Metodi gerarchici
- Metodi non gerarchici o partitivi

I metodi gerarchici si basano su una procedura costituita da stadi successivi, il cui prodotto finale è un insieme di partizioni e non un'unica partizione. Essi si dividono ulteriormente in:

- Metodi agglomerativi
- Metodi divisivi

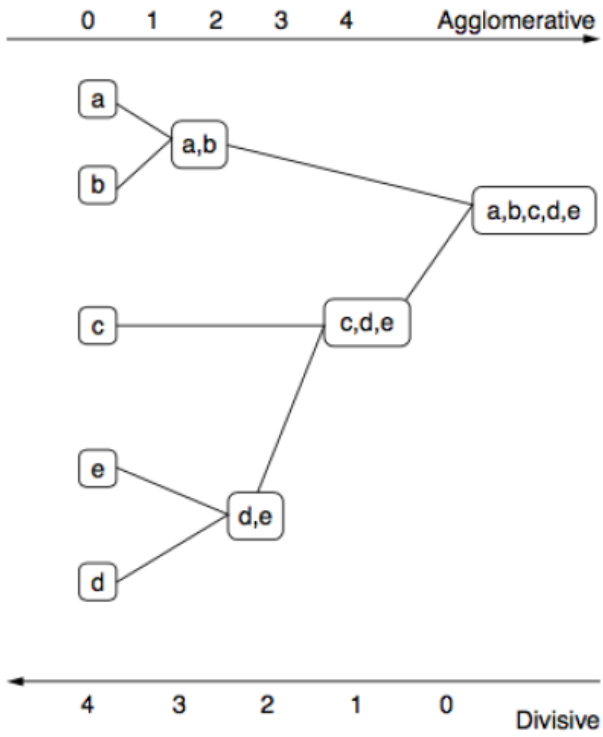


Figura 34 metodi agglomerativi e divisivi

Per quanto riguarda i metodi agglomerativi il punto di partenza sono n gruppi, ciascuno formato da una sola unità statistica. Questi gruppi vengono poi aggregati tra di loro attraverso passaggi successivi. Per quanto riguarda i metodi divisivi invece, il punto di partenza è un solo gruppo contenente tutte ed n le unità statistiche. Tramite una serie di partizioni successive si arrivano ad ottenere gruppi di dimensione unitaria.

I metodi non gerarchici ci portano ad un'unica partizione delle n unità statistiche attraverso due fasi:

- Si determina una partizione delle n unità in un certo numero di gruppi (la scelta di questo numero iniziale può essere effettuata sulla base di una precedente analisi gerarchica)
- Le unità vengono spostate da un gruppo all'altro secondo una strategia volta a massimizzare una prefissata funzione obiettivo.

2.3.1 I metodi gerarchici

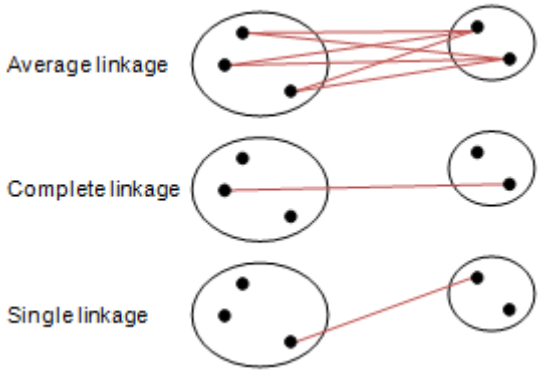
La procedura si articola in tre fasi:

- Data la matrice delle distanze D, si individuano le due unità aventi distanza minima. Queste andranno a formare il primo gruppo.
- Si calcolano ora le distanze fra il gruppo appena formato e le altre unità (nelle fasi più avanzate si farà la stessa cosa, solamente che invece di unità saranno gruppi).

- Si ripetono queste operazioni per (n-1) volte. Il processo si interrompe quando tutte le unità sono parte di un unico gruppo.

In base a come viene ricalcolata la distanza tra i gruppi ad ogni iterazione si distinguono cinque metodi:

- Metodo del legame singolo (*nearest neighbour*): come suggerisce il nome la distanza tra due gruppi è posta pari alla più piccola delle distanze calcolabili a due a due tra tutti gli elementi dei due gruppi. Questo metodo privilegia l'omogeneità tra gli elementi del gruppo a scapito della differenziazione netta tra gruppi.
- Metodo del legame completo (*furthest neighbour*): si considera la maggiore delle distanze calcolate a due a due tra tutti gli elementi dei due gruppi. Questo metodo privilegia la differenza tra i gruppi piuttosto che l'omogeneità degli elementi di ogni gruppo.
- Metodo del legame medio (*average linkage*): si considera come distanza tra due gruppi la media di tutte le distanze calcolate a due a due tra tutti gli elementi dei due gruppi. È considerato come una situazione di compromesso.
- Metodo del centroide: si considera, come distanza tra due gruppi, la distanza tra i rispettivi centroidi (o baricentri), ovvero le medie dei valori assunti dalle unità facenti parte di ciascun gruppo.
- Metodo di Ward: si uniscono i gruppi dalla cui unione deriva il minimo incremento possibile della devianza *within*.



[46]

Figura 35 legame medio, legame completo, legame singolo

La rappresentazione grafica di tali metodi viene chiamata dendrogramma o diagramma ad albero. Sull'asse orizzontale abbiamo gli elementi raggruppati e sull'asse verticale abbiamo la distanza alla quale avviene la fusione. Riportiamo il dendrogramma relativo all'esempio precedente.

[46] *https://www.multid.se/genex/onlinehelp/clustering_distances.png*

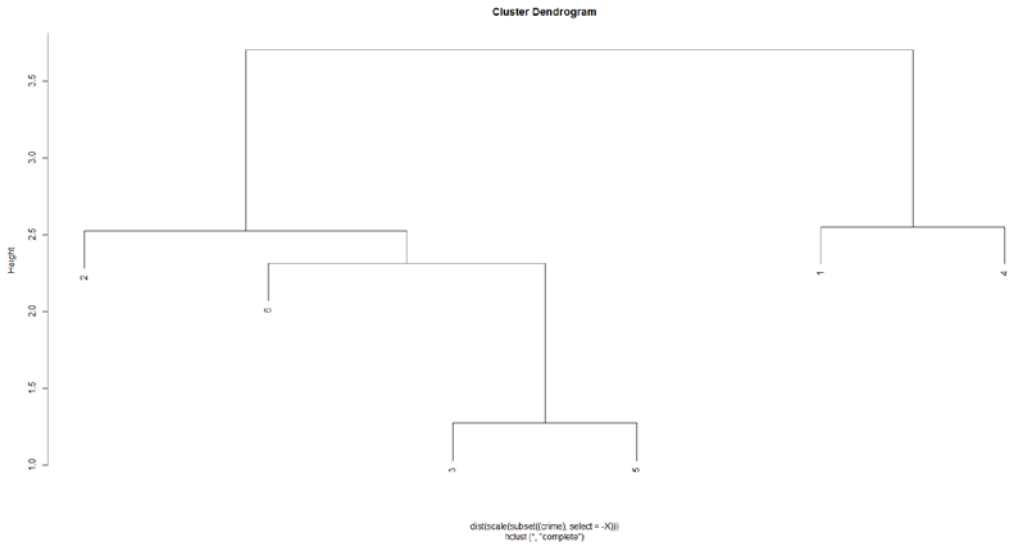


Figura 36 dendrogramma

Esso ci mostra le partizioni che si ottengono a livelli crescenti di distanza. Per l'individuazione del numero ottimo di gruppi possiamo proporre un taglio analizzando la distanza di fusione. Il dendrogramma si taglia laddove presenta un cosìddetto "*salto*". Ecco un possibile taglio:

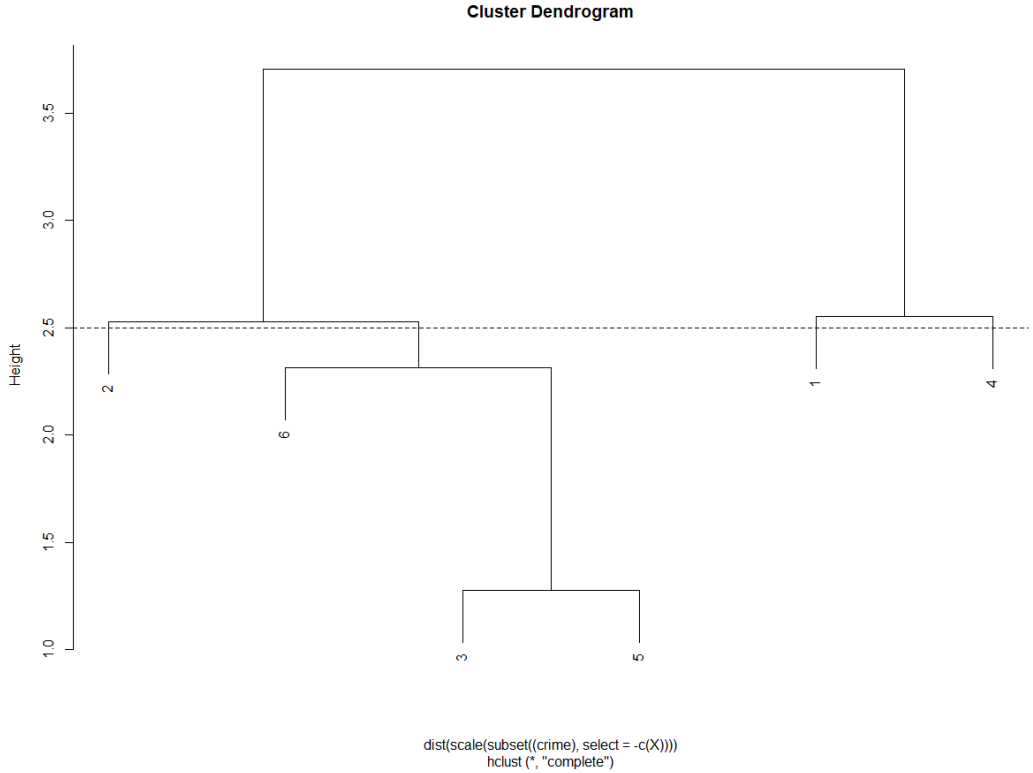


Figura 37 taglio del dendrogramma

2.3.2 I metodi non gerarchici

Tra i metodi non gerarchici, il più noto è l'algoritmo di partizione delle k-medie (*k-means*). Si compone varie fasi:

- Individuazione della partizione del cluster. L'individuazione della partizione ottimale comporterebbe a rigore l'esame di tutte le possibili assegnazioni distinte delle n unità statistiche a G gruppi. Quindi, una partizione formata da tre gruppi (G=3), il numero *P* di possibili soluzioni è

  $\mathbf{3^{n-1} - 1}$ dove n rappresenta il numero di unità da classificare. Quindi ad esempio con n = 5 e G = 3 avremo $\boldsymbol{P = 3^{5-1} - 1}$ = 81 – 1 = 80.

  Il costo computazionale della procedura già con n = 1000 è immenso (per le macchine odierne). Per questo si tende a scegliere un compromesso: si definisce in partenza il numero di cluster da generare. Tale valore, indicato con *k,* può essere determinato sulla base di risultati di una precedente analisi gerarchica.
- Si selezionano casualmente dalla matrice dei dati un numero *k* di unità che costituiranno i centri dei cluster. Una sorta di *"centri provvisori"* che inducono una prima partizione temporanea.
- L'aggregazione avviene sulla base della *minima distanza* da uno di questi centri. La distanza più utilizzata è quella euclidea. Questi verranno considerati come "nuovi centri provvisori".
- Si ripete il procedimento di allocazione delle unità ai centri sulla base della minima distanza. Si itera la partizione tornando al passo 2.

  Se tra un passo e il successivo non vi sono riallocazioni dei punti tra un gruppo e un altro (lo vediamo dal fatto che le distanze non si sono modificate), la procedura si arresta e la partizione può ritenersi soddisfacente.

2.4 La valutazione della partizione

2.4.1 Devianza interna e devianza esterna

Ora sappiamo come suddividere le unità in gruppi. Il passo successivo è verificare la bontà della suddivisione effettuata. Un gruppo è "buono" quando i gruppi sono omogenei al loro interno ed eterogenei rispetto agli altri. Dobbiamo quindi introdurre due concetti fondamentali: la Devianza interna (Within) e la Devianza esterna. La Devianza interna misura il livello di omogeneità interna e la Devianza esterna misura il livello di eterogeneità esterna. La somma di questi due valori rappresenta la Devianza totale. In formule:

$$\boldsymbol{Dev(T) = Dev(W) + Dev(B)}$$

La devianza Within è la somma dei quadrati degli scarti tra i punteggi di ogni soggetto e la relativa media di gruppo, per ognuna delle *p* variabil, per ogni unità e per ogni gruppo. In formule:

$$\boldsymbol{Dev(W) = \sum_{j=1}^{g} \sum_{k=1}^{p} \sum_{i=1}^{n(j)} (x_{ik} - \overline{x}_{jk})^2}$$

$\boldsymbol{g}$ = numero dei gruppi

$\boldsymbol{j}$ = generico gruppo

$\boldsymbol{p}$ = numero delle variabili

$\boldsymbol{k}$ = generica variabile

$\boldsymbol{n(j)}$ = numero di unità appartenenti al generico gruppo $\boldsymbol{j}$

$\boldsymbol{i}$ = generica unità appartenente al generico gruppo $\boldsymbol{j}$

$\boldsymbol{x_{ik}}$ = valore riportato dalla generica unità $\boldsymbol{i}$ rispetto alla generica variabile $\boldsymbol{k}$

$\boldsymbol{\overline{x}_{jk}}$ = media dei valori riportati dalle unità appartenenti al generico gruppo $\boldsymbol{j}$ rispetto alla generica variabile $\boldsymbol{k}$

La devianza Between è la somma dei quadrati degli scarti, in questo caso la differenza tra i punteggi medi di gruppo e la media generale, rispetto a ciascuna delle p variabili. È la media ponderata dei valori calcolati precedentemente, in questo caso i pesi sono rappresentati dal numero delle unità facenti parte di ciascun gruppo. In formule (la terminologia è la stessa della formula precedente):

$$\boldsymbol{Dev(B) = \sum_{k=1}^{p} \sum_{j=1}^{g} n(j)(\overline{x}_{jk} - \overline{x}_k)^2}$$

Minore devianza Within significa maggiore omogeneità all'interno dei gruppi, maggiore devianza Between significa maggiore eterogeneità tra i gruppi. La devianza totale si calcola come somma dei quadrati delle differenze tra i valori riportati da ciascuna unità e la media generale, ovviamente per tutte le p variabili. In formula:

$$\boldsymbol{Dev(T) = \sum_{k=1}^{p} \sum_{i=1}^{n} (x_{ik} - \overline{x}_k)^2}$$

2.4.2 L'indice R^2

La bontà di una partizione dipende da quanto si è riusciti a minimizzare la Devianza Within e da quanto si è riusciti a massimizzare la Devianza Between. Un indice sintetico è quindi l'$\boldsymbol{R^2}$ che altro non è che il rapporto tra la Devianza Between e la Devianza Totale, in simboli rispettivamente Dev(B) e Dev(T).

$$\boldsymbol{R^2 = \frac{Dev(B)}{Dev(T)} = 1 - \frac{Dev(W)}{Dev(T)}}$$

Tale indice consente di confrontare tra di loro partizioni costituite da un diverso numero di gruppi o addirittura determinate attraverso l'applicazione di metodi diversi tra loro. In base alla formula di scomposizione della devianza tale indicatore varia tra 0 e 1 (1 per i gruppi perfettamente omogenei al loro interno e ben separati all'esterno). Questo ci porta a fare una considerazione: secondo tale formula la partizione migliore è una partizione che abbia $\boldsymbol{R^2}$ = 1 e che quindi deve avere Dev(B) = Dev(T). Questo ci porterebbe ad avere una partizione costituita da n gruppi, ciascuno formato da una sola unità, partizione che per uno studioso è assolutamente priva di significato. Dobbiamo quindi trovare il giusto equilibrio tra due esigenze contrapposte:

- Avere un elevato grado di omogeneità interna ai gruppi
- Avere un elevato grado di sintesi della partizione considerata.

Capitolo 3: La comunicazione politica, nuovi mezzi e nuovi strumenti di analisi

3.1 Analisi delle variabili di aggregazione

Nel capitolo precedente è stata esaminata la metodologia della cluster analysis che sarà il nostro strumento fondamentale nella analisi seguente. Il campione è composto da 97 tweet riconducibili a 90 utenti. L'arco temporale della raccolta dati va dal 29 al 5 dicembre. Nel periodo considerato sono stati scaricati i tweet contenenti le due keyword "referendum" e "costituzionale". Dai dati ottenuti sono stati selezionati i tweet contenenti hashtag caratterizzati in senso "partisan" ("iovotosi, iovotono, bastaunsi, iodicono). Il nostro intento è suddividere questi potenziali influencer e brand advocate in gruppi il più possibile omogenei al loro interno. Le variabili considerate sono per ciascun utente sono: produttività (media), numero (medio) di follower, numero di amici, se il tweet era un retweet o meno, il giorno di generazione del tweet, il fatto che fosse favorevole o contrario alla riforma.

- tweet_id_str (rinominato "utente"). Corrisponde ad una codifica numerica del nome utente,
- numero_tweet_id_str. Corrisponde al numero di tweet presi in considerazione per ogni utente
- produttivita. Corrisponde al numero di tweet postati nei giorni presi in considerazione.
- media_user_followers_count. Corrisponde al numero di follower dell'utente preso in considerazione.
- media_user_friends_count. Corrisponde al numero di amici dell'utente preso in considerazione.
- max_tweet_retweeted_01. Indica se il tweet preso in considerazione era un retweet.
- max_giorno. Indica l'ultimo giorno in cui l'utente ha trasmesso un tweet tra i giorni presi in considerazione.
- max_scelta_01. Indica se l'utente era per il SI o per il NO. È stato ottenuto analizzando gli hashtag.

Stiamo prendendo in considerazione il terzo referendum costituzionale nella storia della Repubblica Italiana, che ebbe luogo il 4 dicembre 2016. La maggioranza dei votanti respinse il testo di legge costituzionale della cosiddetta riforma Renzi-Boschi, approvato in via definitiva dalla Camera il 12 aprile 2016 e recante modifiche alla parte seconda della Costituzione[47] (fig. 38). La riforma era nata su iniziativa del Governo Renzi, guidato dal leader del Partito Democratico Matteo Renzi, che ha legato al risultato del referendum il proprio destino politico. Tra le forze politiche che sostenevano il Governo in Parlamento, e che votarono quindi la riforma, rientravano, oltre ai parlamentari del PD, i gruppi di Area Popolare, formato da iscritti a Nuovo Centrodestra e UdC, di ALA, in gran parte formato da ex iscritti a Forza Italia guidati da Denis Verdini, e altre formazioni minori come Centro Democratico, Partito Socialista Italiano e Scelta Civica. Tra quelle che si opposero alle modifiche costituzionali figuravano invece Movimento 5 Stelle, Sinistra Italiana - Sinistra Ecologia Libertà, Lega Nord e Fratelli d'Italia, alle quali si aggiunge Forza Italia, che nelle prime fasi del cammino della riforma

[47] *http://www.gazzettaufficiale.it/eli/id/2016/04/15/16A03075/sg*

in Parlamento le aveva sostenute.

Referendum 25/06/2006 ▸ Area ITALIA + ESTERO

Elettori	49.772.506				
1.REFERENDUM COSTITUZIONALE. Approvazione legge di modifica alla parte seconda della Costituzione					
Votanti	26.110.925	52,46%	Voti validi	25.753.782	
Schede bianche	101.429		Schede non valide (bianche incl.)	357.143	
Sì	9.970.513	38,71%	No	15.783.269	61,29%

[48]

Figura 38 dati referendum

In seguito ricercheremo i caratteri di similarità che accomunano i profili appartenenti a ciascun gruppo.

Riportiamo di seguito la matrice dei dati a valori originali.

[48] *http://elezionistorico.interno.gov.it/index.php?tpel=F&dtel=25/06/2006&tpa=Y&tpe=A&lev0=0&levsut0=0&es0=S&ms=S*

Etichette di riga	Conteggio di tweet_id_str	Media di produttività	Media di user_followers_count	Media di user_friends_count	Max di tweet_retweeted_01	Max di giorno	Max di scelta_01
20689573	1	16	1102	1978	0	2	0
26295447	1	88	13824	2299	1	2	0
32758692	1	13	971	1747	0	3	0
55238295	1	23	73526	2521	0	2	1
70409661	1	1	119	345	0	2	1
103606685	1	1	297	533	0	3	1
117363440	2	237	2890,5	3136	0	4	0
133461180	1	625	4135	1929	1	4	1
151578963	1	17	100	207	0	3	1
174249848	1	7	4846	4525	0	4	1
347497240	1	25	7709	5275	1	1	0
358808488	1	55	556	183	0	3	0
362516905	1	5	327	298	0	1	0
366698775	1	11	65	249	0	4	1
371781632	1	370	2376	2733	0	3	1
376234197	1	1	278	866	0	3	1
378706057	1	108	2278	2092	0	30	1
396503965	1	2	232	624	0	4	1
436640636	1	525	4460	1766	1	4	1
456636936	1	10	2565	2508	0	1	0
465765184	1	6	2121	2045	1	3	1
473848168	1	63	417	45	0	2	0
537412181	1	3	1282	992	0	2	1
579147580	1	8	124	731	0	4	0
595878183	1	60	299	266	1	2	1
616372545	1	4	27	50	1	4	0
764392598	1	1	37	87	1	4	1
989315192	1	3	52	106	0	4	0
999469686	1	124	829	377	0	4	1
1039274630	1	15	16	54	0	2	1
1059881101	2	73	2448,5	2384	1	30	0
1070392579	1	80	7959	7775	0	29	1
1077269276	1	11	2312	1802	1	4	0
1212147877	2	69	3554,5	3019,5	0	2	0
1222000628	1	123	402	618	0	1	1
1254291416	1	25	1079	2110	0	3	0
1327639712	1	59	73076	64409	0	2	1
1331128747	1	15	389	301	1	3	0
1367363988	1	30	267	1367	0	2	0
1384894520	1	296	10076	5390	1	3	0
1438322906	1	2	3077	1738	0	5	0
1476948512	1	61	720	661	0	4	0
1536461521	1	1	228	160	0	4	1
1546547437	1	13	184	451	0	4	1
1586227562	1	147	1064	2559	0	2	1
1601298356	1	5	1665	292	0	4	0
1638143131	1	7	2473	2563	1	3	1
1650257054	1	6	227	496	0	5	0
1701198374	1	10	1128	1807	1	3	0
1709866242	1	24	17913	10581	1	3	0
2307171263	1	2	0	3	0	3	1
2340110677	1	31	392	12	0	2	0
2341872502	1	31	516	237	0	5	0
2368943517	1	8	40	83	0	4	1
2427964266	1	51	962	1975	1	2	0
2436608759	1	420	1045	202	1	1	1
2499858601	1	1	40	180	0	4	1
2553289698	2	34	346	316	0	3	0
2607531334	1	10	887	197	0	4	0
2695632918	3	231	10141	9400,333333	0	4	0
2749566881	1	71	134	106	0	4	1
2858759915	1	38	2429	1374	1	4	0
2866161418	1	4	51	77	0	4	0
2935824485	1	20	156	133	0	2	1
2950839160	1	3	11	93	0	4	0
2962475685	1	370	129	70	0	1	0
2990358471	1	2	21	94	0	1	1
3015774443	1	18	265	1023	0	2	0
3020578613	1	13	78	70	0	1	0
3072811918	1	29	3	37	0	2	1
3300962806	1	1	7	38	0	1	0
3354137837	1	49	364	140	1	1	0
3368018859	1	57	1339	1610	1	4	0
4474043243	1	285	692	761	1	2	0
4655248997	1	338	317	2511	0	4	1
4900930252	1	4	1167	527	0	4	0
7,02129E+17	1	233	660	330	1	2	0
7,06068E+17	1	243	780	1221	1	2	0
7,11197E+17	1	2	276	766	0	1	1
7,12407E+17	1	3	819	841	0	4	1
7,13001E+17	1	28	253	60	0	3	1
7,47328E+17	1	16	594	2525	1	4	0
7,528E+17	1	125	68	315	0	30	1
7,72209E+17	1	2	55	270	0	3	0
7,76686E+17	1	4503	198	5	0	3	0
7,76755E+17	1	1	18	202	0	2	1
7,80807E+17	1	5	37	156	1	1	0
7,96376E+17	1	1	27	207	0	2	0
7,98743E+17	2	89	49	107,5	1	2	0
8,05387E+17	1	40	11	12	1	4	0

Per la prima parte della nostra analisi prenderemo in considerazione solamente le colonne da 1 a 5. Nella seconda parte prenderemo anche in considerazione la variabile relativa alla scelta SI o NO.

Qui vengono riportati gli *user name* e gli *user screen name*, che utilizzeremo in seguito durante l'analisi.

	tweet_id_str	user_name	user_screen_name
1	20689573	Sergio Della Lena	SergioDL
2	26295447	Ross	RossellaFidanza
3	32758692	An Italian	Be_Italian_
4	55238295	anna paola concia	annapaolaconcia
5	70409661	Claudio Longo	claudioit9cbe
6	103606685	Elena Perotti	E_Perotti
7	117363440	Davide	ricci_davide77
8	133461180	Antonio Gentile	antgentile
9	151578963	Michele Povoli	MichelePovoli
10	174249848	Maurizio Amoroso	avvocato2punto0
11	347497240	Bubi	calygola
12	358808488	Giovanni Nappi	Giovanni_Nappi
13	362516905	WinterMute	crazybalzano
14	366698775	Wasim	Wasimj96
15	371781632	Raffaele Pizzati	RaffaelePizzati
16	376234197	Andrea Lion	andrealion1
17	378706057	à¸„driÃ¡à¸	Adrian_in_it
18	396503965	Luca Valdrighi	LValdrighi
19	436640636	Francesco Balsamo	TheLambkin_
20	456636936	Eros Forenzi	EForenzi
21	465765184	Chiara Raimondi	chiaramondi
22	473848168	Scugnizz'e Brigante	Scugnizzobrigan
23	537412181	Lorenzo Pelliconi	LorenzPellico
24	579147580	Mozzini Edoardo G.	Ed96webchannel
25	595878183	daniela bert.	danbertsamp
26	616372545	GIORGIOBELLINO	GIORGIOBELLINO2
27	764392598	LAtoB__	LAtoB__
28	989315192	Ape MagÃ	ApeMag
29	999469686	Antonella	Antonella180262
30	1039274630	Giuseppe Zamperetti	BeppeZamperetti
31	1059881101	Mauro Beltramo	MauroBeltramo
32	1070392579	Giuseppe Sama	BeppeSama
33	1077269276	Currenti Calamo	CurrentiCalamo_
34	1212147877	#IOvotoNO	paceinterra_it
35	1222000628	rossana delpiccolo	stemar9288
36	1254291416	cristina atzeri	catzeri3235
37	1327639712	#IOVOTOSI	paparcura
38	1331128747	Sandra Abbondandolo	Sandra_AbbDR
39	1367363988	barbara sardella	SardellaBarbara
40	1384894520	Mauro Barin	MauroBarin

41	1438322906	Elisa Bellino	elfiegnomi
42	1476948512	Sandra Moro	SandraM_Tcon0
43	1536461521	maryshark	mariateresabru
44	1546547437	nicola	nicola1691
45	1586227562	rugaskipper	rugaskipper
46	1601298356	Marta Saitta	SaittaMarta
47	1638143131	Corrado Petrocelli	CorrPetrocelli
48	1650257054	StefaniaPernisa	LaStefi_P
49	1701198374	Uff Post	UffPost
50	1709866242	RetwittatorCortese	RETWITTATORc
51	2307171263	Salvatore Pomara	salgiupom
52	2340110677	acquo	acquodario
53	2341872502	Shlomo	Shlomo_75
54	2368943517	Marco Marinoni	MarcoM_Marinoni
55	2427964266	Forza Italia Sanremo	FISRemo
56	2436608759	Ginill ðŸ”´âš«ï¸?	Uleprr
57	2499858601	Roberto Giacomelli	xetibor
58	2553289698	masterofmate	masterofmate
59	2607531334	ITALIA_FASCISTA	PREDAPPIO98
60	2695632918	lega nord	lega_nord
61	2749566881	Alessandra Estatico	alexa5313
62	2858759915	MassimoLimonta71	massimo_limonta
63	2866161418	Mario Grasso	milazzo1987
64	2935824485	Luca Soldini	lucasoldini_93
65	2950839160	Francesca Di Valerio	FranciDiValerio
66	2962475685	Luigi Leonardi	_LuigiLeonardi
67	2990358471	Mauro Fontana	mrmafont
68	3015774443	silvia carcione	nove_silvia
69	3020578613	La grullaia	lagrullaia
70	3072811918	Mirco Lupi	LupiMirco
71	3300962806	Fabio	Fabio84V
72	3354137837	FI Regione Campania	fi_regcampania
73	3368018859	Michele Bobbio	Bobbio65M
74	4474043243	ParteCivile	ParteCivile
75	4655248997	Francesco Bianchini	fbianco91
76	4900930252	AntonioMarrapeseBarr	amarrapese_barr
77	7,02129E+17	Laura #IoDicoNo!!	LauraGio_75
78	7,06068E+17	Mauro 55	Mauro5514
79	7,11197E+17	Mostro Alfonso	DragoRosso_
80	7,12407E+17	Lilly Tagloff	iceflaws
81	7,13001E+17	Marzia Cappelli	marzia_cappelli

82	7,47328E+17	Henk Nexit	NEXITIUS
83	7,528E+17	NicolÃ²	izoon2
84	7,72209E+17	Antonio Maggio	MaggioRLIPz
85	7,76686E+17	Schiforma	Schiforma
86	7,76755E+17	Annalisa	annpn83
87	7,80807E+17	Tamara #IOVOTOSI	_Referendum_
88	7,96376E+17	Victor div	Naiandiv
89	7,98743E+17	Melania	Melania11564076
90	8,05387E+17	quantmint	quantmint1

Prima di iniziare una analisi approfondita trovo utile effettuare una rapida analisi descrittiva delle variabili prettamente quantitative (produttività, numero di follower, numero di amici). Le statistiche descrittive delle variabili sono rappresentate in tabella e in forma standardizzata mediante box-plot. Un utile strumento per la descrizione e il confronto di molteplici distribuzioni è il box-plot, un grafico costruito su cinque valori di sintesi:

- Mediana: rappresenta il valore assunto dalle unità che si trovano al centro della distribuzione e costituisce la linea che divide la scatola in due parti
- Primo quartile: rappresenta il valore che si lascia a sinistra il 25% dei dati e rappresenta l'estremo inferiore della scatola.
- Terzo quartile: rappresenta il valore che si lascia alla sinistra il 75% dei dati e rappresenta l'estremo superiore della scatola.
- Il valore minimo, tramite il quale viene tracciato il baffo inferiore.
- Il valore massimo, tramite il quale viene tracciato il baffo superiore.

La lunghezza dei baffi serve a misurare il campo di variazione (la differenza tra il valore massimo e il valore minimo), mentre l'altezza della scatola è data dalla differenza interquartile (la differenza tra il terzo ed il primo quartile).

	Conteggio di tweet_id_str	Media di produttività	Media di user_followers_count	Media di user_friends_count	Max di tweet_retweeted_01
Media	1,08	121,07	3149,77	1992,64	0,30
Errore standard	0,03	50,88	1163,91	729,46	0,05
Mediana	1	19	397	511,5	0
Moda	1	1	27	207	0
Deviazione standard	0,31	482,71	11041,81	6920,23	0,46
Curtosi	19,90	78,55	36,12	76,49	-1,24
Asimmetria	4,31	8,62	5,93	8,47	0,89
Intervallo	2	4502	73526	64406	1
Minimo	1	1	0	3	0
Massimo	3	4503	73526	64409	1
Conteggio	90	90	90	90	90

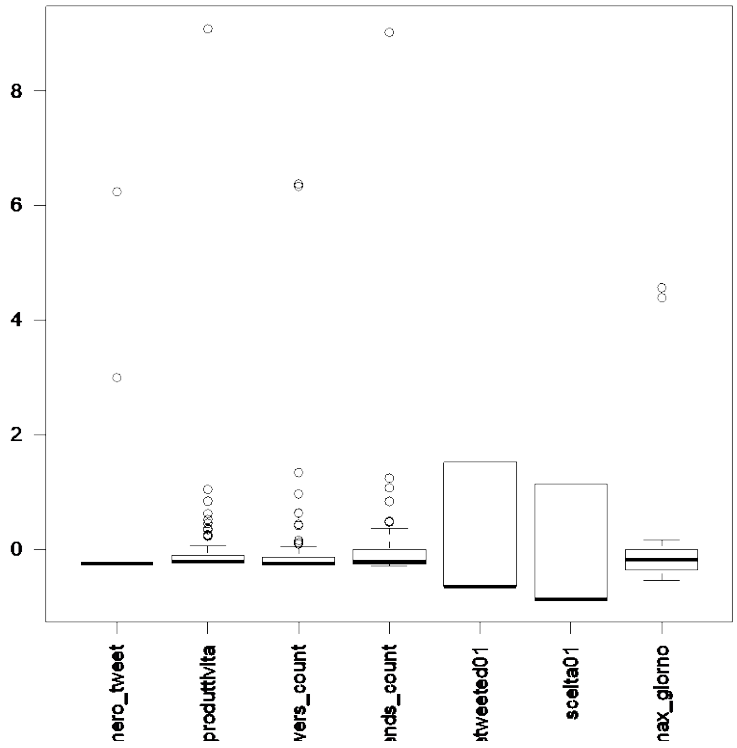
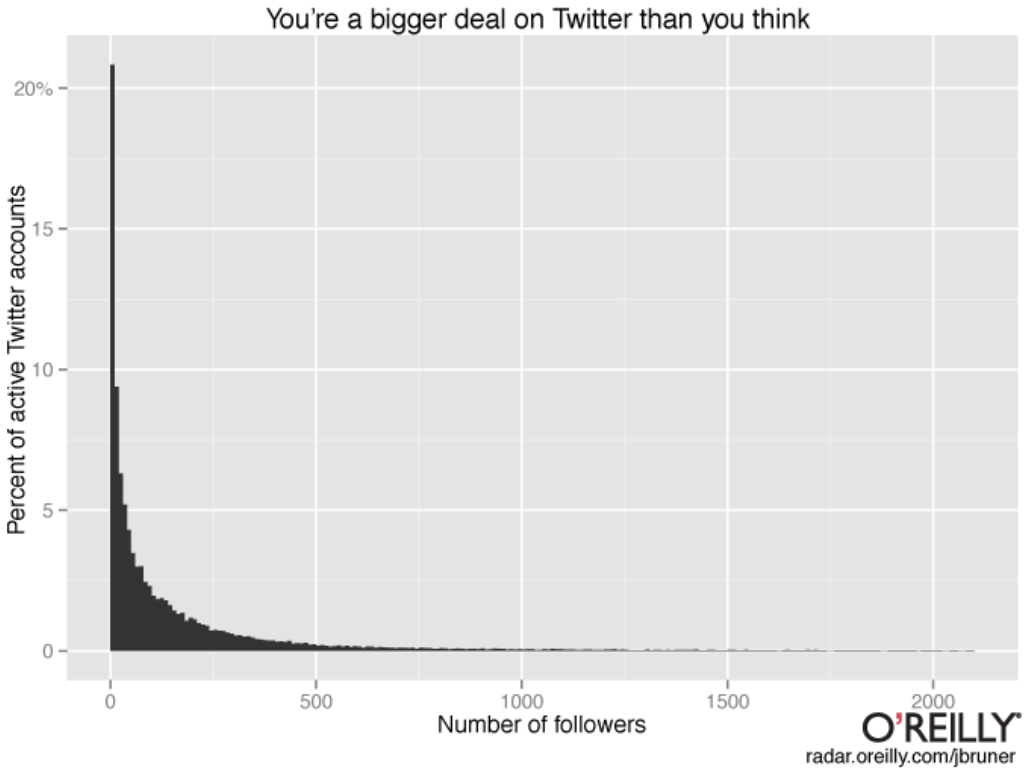


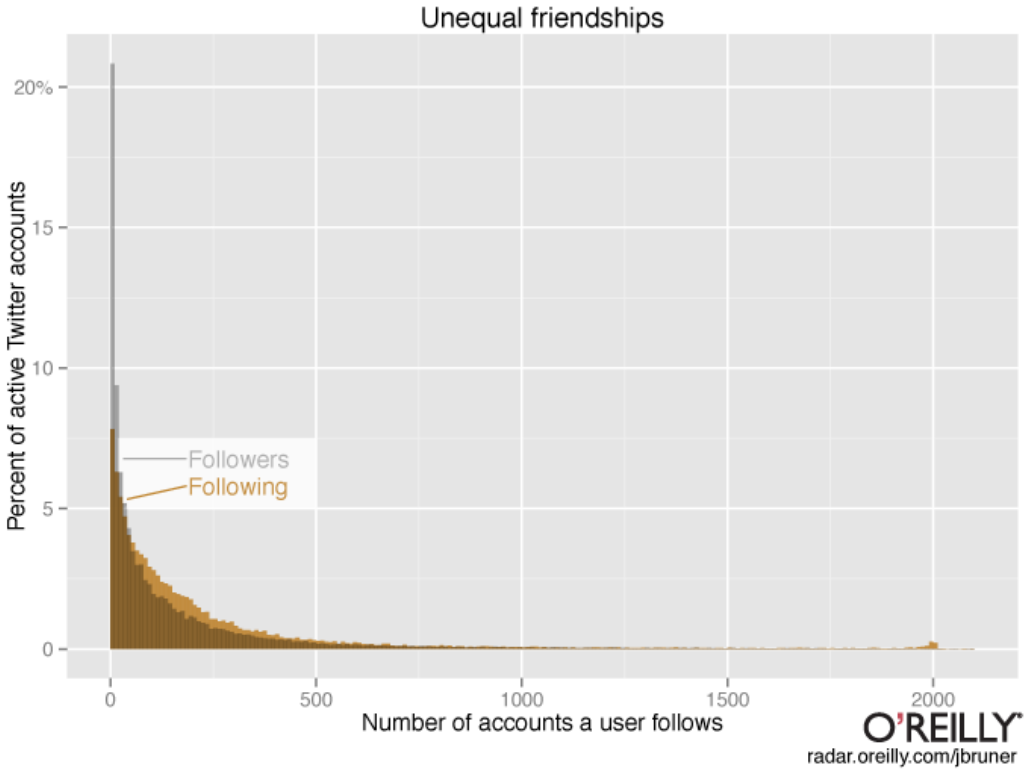
Figura 39 box plot

Il grafico è parzialmente in linea con quanto scoperto da studi precedenti[49]. Secondo l'articolo di Jon Bruner pubblicato per O'Reily Radar il 18 dicembre 2013, su un campione casuale di 400,000 utenti l'account mediano ha un singolo follower (prendendo in considerazione gli account che si sono loggati almeno una volta al mese). Se invece prendiamo in considerazione gli account che hanno postato almeno una volta in un mese l'account mediano ha 61 follower. Un account con 1000 follower si trova già nel 96esimo percentile. Il 76% segue più persone di quante poi seguano loro.

Come possiamo vedere anche nel nostro caso la maggioranza degli utenti segue più persone di quante poi seguano indietro, tuttavia l'account mediano ha molti più follower e amici di quello dello studio di Jon Bruner (il nostro campione è però molto più piccolo).

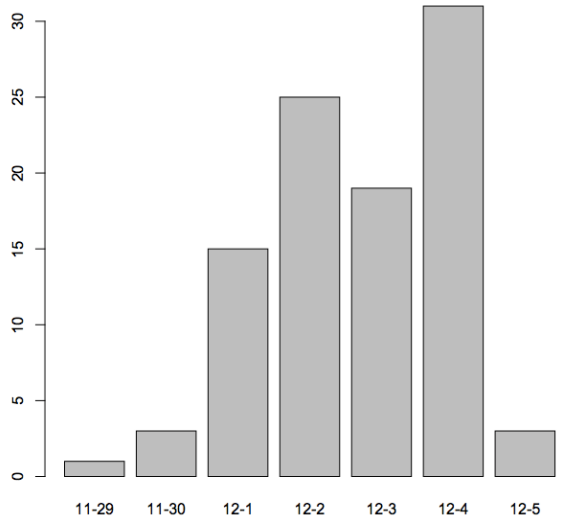
[49] *https://www.oreilly.com/ideas/tweets-loud-and-quiet*


You're a bigger deal on Twitter than you think
20%
15
10
5
0
Percent of active Twitter accounts
0
500
1000
1500
2000
Number of followers
O'REILLY
radar.oreilly.com/jbruner


Unequal friendships
20%
15
10
5
0
Percent of active Twitter accounts
Followers
Following
0
500
1000
1500
2000
Number of accounts a user follows
O'REILLY
radar.oreilly.com/jbruner

Percentile of active Twitter accounts	Number of follwers
10	3
20	9
30	19
40	36
50	**61**
60	98
70	154
80	246
90	458
95	819
96	978
97	1,211
98	1,675
99	2,991
99.9	24,964

Si riporta anche la distribuzione di frequenza del numero di tweet per giorno nel periodo considerato.



frequenza di tweet nei giorni dal 30 novembre al 5 dicembre

3.2 Applicazione della cluster analysis agli influencer italiani su Twitter durante la campagna referendaria

Sulla matrice delle distanze, calcolata a partire dai valori standardizzati, è stato applicato il metodo di Ward.

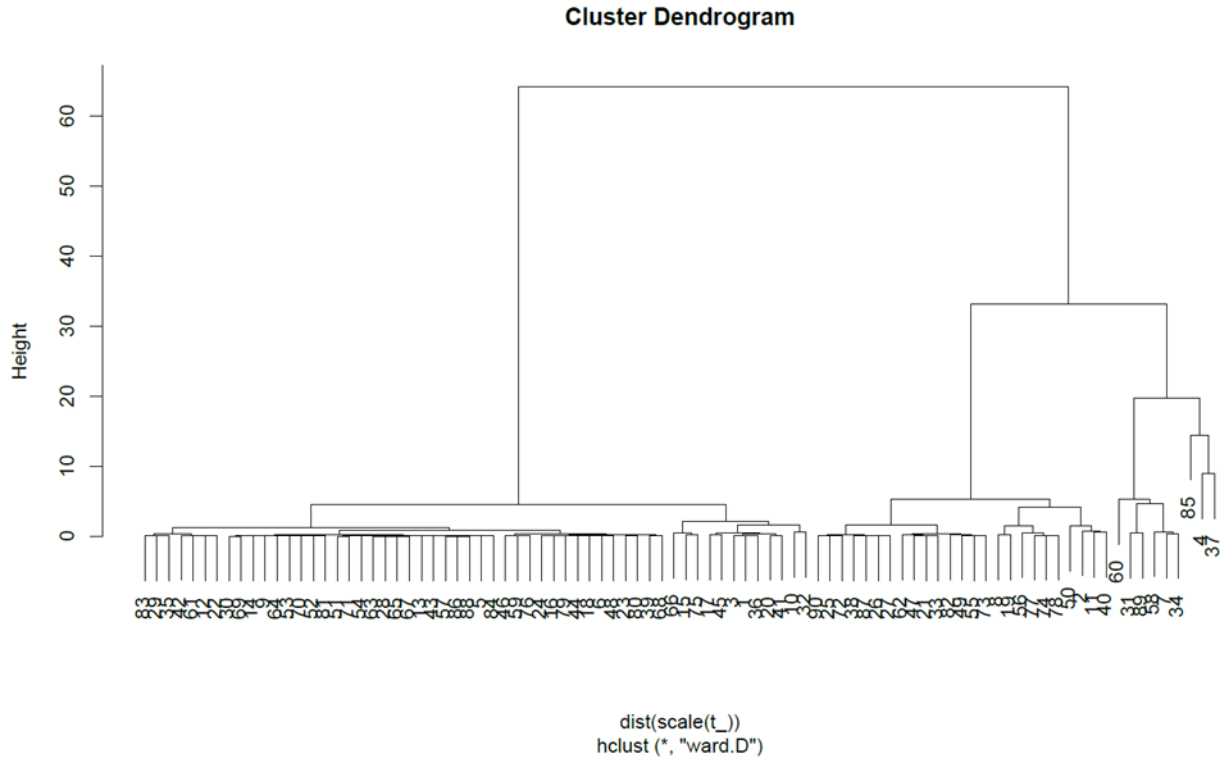


Figura 40 distanza euclidea, metodo di Ward

I metodi del legame completo, del legame medio e di Ward forniscono cluster più stabili rispetto agli altri due in cui le aggregazioni appaiono meno visibili. I rami più lunghi e la numerosità relativamente bilanciata dei gruppi permettono di tagliare il dendrogramma con maggior facilità.

Il metodo del legame singolo invece presenta il problema delle concatenazioni, che si manifestano in raggruppamenti di forma allungata. Le unità si aggiungono volta per volta al primo gruppo formato.

Con il metodo del centroide la distanza di fusione ha un andamento non monotòno, il che significa che, con l'avanzare del processo di aggregazione, essa può aumentare o diminuire generando delle inversioni.

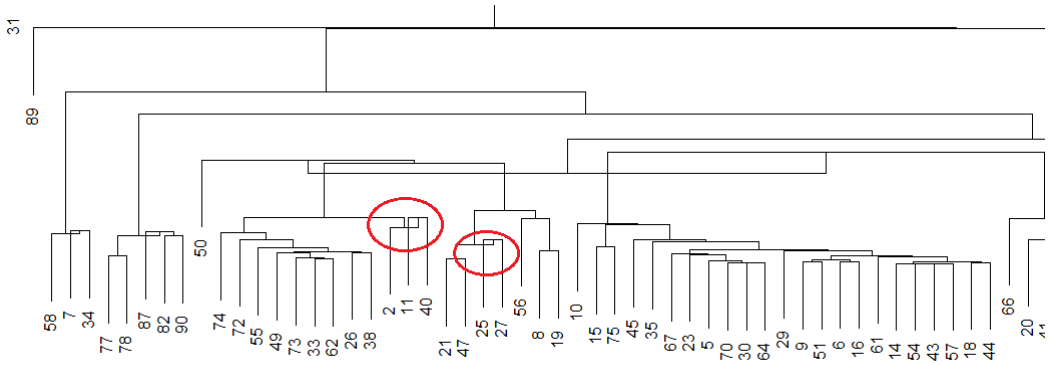


Figura 41 esempio inversioni

Un altro problema è costituito dalla possibilità di fenomeni gravitazionali: i gruppi di dimensioni ridotte vengono catturati da quelli costituiti da un maggior numero di unità.

3.3 Determinazione del numero di gruppi

Per determinare il numero di gruppi useremo NbClust[50], un pacchetto di R per determinare il numero rilevante di cluster in un data set attraverso l'uso di ben 30 indici.

- CH index (Calinski e Harabasz 1974): $\boldsymbol{CH(q)} = \frac{\boldsymbol{trace(B_q)\,/\,(q-1)}}{\boldsymbol{trace(W_q)/(n-q)}}$
- Duda index (Duda e Hart 1973): $\boldsymbol{Duda} = \frac{\boldsymbol{Je\,(2)}}{\boldsymbol{Je(1)}} = \frac{\boldsymbol{W_k + W_l}}{\boldsymbol{W_m}}$
- Pseudot2 index (Duda e Hart 1973): $\boldsymbol{Pseudot2} = \frac{\boldsymbol{W_{kl}}}{\frac{\boldsymbol{W_k+W_l}}{\boldsymbol{n_k+n_l-2}}}$
- Cindex index (Hubert e Levin 1976): $\boldsymbol{Cindex} = \frac{\boldsymbol{S_w - S_{min}}}{\boldsymbol{S_{max} - S_{min}}}, \boldsymbol{S_{min} \neq S_{max}}$
- Gamma index (Baker e Hubert 1975): $\boldsymbol{Gamma} = \frac{\boldsymbol{s(+)-s(-)}}{\boldsymbol{s(+)+s\,(-)}}$
- Beale index (Beale 1969): $\boldsymbol{Beale = F} = \frac{(\frac{\boldsymbol{V_{kl}}}{\boldsymbol{W_k+W_l}})}{((\frac{\boldsymbol{n_m-1}}{\boldsymbol{n_m-2}})\boldsymbol{2^{\frac{2}{p}}-1})}$
- Cubic Clustering Criterion (CCC) index (Sarle 1983):

$$\boldsymbol{CCC} = \boldsymbol{ln}\left[\frac{\boldsymbol{1 - E(R^2)}}{\boldsymbol{1 - R^2}}\right]\frac{\sqrt{\frac{\boldsymbol{np*}}{\boldsymbol{2}}}}{\boldsymbol{(0.001 + E\,(R^2))}}$$

- Ptbiserial index (Milligan (1980, 1981) e Kraemer (1982)):

$$\boldsymbol{Ptbiserial} = \frac{[\boldsymbol{\bar{S}_b - \bar{S}_w}]\left[\boldsymbol{N_w N_b}/\boldsymbol{N_T^2}\right]^{\frac{1}{2}}}{\boldsymbol{s_d}}$$

[50] *Journal of Statistical Software, October 2014, Volume 61, Issue 6.*
https://www.jstatsoft.org/article/view/v061i06/v61i06.pdf

- Gplus index (Rohlf 1974): $\boldsymbol{Gplus} = \frac{2_s(-)}{N_t(N_t-1)}$
- DB index (Davies e Bouldin 1979): $\boldsymbol{DB(q)} = \frac{1}{q}\sum_{k=1}^{q} \underset{k\neq l}{\boldsymbol{max}} \left(\frac{\delta_k+\delta_l}{d_{kl}}\right)$
- Frey index (Frey e Van Groenewoud 1972): $\boldsymbol{Frey} = \frac{\bar{S}_{bj+1}-\bar{S}_{bj+1}}{\bar{S}_{wj+1}-\bar{S}_{wj}}$
- Hartigan index (Hartigan 1975):

$$\boldsymbol{Hartigan} = \left(\frac{\boldsymbol{trace}(\boldsymbol{W}_q)}{\boldsymbol{trace}(\boldsymbol{W}_{q+1})} - \boldsymbol{1}\right)(\boldsymbol{n-q-1})$$

- Tau index (Rohlf 1974): $\boldsymbol{Tau} = \frac{s(+)-s(-)}{\left[\left(N_t(N_{t-1})/_{2-t}\right)\left(N_t(N_{t-1})/_{2}\right)\right]^{\frac{1}{2}}}$
- Ratkowsky index (Ratkowsky e Lance 1978): $\boldsymbol{Ratkowsky} = \frac{\bar{S}}{q^{\frac{1}{2}}}$
- Scott index (Scott e Symons 1971): $\boldsymbol{Scott} = \boldsymbol{n\,log}\ \ \frac{det(T)}{det(W_q)}$
- Marriot index (Marriot 1971): $\boldsymbol{Marriot} = \boldsymbol{q^2 det}(\boldsymbol{W_q})$
- Ball index (Ball e Hall 1965): $\boldsymbol{Ball} = \frac{W_q}{q}$
- Trcovw index (Milligan e Cooper 1985): $\boldsymbol{Trcovw} = \boldsymbol{trace}(\boldsymbol{COV}(\boldsymbol{W_q}))$
- Tracew index (Milligan e Cooper 1985; Edwards e Cavalli-Sforza 1965; Friedman e Rubin 1967; Orloci 1967; Fukunaga e Koontz 1970):
  $\boldsymbol{Tracev} = \boldsymbol{trace}(\boldsymbol{W_q})$
- Friedman index (Friedman e Rubin 1967): $\boldsymbol{Friedman} = \boldsymbol{trace}(\boldsymbol{W_q^{-1}B_q})$
- McClain index (McClain e Rao 1975): $\boldsymbol{McClain} = \frac{\bar{S}_w}{\bar{S}_b}$
- Rubin index (Rubin e Friedman 1967): $\boldsymbol{Rubin} = \frac{det(T)}{det(W_q)}$
- KL index (Krzanowski and Lai 1988): $\boldsymbol{KL}_{(q)} = \left|\frac{\boldsymbol{DIFF_q}}{\boldsymbol{DIFF_{q+1}}}\right|$
- Silhouette index (Rousseeuw 1987): $\boldsymbol{Silhouette} = \frac{\sum_{i=1}^{n} S(i)}{n}$
- Gap index (Tibshirani et al. 2001): $\boldsymbol{Gap(q)} = \frac{1}{B}\sum_{b=1}^{B} \boldsymbol{log}\, \boldsymbol{W_{qb}} - \boldsymbol{log}\, \boldsymbol{W_q}$
- Dindex (Lebart et al. 2000): $\boldsymbol{w(P^q)} = \frac{1}{q}\sum_{k=1}^{q}\frac{1}{n_k}\sum_{x_i\in C_k} \boldsymbol{d}(\boldsymbol{x_i}, \boldsymbol{c_k})$
- Dunn index (Dunn 1974): $\boldsymbol{Dunn} = \frac{\underset{1\leq i<j\leq q}{\boldsymbol{min}}\, \boldsymbol{d}(\boldsymbol{C_i},\boldsymbol{C_j})}{\underset{1\leq k\leq q}{\boldsymbol{max}}\, \boldsymbol{diam}(\boldsymbol{C_k})}$
- Hubert Statistic (Hubert e Arabie 1985): $\boldsymbol{\Gamma(P,Q)} = \frac{1}{N_t}\sum_{\substack{i=1\\ i<j}}^{n-1} \boldsymbol{P_{ij}Q_{ij}}$
- SDindex: $\boldsymbol{SDindex(q)} = \boldsymbol{\alpha Scat(q)} + \boldsymbol{Dis(q)}$
- SDbw index: $\boldsymbol{SDbw(q)} = \boldsymbol{Scat(q)} + \boldsymbol{Density.bw(q)}$

Utilizzeremo l'indice Silhouette, la cui formula può essere riscritta come segue:

$$s_i = \frac{b_i - a_i}{max(a_i, b_i)}$$

a_i è la distanza media fra l'unità i e le altre unità all'interno dello stesso cluster dell'unità i. Il valore b_i è invece la distanza media tra l'unità i e le unità del più vicino degli altri cluster.

In base all'indice Silhouette il numero di cluster ottimale è 7, poiché con 7 cluster abbiamo un valore pari a 0,8132**.** Il valore della silhouette è una misura di quanto simile un oggetto è rispetto al cluster di appartenenza (*cohesion*) comparato ad altri cluster (*separation*). Il range va da -1 a +1, dove un alto valore indica che un oggetto è ben abbinato all'interno della propria cluster e poco abbinato con le cluster circostanti. Se la maggior parte degli elementi hanno un valore elevato allora la configurazione della cluster è appropriata. Se molti elementi hanno valori bassi o negativi ciò significa che la configurazione ha troppi o troppo pochi cluster.

```
$All.index
     4      5      6      7      8      9     10
0.7583 0.7765 0.7993 0.8132 0.7414 0.7557 0.5905

$Best.nc
Number_clusters     Value_Index
         7.0000          0.8132

$Best.partition
 [1] 1 2 1 3 1 1 4 2 1 1 2 1 1 1 1 1 1 1 2 1 2 1 1 1 2 2 2 1 1 4 1 2 4 1 1 5 2 1 2 1 1 1 1 1 1 2 1 2 2 1 1 1 1 2 2 1 4
[59] 1 6 1 2 1 1 1 1 1 1 1 1 1 2 2 2 1 1 2 2 1 1 1 2 1 1 7 1 2 1 4 2
```

Figura 42 analisi effettuata con Nbclust, indice Silhouette

Abbiamo chiesto a Nbclust di utilizzare l'indice Silhouette e questo è il risultato: il numero di cluster ottimale è 7, poiché con 7 cluster abbiamo un valore pari a 0,8132. Con 8 cluster avremmo avuto un valore pari a 0,7414 e con 6 un valore pari a 0,7993. Possiamo chiaramente vedere che anche 6 è un numero accettabile, tuttavia 7 lo migliora. Con 8 invece abbiamo un calo notevole, quindi non va preso in considerazione. Raccogliendo i dati di Nbclust ecco dunque la migliore partizione (numero del gruppo; numerosità del gruppo).

1	2	3	4	5	6	7
56	25	1	5	1	1	1

Ricollegando tali dati agli utenti:

- GRUPPO 1: SergioDL, Be_Italian_, claudioit9cbe, E_Perotti, MichelePovoli, avvocato2punto0, Giovanni_Nappi, crazybalzano, Wasimj96, RaffaelePizzati, andrealion1, Adrian_in_it, LValdrighi, EForenzi, Scugnizzobrigan, LorenzPellico, Ed96webchannel, ApeMag, Antonella180262, BeppeZamperetti, BeppeSama, stemar9288, catzeri3235, SardellaBarbara, elfiegnomi, SandraM_Tcon0, mariateresabru, nicola1691, rugaskipper, SaittaMarta, LaStefi_P, salgiupom, acquodario, Shlomo_75, MarcoM_Marinoni, xetibor, PREDAPPIO98, alexa5313, massimo_limonta, milazzo1987, lucasoldini_93, FranciDiValerio, _LuigiLeonardi, mrmafont, nove_silvia, lagrullaia, LupiMirco, Fabio84V, fbianco91, amarrapese_barr, DragoRosso_, iceflaws, marzia_cappelli, izoon2, MaggioRLIPz, annpn83, Naiandiv
- GRUPPO 2: RossellaFidanza, antgentile, calygola, TheLambkin_, chiaramondi, danbertsamp, GIORGIOBELLINO2, LAtoB__, CurrentiCalamo_, Sandra_AbbDR, MauroBarin, CorrPetrocelli, UffPost, RETWITTATORc, FISRemo, Uleprr, fi_regcampania, Bobbio65M, ParteCivile, LauraGio_75, Mauro5514, NEXITIUS, _Referendum_, quantmint1
- GRUPPO 3: annapaolaconcia,
- GRUPPO 4: ricci_davide77, MauroBeltramo, paceinterra_it, masterofmate, Melania11564076,
- GRUPPO 5: paparcura,
- GRUPPO 6: lega_nord
- GRUPPO 7: Schiforma

Riportiamo per ciascun raggruppamento le medie relative alle variabili di aggregazione

```
> aggregate((twi), list(cutree(hclust(dist(scale(twi), method = "euclidean"), method = "ward.D"), k = 7)), mean)
  Group.1 numero_tweet_id_str produttivita media_user_followers_count media_user_friends_count max_tweet_retweeted_01
1       1                   1     43.64286                  739.8929                 849.500                    0.0
2       2                   1    125.36000                 3032.6400                1858.680                    1.0
3       3                   1     23.00000                73526.0000                2521.000                    0.0
4       4                   2    100.40000                 1857.7000                1792.600                    0.4
5       5                   1     59.00000                73076.0000               64409.000                    0.0
6       6                   3    231.00000                10141.0000                9400.333                    0.0
7       7                   1   4503.00000                  198.0000                   5.000                    0.0
```

Figura 43 medie relative alle variabili di aggregazioni

3.4 L'analisi non gerarchica: il metodo delle k-medie

Applichiamo in ultima analisi l'algoritmo delle *k*-medie, utilizzando come dati di input la matrice dei dati standardizzata e un numero iniziale di centri pari a 7.

Riportiamo dunque i centri ottenuti:

```
  numero_tweet produttivita media_user_followers_count media_user_friends_count max_tweet_retweeted01
1   -0.2523361  0.169441683               -0.039840838               0.17168176           -0.65100655
2    3.5327060  0.002278821                0.008012976               0.15431836            0.07233406
3   -0.2523361 -0.203160341                6.373611739               0.07635049           -0.65100655
4   -0.2523361 -0.207515046               -0.243737408              -0.21331193           -0.65100655
5   -0.2523361  0.008894307               -0.010608061              -0.01935731            1.51901528
6   -0.2523361 -0.128580746                6.332857562               9.01940580           -0.65100655
7   -0.2523361  9.077855929               -0.267326776              -0.28722124           -0.65100655
```

Figura 44 centri ottenuti con 7 gruppi

Il vettore contenente l'allocazione degli elementi:

```
Clustering vector:
 [1] 4 5 4 3 4 4 2 5 4 1 5 4 4 4 1 4 1 4 5 4 5 4 4 4 4 5 5 5 4 4 4 2 1 5 2 4 4 6 5 4 5 4 4 4 4 4 1 4 5 4 5 5 4 4 4 4 5
[56] 5 4 2 4 2 4 5 4 4 4 1 4 4 4 4 4 4 5 5 5 1 4 5 5 4 4 4 5 4 4 7 4 5 4 2 5
```

Figura 45 vettore contenente l'allocazione degli elementi

Per renderlo più comprensibile delimitiamo la dimensione di ognuno dei 7 cluster ottenuti:

```
K-means clustering with 7 clusters of sizes 7, 6, 1, 49, 25, 1, 1
```

Figura 46 dimensione delle cluster

Riportiamo la devianza interna di ciascun gruppo e l'indice $\boldsymbol{R^2}$ pari a 93,4%

```
Within cluster sum of squares by cluster:
[1]  1.7161930 16.9561815  0.0000000  0.7300259 10.1183414  0.0000000  0.0000000
 (between_SS / total_SS =  93.4 %)
```

Figura 47 devianza interna di ciascun gruppo e indice R^2

Per rendere il tutto ancora più comprensibile ricolleghiamo anche i singoli utenti al cluster di appartenenza:

	user_screen_name	cluster
1	SergioDL	4
2	RossellaFidanza	5
3	Be_Italian_	4
4	annapaolaconcia	3
5	claudioit9cbe	4
6	E_Perotti	4
7	ricci_davide77	2
8	antgentile	5
9	MichelePovoli	4
10	avvocato2punto0	1
11	calygola	5
12	Giovanni_Nappi	4
13	crazybalzano	4
14	Wasimj96	4
15	RaffaelePizzati	1
16	andrealion1	4
17	Adrian_in_it	1

18	LValdrighi	4
19	TheLambkin_	5
20	EForenzi	4
21	chiaramondi	5
22	Scugnizzobrigan	4
23	LorenzPellico	4
24	Ed96webchannel	4
25	danbertsamp	5
26	GIORGIOBELLINO2	5
27	LAtoB__	5
28	ApeMag	4
29	Antonella180262	4
30	BeppeZamperetti	4
31	MauroBeltramo	2
32	BeppeSama	1
33	CurrentiCalamo_	5
34	paceinterra_it	2
35	stemar9288	4
36	catzeri3235	4
37	paparcura	6
38	Sandra_AbbDR	5
39	SardellaBarbara	4
40	MauroBarin	5
41	elfiegnomi	4
42	SandraM_Tcon0	4
43	mariateresabru	4
44	nicola1691	4
45	rugaskipper	1
46	SaittaMarta	4
47	CorrPetrocelli	5
48	LaStefi_P	4
49	UffPost	5
50	RETWITTATORc	5
51	salgiupom	4
52	acquodario	4
53	Shlomo_75	4
54	MarcoM_Marinoni	4
55	FISRemo	5
56	Uleprr	5
57	xetibor	4
58	masterofmate	2
59	PREDAPPIO98	4
60	lega_nord	2
61	alexa5313	4
62	massimo_limonta	5
63	milazzo1987	4
64	lucasoldini_93	4
65	FranciDiValerio	4
66	_LuigiLeonardi	1
67	mrmafont	4
68	nove_silvia	4

69	lagrullaia	4
70	LupiMirco	4
71	Fabio84V	4
72	fi_regcampania	5
73	Bobbio65M	5
74	ParteCivile	5
75	fbianco91	1
76	amarrapese_barr	4
77	LauraGio_75	5
78	Mauro5514	5
79	DragoRosso_	4
80	iceflaws	4
81	marzia_cappelli	4
82	NEXITIUS	5
83	izoon2	4
84	MaggioRLIPz	4
85	Schiforma	7
86	annpn83	4
87	_Referendum_	5
88	Naiandiv	4
89	Melania11564076	2
90	quantmint1	5

3.5 Analisi dei gruppi

3.5.1 Silhouette analysis

Abbiamo visto che utilizzando l'indice silhouette Nbclust ha determinato che il numero ottimale di cluster è 7. Analizziamo però da vicino tale risultato.

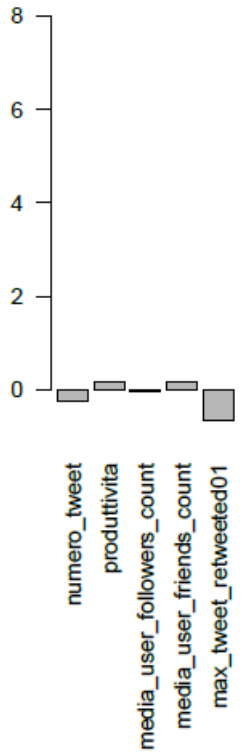
Utilizzeremo la cosiddetta *average silhouette widht*. Essa può assumere, come già detto, un valore compreso fra -1 e +1. Un valore negativo non è desiderabile, poiché ciò corrisponde al caso in cui $\boldsymbol{a_i}$, la distanza media nei confronti dei punti nel cluster, è superiore a $\boldsymbol{b_i}$, la minima distanza media nei confronti dei punti in un altro cluster. Vogliamo che il coefficiente sia positivo ($\boldsymbol{a_i < b_i}$) e per $\boldsymbol{a_i}$ vogliamo che esso sia il più possibile vicino a 0 poiché il coefficiente assume il suo valore massimo,1, quando $\boldsymbol{a_i = 0}$. L'*average silhouette coefficient* si calcola semplicemente facendo la media dei *silhouette coefficient* di tutti i punti appartenenti al cluster (fig. 45). Una misura della bontà di un clustering può essere calcolata calcolando l'*average silhouette coefficient* di tutti i punti. Nel grafico vogliamo che la silhouette sia il più larga possibile. Questo ci permette di distinguere un "taglio pulito" rispetto a cluster "deboli" all'interno dello stesso grafico: cluster con una averege silhouette widht più grande sono più pronunciati. Questo è chiarissimo all'interno della fig. 44: i cluster 2 (0,41), 4 (0,79), e 5 (0,68) sono enormemente più pronunciati dei cluster 1 (0,003), 3 (0,00), 6 (0,00) e 7 (0,00).

Analizziamo più da vicino dunque i cluster 1, 3, 6, 7 riunendo insieme vari dati che abbiamo a disposizione. Per i cluster rimanenti mi limiterò ad analizzare i centroidi (pezzi della fig. 47):

- **CLUSTER 1:** avvocato2punto0, RaffaelePizzati, Adrian_in_it, BeppeSama, rugaskipper, _LuigiLeonardi, fbianco9.
- **CLUSTER 3:** annapaolaconcia,
- **CLUSTER 6:** paparcura
- **CLUSTER 7:** Schiforma

Analisi utenti cluster 1:

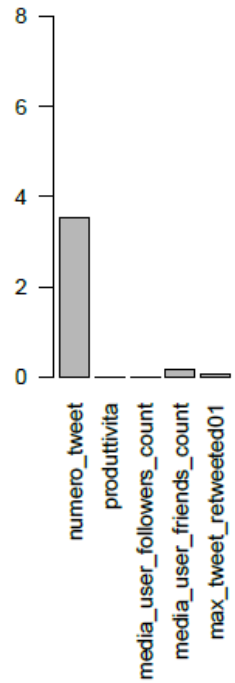
- avvocato2punto0: nella cluster gerarchica messo nel gruppo 1. Silhouette width: 0,019.
- RaffaelePizzati: nella cluster gerarchica messo nel gruppo 1. Silhouette width: 0,272.
- Adrian_in_it: nella cluster gerarchica è stato messo nel gruppo 1. Silhouette width: -0,363.
- BeppeSama: nella cluster gerarchica è stato messo nel gruppo 1. Silhouette width: 0,177.
- Rugaskipper: nella cluster gerarchica è stato messo nel gruppo 1. Silhouette width: -0,21.
- _LuigiLeonardi: nella cluster gerarchica è stato messo nel gruppo 1. Silhouette width: -0,035.
- fbianco91: nella cluster gerarchica è stato messo nel gruppo 1. Silhouette width: 0,199.



Decisamente un cluster interessante. Possiamo subito vedere dopo una analisi più approfondita che è caratterizzato esclusivamente da persone fisiche. Non abbiamo alcun partito, alcun giornale, alcuna associazione o fondazione. Quello che mi è balzato subito all'occhio è che il valore relativo al numero di tweet

totali, una volta standardizzato, è enormemente inferiore rispetto alla produttività. Stiamo parlando dunque di persone che si sono mobilitate appositamente per la campagna, con un numero di amici superiore al numero di follower, bassissimo numero di retweet.

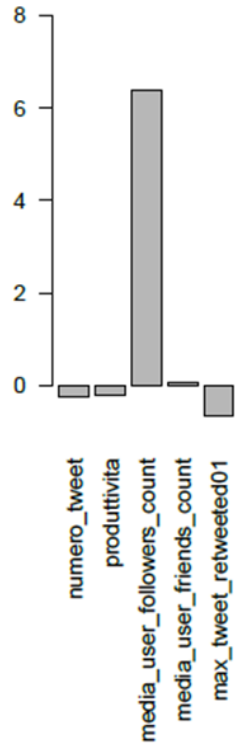
Analisi cluster 2:



Come abbiamo già visto questa cluster ha una mediocre silhouette widht. Passando all'analisi dei centroidi anche questo gruppo è interessante: numero di tweet enormemente più alto della produttività. Siamo dunque di fronte a gente che non si è mobilitata appositamente per la campagna. Numero di amici superiore al numero di follower, alto numero di retweet.

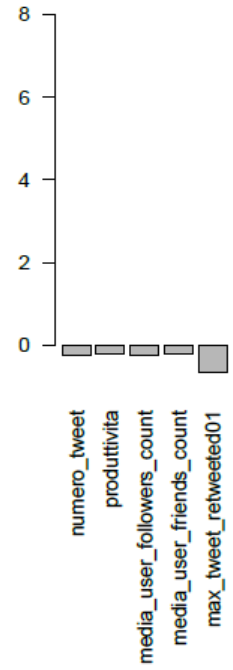
Analisi utenti cluster 3:

- annapaolaconcia: nella cluster gerarchica è stato messo nel gruppo 3 (da sola). Silhouette width: 0,000.



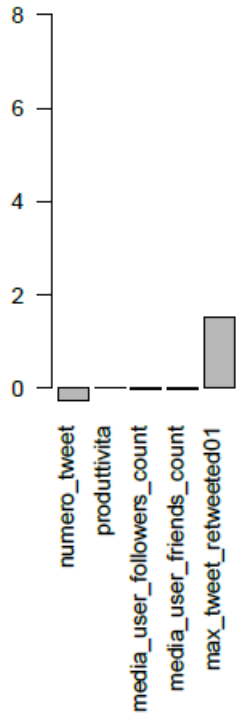
Bassissimo numero di tweet, bassa produttività, altissimo numero di follower, bassissimo numero di amici. Bassissimo numero di retweet. Siamo infatti di fronte ad un politico. Non è la persona comune che poi influenzerà i suoi amici e conoscenti.

Analisi utenti cluster 4:



Qui tutti gli indicatori sono bassi.

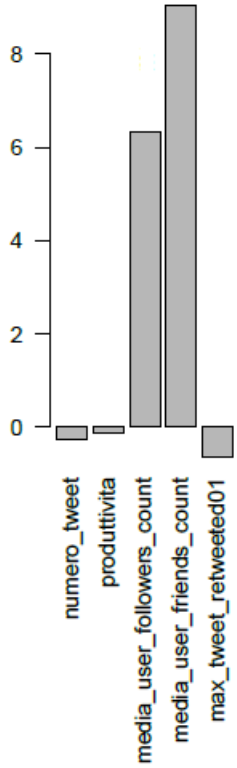
Analisi utenti cluster 5:



Cluster decisamente interessante. Numero di tweet inferiore rispetto alla produttività durante la campagna. Follower e amici quasi uguali (e il numero è basso). Altissimo numero di retweet (il più alto fra tutti). Potremmo essere di fronte al gruppo con più influenza nel mondo reale.

Analisi utenti cluster 6:

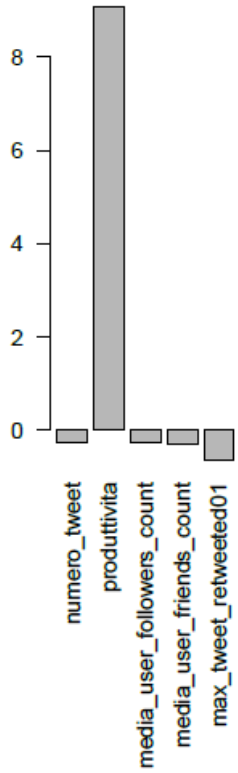
- paparcura: nella cluster gerarchica è stato messo nel gruppo 5 (da solo). Silhouette width: 0,000.



Bassissimo numero di tweet, bassa produttività. Alto numero di follower, numero di amici decisamente superiore. Bassissimo numero di retweet.

Analisi utenti cluster 7:

- Schiforma: nella cluster gerarchica è stato messo nel gruppo 7 (da solo). Silhouette width: 0,000. Produttività: 4503. media_user_followers_count: 198. media_user_friends_count: 5.



Basso numero di tweet, enorme produttività durante la campagna, basso numero di follower, basso numero di amici, basso numero di retweet. Guardando il profilo si legge che l'utente Schiforma ha generato il suo profilo appositamente osteggiare la riforma elettorale. Dubito abbia avuto una influenza pesante nella vita reale.

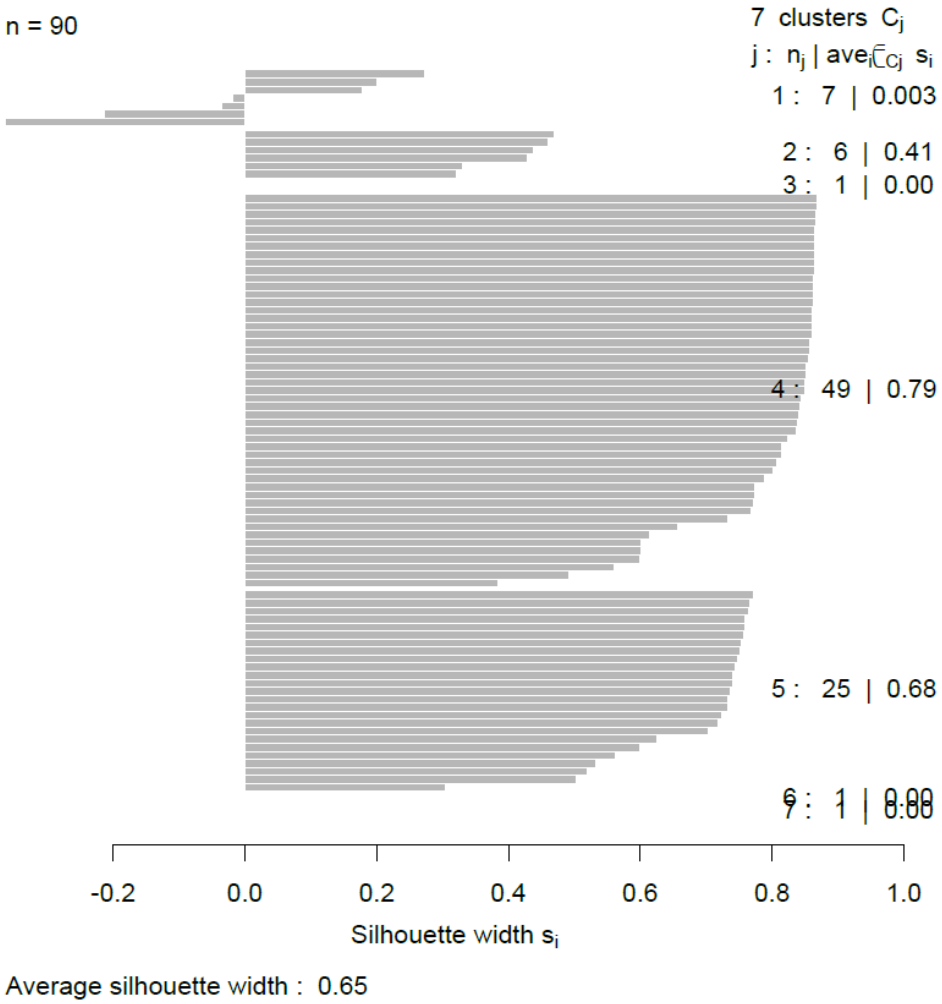


Figura 48 grafico silhouette

	cluster	neighbor	sil_width
[1,]	4	1	0.59911120
[2,]	5	1	0.51874118
[3,]	4	1	0.65644690
[4,]	3	1	0.00000000
[5,]	4	1	0.86527944
[6,]	4	1	0.85748897
[7,]	2	1	0.45994426
[8,]	5	1	0.50226544
[9,]	4	1	0.86664044
[10,]	1	4	-0.01903641
[11,]	5	1	0.59888717
[12,]	4	1	0.80680270
[13,]	4	1	0.86779125
[14,]	4	1	0.86871397
[15,]	1	4	0.27296196
[16,]	4	1	0.82473551
[17,]	1	4	-0.36354876
[18,]	4	1	0.85128977
[19,]	5	1	0.56266834
[20,]	4	1	0.38316653
[21,]	5	4	0.75653125
[22,]	4	1	0.78781475
[23,]	4	1	0.76775922
[24,]	4	1	0.84104787
[25,]	5	4	0.75323234
[26,]	5	4	0.73385849
[27,]	5	4	0.73384578
[28,]	4	1	0.86284337
[29,]	4	1	0.60172372
[30,]	4	1	0.85790974
[31,]	2	5	0.33032395
[32,]	1	4	0.17747269
[33,]	5	4	0.75983879
[34,]	2	4	0.46983916
[35,]	4	1	0.60151799
[36,]	4	1	0.56024843
[37,]	6	3	0.00000000
[38,]	5	4	0.74813324
[39,]	4	1	0.73378627
[40,]	5	1	0.53163789
[41,]	4	1	0.49204852
[42,]	4	1	0.77294522
[43,]	4	1	0.86436753
[44,]	4	1	0.86466378
[45,]	1	4	-0.21351808
[46,]	4	1	0.77427926
[47,]	5	4	0.74329539
[48,]	4	1	0.86244799
[49,]	5	4	0.75961847
[50,]	5	1	0.30370559
[51,]	4	1	0.85156471
[52,]	4	1	0.84168368
[53,]	4	1	0.84914481
[54,]	4	1	0.86246148
[55,]	5	4	0.76482350
[56,]	5	1	0.62478299
[57,]	4	1	0.86430513
[58,]	2	4	0.43769360
[59,]	4	1	0.83941697
[60,]	2	1	0.42910367
[61,]	4	1	0.77183649
[62,]	5	4	0.76582267
[63,]	4	1	0.86130629
[64,]	4	1	0.86270923
[65,]	4	1	0.86086582
[66,]	1	4	-0.03505778
[67,]	4	1	0.86055817
[68,]	4	1	0.80173071
[69,]	4	1	0.86163058
[70,]	4	1	0.84403875
[71,]	4	1	0.85460484
[72,]	5	4	0.75058216
[73,]	5	4	0.77166483
[74,]	5	4	0.70255302
[75,]	1	4	0.19956056
[76,]	4	1	0.81547576
[77,]	5	4	0.71809504
[78,]	5	4	0.72421849
[79,]	4	1	0.83737450
[80,]	4	1	0.81458490
[81,]	4	1	0.85062696
[82,]	5	4	0.74097051
[83,]	4	1	0.61500447
[84,]	4	1	0.86614101
[85,]	7	1	0.00000000
[86,]	4	1	0.86416638
[87,]	5	4	0.73739187
[88,]	4	1	0.86454451
[89,]	2	5	0.32078062
[90,]	5	4	0.74121212

Figura 49 analisi punti silhouette

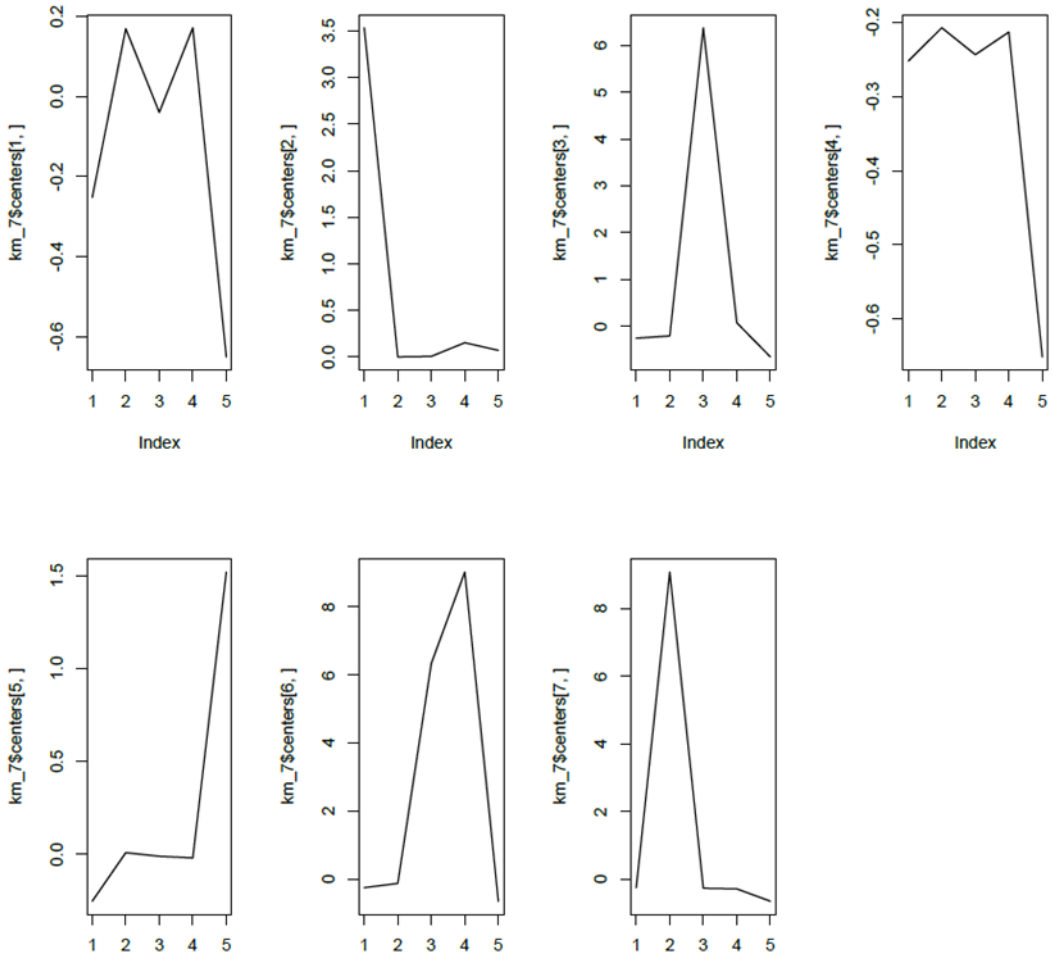


Figura 50 grafico centroidi a linee

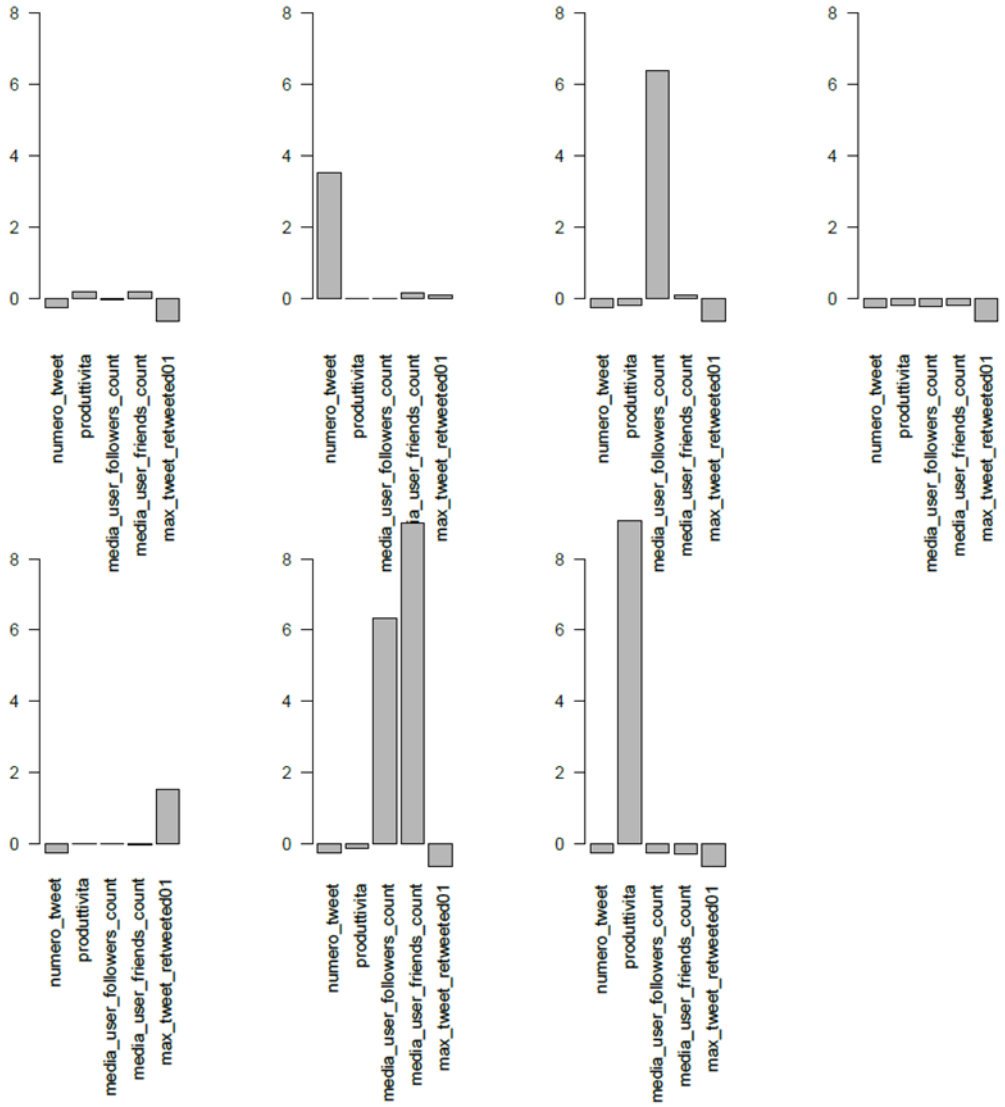


Figura 51 grafico centroidi a barre

Tramite l'analisi degli hashtag siamo riusciti a individuare cosa sostenevano i vari utenti (fig. 48).

CLUSTER 1: quasi eslusivamente favorevole al sì.

CLUSTER 2: esclusivamente favorevole al no.

CLUSTER 3: esclusivamente favorevole al sì.

CLUSTER 4: metà sì e metà no.

CLUSTER 5: larga maggioranza no.

CLUSTER 6: esclusivamente sì.

CLUSTER 7: esclusivamente no.

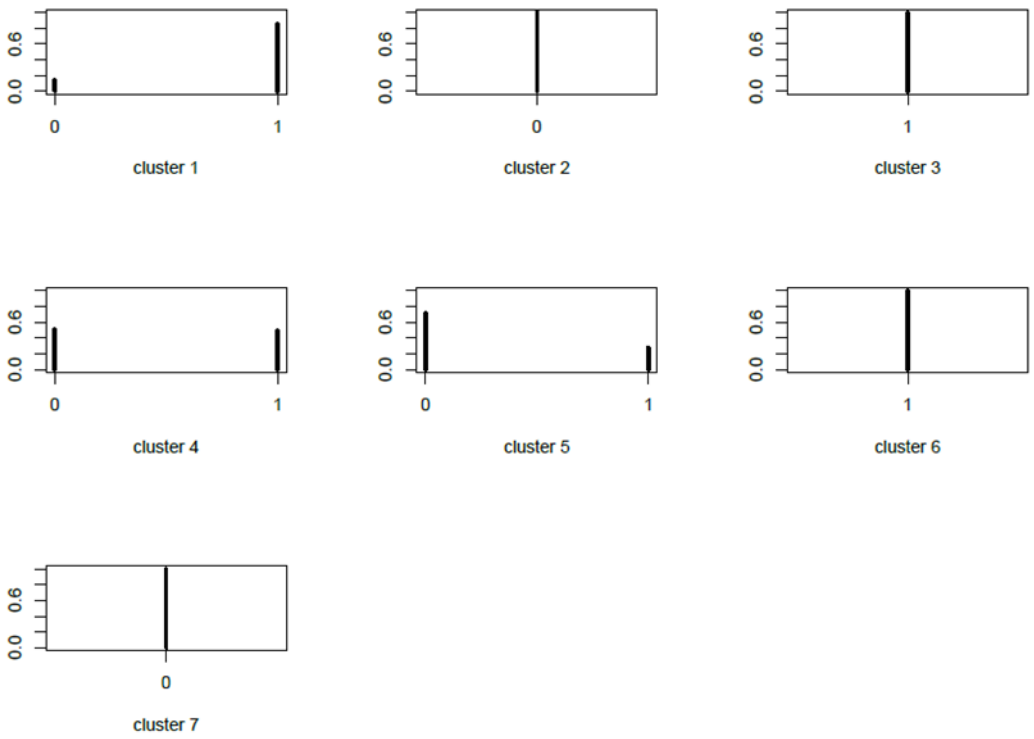


Figura 52 sì o no

3.5.2 Alluvial plot

Costruiamo ora un alluvial plot (fig. 49) che mette in relazione gli utenti e i loro retweet tramite i dati del data frame della fig. 50.

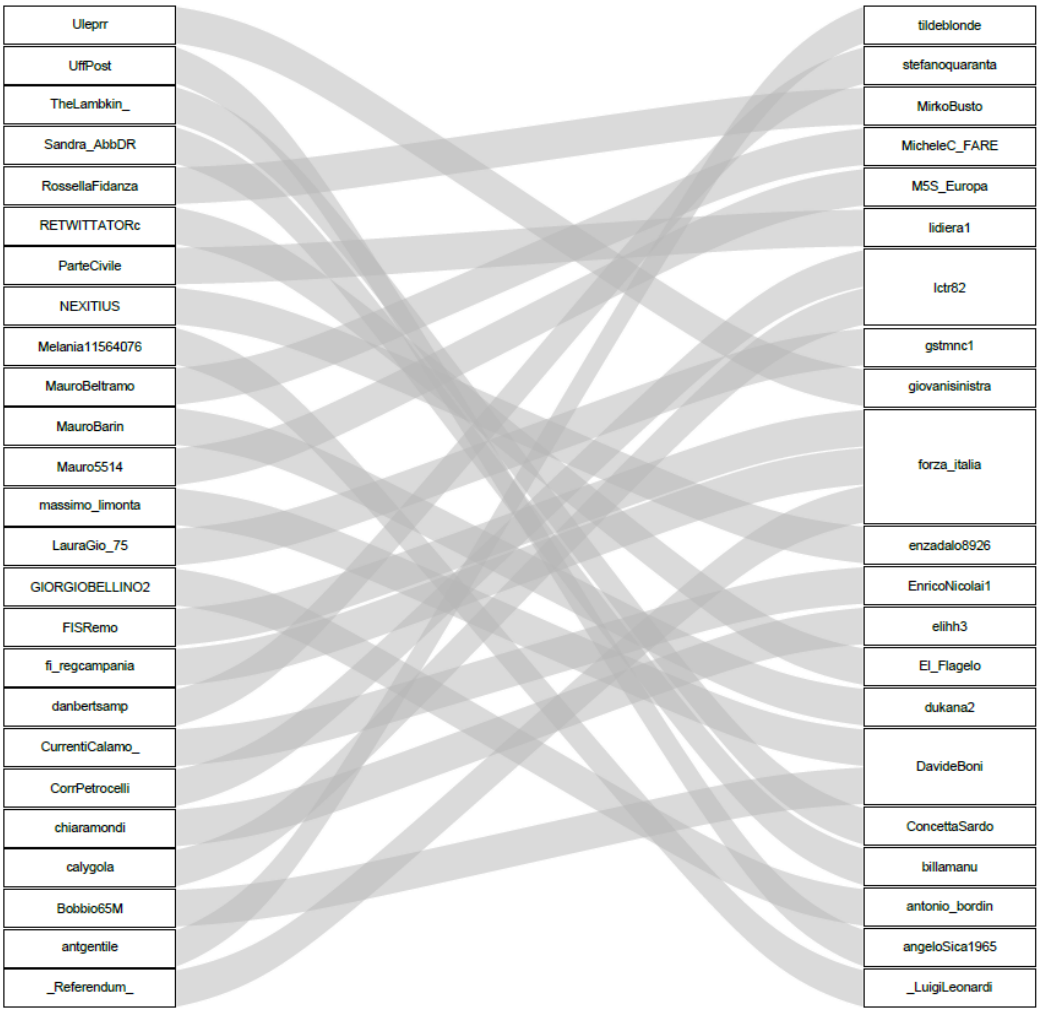


Figura 53 alluvial plot dei retweet

	user_screen_name	tweet_retweeted_status_user_screen_name
1	Sandra_AbbDR	billamanu
2	Mauro5514	M5S_Europa
3	RossellaFidanza	MirkoBusto
4	UffPost	angeloSica1965
5	NEXITIUS	enzadalo8926
6	antgentile	tildeblonde
7	TheLambkin_	ConcettaSardo
8	GIORGIOBELLINO2	antonio_bordin
9	fi_regcampania	forza_italia
10	_Referendum_	forza_italia
11	FISRemo	forza_italia
12	ParteCivile	lidiera1
13	Melania11564076	_LuigiLeonardi
14	massimo_limonta	DavideBoni
15	Bobbio65M	DavideBoni
16	MauroBarin	dukana2
17	LauraGio_75	gstmnc1
18	RETWITTATORc	El_Flagelo
19	chiaramondi	elihh3
20	CurrentiCalamo_	EnricoNicolai1
21	Uleprr	giovanisinistra
22	CorrPetrocelli	lctr82
23	calygola	lctr82
24	MauroBeltramo	MicheleC_FARE
25	danbertsamp	stefanoquaranta

Figura 54 data frame alluvial plot

3.5.3 Word Cloud

Costruiamo ora due diverse categorie di word cloud.

Word cloud degli hashtag:

- Word cloud generale
- Word cloud del no
- Word cloud del sì

Possiamo subito vedere che coloro che hanno sostenuto il no fanno un uso molto più ampio degli hashtag.

Word cloud delle parole dei tweet:

- Parole relative al sì
- Parole relative al no

Figura 55 word cloud di tutti gli hashtag

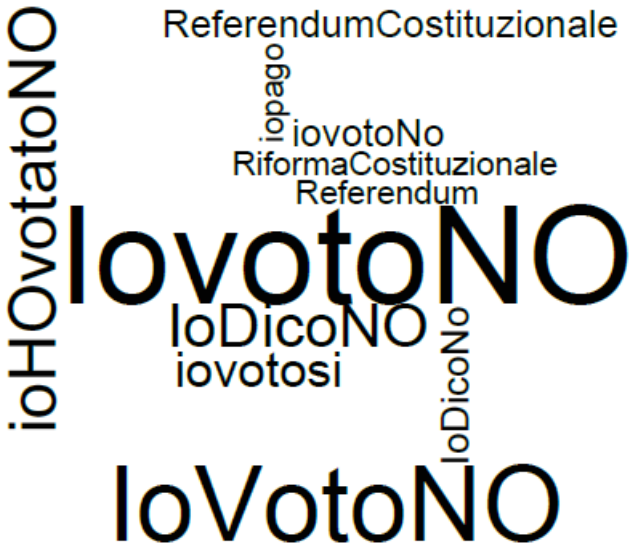


Figura 56 word cloud del no

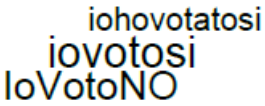


Figura 57 word cloud del sì

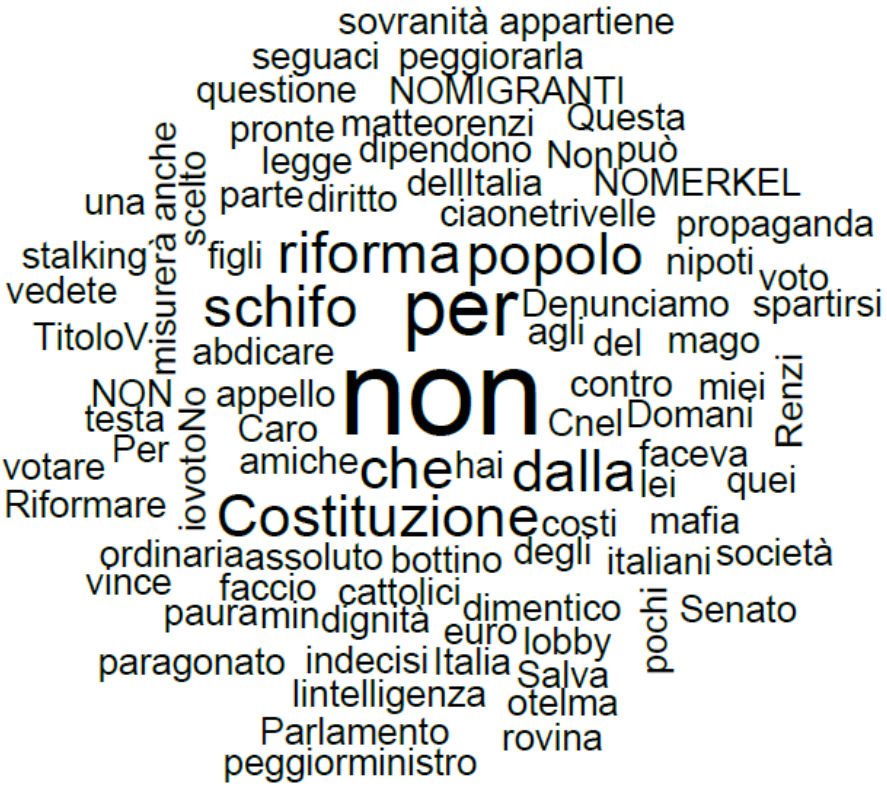


Figura 58 tweet no

Figura 59 tweet sì

Possiamo vedere come fra coloro che abbiano votato NO abbondino riferimenti a messaggi anti UE, mentre fra coloro che hanno votato SI sono frequenti le parole cambiare e cambiamento.

Conclusioni

In base a quanto detto fino ad ora posso finalmente identificare le caratteristiche del brand advocate politico a cui bisogna mirare su Twitter, la persona "comune" il cui messaggio è in grado di mobilitare amici e parenti:

- Il numero dei follower deve essere pari al numero degli amici, oppure il numero degli amici deve essere pari a circa i 2/3 del numero dei follower.
- Deve avere pochi amici e follower (dobbiamo mirare il più possibile alla persona comune)
- Usa il proprio nome
- Ha scritto pochi messaggi dal momento dell'attivazione del profilo.
- Non scrive troppi messaggi. È facile diventare noiosi su internet.
- Si mobilita solamente durante la campagna (quindi dobbiamo vedere che la media di messaggi scritti durante l'arco temporale della campagna sia superiore rispetto alla media di messaggi scritti sin dall'attivazione del profilo). Questo perché in questa maniera si evita che l'utente risulti "pedante" (se la cerchia di amici e parenti lo reputa "pedante" non leggerà i suoi messaggi con attenzione). L'attenzione è un bene prezioso su internet che va centellinato.
- Retwetta. Può retwettare messaggi di politici, giornali ai suoi amici (attiva così il two step flow of communication). Oppure può saltare questo passo e retweettare direttamente messaggi di amici. Se i due hanno ad esempio una cerchia di amici in comune è più probabile che il messaggio dell'amico venga ascoltato dagli altri amici.
- Usa molti hashtag. Gli hashtag aiutano a dare visibilità ai messaggi (specialmente se si usa un trending hashtag) e aiutano subito a "etichettare" il messaggio (sappiamo alla prima occhiata il punto su cui si concentrerà).
- Usa parole triviali, senza però abusarne. Come già detto su internet c'è una soglia di attenzione bassa. Messaggi brevi e divertenti. Se accompagnati da foto divertenti (ad esempio meme) aiutano ancora meglio a catturare l'attenzione.

Il gruppo che più si avvicina a questo è il gruppo numero 5 (che sostiene in larga maggioranza il no, la scelta che ha vinto).

Appendice: il software R

Nello svolgimento della cluster analysis è stato impiegato il softtware R, nello specifico Rstudio. Rstudio è particolarmente comodo poiché permette di importare documenti di vari formati (nel mio caso ho utilizzato file .txt e file Excel con estensione .xlsx).

Nella prima fase per comodità ho utilizzato un documento .xlsx con le variabili numero_tweet_id_str, produttivita, media_user_followers_count, media_user_friends_count, max_tweet_retweeted_01.

Per prima cosa apriamo la matrice dati (il documento è stato rinominato twitter).

twitter <- read_excel("twitter.xlsx")

I dati vengono standardizzati utilizzando il comando scale.

Per effettuare l'analisi gerarchica dobbiamo prima assegnare il tipo di distanza che intendiamo utilizzare e poi il tipo di legame.

Y<- scale(twitter)

Le varie distanze:

- Euclidean, per la distanza euclidea
- Maximum, per la distanza di Lagrange
- Manhattan per la distanza di Manhattan
- Canberra, per la distanza di Canberra
- Binary, per la distanza di Jaccard

d<- Dist(y, method = "").

I vari legami:

- Single, per il metodo del legame singolo
- Average, per il metodo del legame medio
- Centroid, per il metodo del centroide
- Ward.D, per il metodo di Ward.

Il comando risulta:

hc<- hclust(d, method = "")

hc$merge: evidenzia le aggregazioni avvenute per ogni fase del processo di aggregazione. Gli elementi negativi rappresentano singole unità, mentre quelli positivi rappresentano cluster già formati.

Hc$height: restituiscono un vettore che esprime la distanza alla quale è avvenuta la fusione in ciascuna fase del processo di agglomerazione.

Per visualizzare il dendrogramma si usa il comando plot:

plot(hc).

È possibile tagliare il dendrogramma tramite il comando cutree decidendo il numero di cluster che si vuole ottenere oppure definendo una specifica altezza.

Id<- Cutree(hc, k =NULL, h = NULL)

K rappresenta il numero di cluster (bisogna scrivere il numero che si vuole ottenere al posto di NULL). H rappresenta l'altezza a cui lo si vuole tagliare.

Il comando list ci permette di avere una lista degli elementi di ciascun cluster.

id<- list(id)

Per determinare il numero di gruppi è stato utilizzato il pacchetto chiamato NbClust. L'indice utilizzato è l'indice Silhouette.

Bisogna inserire il comando seguente:

nb_silhouette<-
NbClust(scale(twi),distance="euclidean",min.nc=4,max.nc=10,method="ward.D",index="silhouette")

Sottolineiamo gli elementi importanti della riga di comando.

NbClust::NbClust(scale(twi),distance="euclidean",min.nc=4,max.nc=10,method="ward.D",index="silhouette")

min.nc=4 indica che vogliamo ottenere come minimo 4 cluster

max.nc=10 indica che vogliamo ottenere come massimo 10 cluster

index="silhouette" abbiamo specificato che intendiamo utilizzare come indice solamente l'indice Silhouette.

Per ottenere la migliore partizione inseriamo il comando seguente:

table(nb_silhouette$Best.partition)

ora sappiamo dove tagliare il dendrogramma

Per effettuare l'analisi delle k-medie si riparte dalla matrice standardizzata originale, su cui si applica il comando:

kmeans(y, centers)

Centers rappresenta il numero di cluster che si vuole ottenere.

Per visualizzare i risultati prodotti si utilizzano i seguenti comandi:

km<- kmeans(y, centers)

km$cluster: restituisce un vettore di allocazione che assegna ciascuna unità al cluster di appartenenza

km$center : restuisce la matrice dei centroidi

km$withins : evidenzia la devianza interna di ciascun gruppo

km$size : evidenzia le dimensioni dei gruppi

grafico Silhouette:

km_7<-kmeans(scale(twi), 7)

silhouette(x = km_7$cluster, dist = dist(scale(twi)))

Facciamo ora il grafico dei centroidi di km_7

Scarichiamo il pacchetto e1071 che ci permette di effettuare Fuzzy clustering utilizzando la funzione cmeans .

cmt<-cmeans(scale(twi), 7)

Il numero 7 indica il numero di cluster.

centers<-cmt$centers

questi sono i centri finali delle cluster.

Disegniamo ora il grafico. Dobbiamo creare una multi-paneled plotting window.

Tramite il comando par(mfrow) chiediamo di sistemare le figure in due righe e tre colonne:

par(mfrow=c(2,3))

Creiamo ora i grafici da sistemare:

> plot(centers[“”,],type="l",ylim=c(-1,1))

- centers[“”,] al posto delle virgolette dobbiamo mettere il numero del cluster.
- type="l" la l sta per linee
- ylim=c(-1,1) indica il range dell’asse delle y

Lavoriamo ora sulla silhouette:

utilizziamo la funzione silhouette {cluster} che ci permette di calcolare le informazioni relative alla silhouette di un dato cluster in k cluster:

> silhouette(km_7$cluster,dist(scale(twi)))

Ora costruiamo un grafico per vedere la frequenza dei sì e dei no all’interno dei cluster.

Dobbiamo aprire il dataset originale:

twitter <- read.delim("E:/twitter.txt")

utilizziamo la funzione subset per creare 7 sottogruppi, uno per cluster

```
gruppo1<-subset(twitter,cluster==1)

gruppo2<-subset(twitter,cluster==2)

gruppo3<-subset(twitter,cluster==3)

gruppo4<-subset(twitter,cluster==4)

gruppo5<-subset(twitter,cluster==5)

gruppo6<-subset(twitter,cluster==6)

gruppo7<-subset(twitter,cluster==7)
```

utilizziamo ora la funzione già incontrata prima:

```
par(mfrow=c(3,3))
```

costruiamo ora il grafico:

```
> plot(table(gruppo1$scelta01))

> plot(table(gruppo2$scelta01))

> plot(table(gruppo3$scelta01))

> plot(table(gruppo4$scelta01))

> plot(table(gruppo5$scelta01))

> plot(table(gruppo6$scelta01))

> plot(table(gruppo7$scelta01))
```

Per fare l'alluvial plot dei retweet utilizziamo il pacchetto alluvial .

Dobbiamo generare una matrice su cui lavorare. Il mio documento si chiama "alluvial" ed è la tabella della fig. 50. Dobbiamo fornire il data frame come primo argomento nella riga di comando e un vettore di frequenze nell'argomento freq.

Forniamo il data frame:

```
> alluvial <- read.csv("H:/alluvial/alluvial.txt", sep="")
```

> View(alluvial)

Chiamando il data frame “ma” nella riga che segue:

alluvial(ma, freq=c(rep(1,25)), cex=0.5)

freq è il vettore di frequenza citato all’inizio.

wordcloud:

abbiamo diviso il lavoro in tre fasi:

- tutti gli hashtag
- hashtag favorevoli al sì
- hashtag favorevoli al no

TUTTI GLI HASHTAG:

```
hashtag<-c("IoDicoNO" ,"IoVotoNO", "ReferendumCostituzionale"
,"IoDicoNO" , "IoVotoNO", "referendumcostituzionale", "M5S"
,"IoDicoNO" , "IoVotoNO", "referendumcostituzionale", "matteorenzi"
,"IoDicoNO" , "referendumcostituzionale"
,"IoDicoNO" , "ReferendumCostituzionale"
,"IoDicoNO" , "referendumcostituzionale"
,"IoDicoNO" , "riformacostituzionale"
,"IoDicoNO" , "VotoNo", "referendumcostituzionale"
,"ioHOvotatoNO", "ciaone", "referendumcostituzionale"
,"ioHOvotatoNO", "IoVotoNO", "IoDicoNo", "referendumcostituzionale", "Referendum"
,"ioHOvotatoNO", "Referendum", "referendumcostituzionale"
,"ioHOvotatoNO", "referendumcostituzionale"
,"ioHOvotatoNO", "referendumcostituzionale"
,"ioHOvotatoNO", "referendumcostituzionale", "IoVotoNO"
,"ioHOvotatoNO", "referendumcostituzionale", "Referendum", "matitacancellabili"
,"ioHOvotatoNO", "ReferendumCostituzionale", "RiformaCostituzionale"
,"ioHOvotatoNO", "RENZIACASA", "referendumcostituzionale"
,"iohovotatoNO", "RiformaCostituzionale", "ReferendumCostituzionale"
,"iohovotatosi", "bastaunsi", "iovotosi"
,"iohovotatosi ", "bastaunsi", "iovotosi"
,"iohovotatosi"," iovotosi"
,"iohovotatosi", "iovotosi", "iodicosi", "referendumcostituzionale"
,"ionondimentico", "RiformaCostituzionale", "referendumcostituzionale", "IoVotoNO"
,"iopago", "IovotoNO"
,"iopago","IovotoNO"
,"iopago","IovotoNO"
,"iostoconmarino", "IoVotoNO", "referendumcostituzionale", "noiNO", "NOncosi", "ParteCivile"
,"iovoto", "referendumcostituzionale", "riformacostituzuonale"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IoVotoNO", "èNo","referendumcostituzionale", "fateveneunaragione"
,"iovotoNo","IlmioNOëDiverso"
,"iovotoNo", "IoDicoNo"
,"iovotoNo"," iodicono"
,"IoVotoNO", "IoDicoNo", "iovotosi", "iodicosi"
,"iovotono", "iodicono", "referendumcostituzionale", "4dicembre"
,"iovotonoalreferendumcostituzionale"
,"IoVotoNO", "IoDicoNo", "referendumcostituzionale", "referendum"
,"IOVOTONO", "IoVotoNO", "IoHoVotatoNO", "referendumcostituzionale"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"iovotoNo", "stavoltaNO"
,"iovotosi", "iovotoNo"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi")
```

```
hashtag2<-c("iodicono", "iovotosi" , "4dicembre"," bastaunsi" , "ciaone",  "èNo" ,
"fateveneunaragione",                    "IlmioNOèDiverso",                          "iodicono",
                        " IoDicoNo",                          " IoDicoNO",                          " iodicosi",
"iohovotatoNO",                      " ioHOvotatoNO",                    "IoHoVotatoNO",
"iohovotatosi",                      "iohovotatosi",                     "ionondimentico",
"iopago",                     " iostoconmarino",                           "iovoto",

                        " iovotono",                            "iovotoNo",                            "IovotoNO",
"IoVotoNO",                             " IOVOTONO", "iovotonoalreferendumcostituzionale",
"iovotosi" ,                                 "M5S",                 "matitacancellabili" ,
                      "matteorenzi" ,                            "noiNO" ,                           "NOncosì"
                      "ParteCivile",                          "referendum",                        "Referendum",
"referendumcostituzionale",           "ReferendumCostituzionale",                       "RENZIACASA",
"riformacostituzionale",              "RiformaCostituzionale",             "riformacostituzuonale",
"stavoltaNO",                             "VotoNo")
```

HASTAG FAVOREVOLI AL NO:

```
hashtag_no<-c("IoDicoNO" ,"IoVotoNO", "ReferendumCostituzionale"
,"IoDicoNO" , "IoVotoNO", "referendumcostituzionale", "M5S"
,"IoDicoNO" , "IoVotoNO", "referendumcostituzionale", "matteorenzi"
,"IoDicoNO" , "referendumcostituzionale"
,"IoDicoNO" , "ReferendumCostituzionale"
,"IoDicoNO" , "referendumcostituzionale"
,"IoDicoNO" , "riformacostituzionale"
,"IoDicoNO" , "VotoNo", "referendumcostituzionale"
,"ioHOvotatoNO", "ciaone", "referendumcostituzionale"
,"ioHOvotatoNO", "IoVotoNO", "IoDicoNo", "referendumcostituzionale", "Referendum"
,"ioHOvotatoNO", "Referendum", "referendumcostituzionale"
,"ioHOvotatoNO", "referendumcostituzionale"
,"ioHOvotatoNO", "referendumcostituzionale"
,"ioHOvotatoNO", "referendumcostituzionale", "IoVotoNO"
,"ioHOvotatoNO", "referendumcostituzionale", "Referendum", "matitacancellabili"
,"ioHOvotatoNO", "ReferendumCostituzionale", "RiformaCostituzionale"
,"ioHOvotatoNO", "RENZIACASA", "referendumcostituzionale"
,"iohovotatoNO", "RiformaCostituzionale", "ReferendumCostituzionale"
,"ionondimentico", "RiformaCostituzionale", "referendumcostituzionale", "IoVotoNO"
,"iopago", "IovotoNO"
,"iopago","IovotoNO"
,"iopago","IovotoNO"
,"iostoconmarino", "IoVotoNO", "referendumcostituzionale", "noiNO", "NOncosì", "ParteCivile"
,"iovoto", "referendumcostituzionale", "riformacostituzuonale"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IovotoNO"
,"IoVotoNO", "èNo","referendumcostituzionale", "fateveneunaragione"
,"iovotoNo","IlmioNOèDiverso"
,"iovotoNo", "IoDicoNo"
,"iovotoNo"," iodicono"
,"IoVotoNO", "IoDicoNo", "iovotosi", "iodicosi"
,"iovotono", "iodicono", "referendumcostituzionale"
,"4dicembre","iovotonoalreferendumcostituzionale"
,"IoVotoNO", "IoDicoNo", "referendumcostituzionale", "referendum"
,"IOVOTONO"," IoVotoNO", "IoHoVotatoNO", "referendumcostituzionale"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"iovotosi", "iovotoNo")
```

```
hashtag_no2<-c(" iodicono",                          "IoVotoNO",                          "iovotosi",
"4dicembre",                          "ciaone",                                  "èNo",
"fateveneunaragione",                 "IlmioNOèDiverso",                          "iodicono",
"IoDicoNo",                         "IoDicoNO",                          "iodicosi" ,
"iohovotatoNO",                     "ioHOvotatoNO" ,                     "IoHoVotatoNO",
"ionondimentico" ,                          "iopago"  ,                 "iostoconmarino" ,
"iovoto" ,                        "iovotono"     ,                    "iovotoNo" ,
"IovotoNO" ,                        "IoVotoNO"      ,                    "IOVOTONO" ,
"Iovotonoalreferendumcostituzionale",                          "iovotosi"   ,                                "M5S" ,
"matitacancellabili" ,                      "matteorenzi" ,                                "noiNO",
"NOncosì"   ,                  "ParteCivile",                          "referendum" ,
"Referendum" ,          "referendumcostituzionale" ,          "ReferendumCostituzionale" ,
"RENZIACASA"  ,            "riformacostituzionale"  ,            "RiformaCostituzionale" ,
"Riformacostituzuonale",                            "VotoNo")
```

HASHTAG FAVOREVOLI AL SI:

```
hashtag_si<-c("iohovotatosi", "bastaunsi", "iovotosi"
,"iohovotatosi", "bastaunsi", "iovotosi"
,"iohovotatosi"," iovotosi"
,"iohovotatosi", "iovotosi", "iodicosi", "referendumcostituzionale"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"IoVotoNO"," iovotosi"
,"iovotoNo", "stavoltaNO"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi"
,"iovotosi")
```

```
hashtag_si2<-c("iovotosi",              "bastaunsi" ,               "iodicosi" ,           "iohovotatosi" ,
"iovotoNo",                "IoVotoNO",                "iovotosi referendumcostituzionale",
"stavoltaNO")
```

Con le parole invece come prima cosa:

VOTO NO:

creiamo t_si, ci da la frequenza delle parole nei tweet

> t_si<-termFreq(tweet_text_si,control=ctrl)

> wf_si <- data.frame(word=names(t_si), freq=t_si)

> wordcloud(wf_si$word,wf_si$freq,random.order=FALSE,min.freq=1)

Stessa cosa per il no e per le parole in generale.

Bibliografia

CAIAZZO, D., COLAIANNI, A., FEBBRAIO, A., MASI, D., (2009). *Buzz marketing nei social media. Come scatenare il passaparola on-line*, Fausto Lupetti Editore

SORICE, M., (2009) *Sociologia dei mass media*. Carocci editore

PACCAGNELLA, L., (2004) *Sociologia della Comunicazione*. il Mulino

KLAPPER, J. T., (1960). *Effects of Mass Communication*

DE FLEUR, M. L., BALE-ROKEACH, S., (1989). *Theories of Mass Communication*

CAIAZZO, D., COLAIANNI, A., FEBBRAIO, A., LISIERO, U., (2009). *Buzz marketing nei social media*

LASN, K., *Culture Jam: The Uncooling of America*, (1999). Eagle Brook

HARLOW, S. HARP, D., (2012). *Collective action on the Web: A cross-cultural study of social networking sites and online and offline activism in the United States and Latin America. Information, Communication & Society*

KARPF, D., (2010). *Online political mobilization from the advocacy group's perspective: Looking beyond clicktivism. Policy & Internet*

REBER, B. H., KIM, J. K. (2006). *How activist groups use websites in media relations: evaluating online press rooms*. Journal of Public Relations Research

DE BLASIO, E., QUARANTA, M., SANTANIELLO, M., SORICE, M., (2017). *Media, politica e società: le tecniche di ricerca*

Sitografia

http://www.glossariomarketing.it/significato/word-of-mouth/

http://www.unipd.it/ilbo/content/il-mondo-%E2%80%9Cpiccolissimo%E2%80%9D-dei-social-network

http://www.glossariomarketing.it/significato/opinion-leader/

http://www.glossariomarketing.it/significato/public-relations/

http://www.glossariomarketing.it/significato/brand-advocate/

http://www.zuberance.com/downloads/brandAdvocateInsights.pdf

http://www.nielsen.com/us/en/newswire/2012/consumer-trust-in-online-social-and-mobile-advertising-grows.html

https://www.weforum.org/agenda/2016/08/hillary-clinton-or-donald-trump-winning-on-twitter/

http://www.stirista.com/wpcontent/uploads/2016/06/WhosFollowingTrumpAndClinton-1.pdf

https://www.alexa.com/siteinfo/reddit.com

https://www.gruppodigitouch.it/servizi/amplification/social-media-content/

https://www.youtube.com/watch?v=MKH6PAoUuD0

https://www.nytimes.com/2016/11/20/opinion/sunday/reddit-and-the-god-emperor-of-the-internet.html?mcubz=3

https://www.dailydot.com/layer8/donald-trump-inauguration-donations-crowdsourced-journalism-reddit-twitter/

https://www.dariovignali.net/marketing-politico-ed-elettorale/

http://www.repubblica.it/speciali/esteri/presidenziali-usa2016/2016/11/12/news/trump_internet_meme_virali_social_4chan-151826943/

http://thehill.com/homenews/campaign/334897-poll-majority-says-mainstream-media-publishes-fake-news

https://www.tvdigitaldivide.it/2017/09/15/audiweb-32-mln-gli-italiani-online-a-luglio-2017/

http://www.audiweb.it/dati_it/total-digital-audience_it/

http://www.juliusdesign.net/28700/lo-stato-degli-utenti-attivi-e-registrati-sui-social-media-in-italia-e-mondo-2015/

http://www.censis.it/7?shadow_comunicato_stampa=121073

https://www.wired.it/internet/social-network/2016/03/08/italiani-social-media/

http://www.ilsole24ore.com/art/notizie/2017-09-28/su-facebook-testa-testa-grillo-e-salvini-doppiato-renzi-che-si-rifa-twitter-091110.shtml

http://www.ilpost.it/2015/10/09/account-italiani-piu-seguiti-su-twitter/

http://www.ilpost.it/2016/10/10/account-italiani-piu-seguiti-twitter-2/

https://www.youtube.com/watch?v=CTvzvyy3EIk

https://www.oreilly.com/ideas/tweets-loud-and-quiet

https://www.jstatsoft.org/article/view/v061i06/v61i06.pdf

Sommario

Influenzare per governare: chi è effettivamente in grado di farci cambiare idea?

Il passaparola (*Word of mouth* o WOM) "viene spesso riferito al consiglio disinteressato che viene offerto da un consumatore a un altro in merito a un certo prodotto o servizio. Nasce da uno scambio informale di opinioni ed informazioni tra interlocutori che, in linea di principio, non sono mossi da interessi di natura commerciale nel raccomandare un particolare prodotto, trattandosi per lo più di consumatori che, dopo averlo provato ed esserne rimasti soddisfatti, decidono di consigliarlo ai propri conoscenti".

Le aziende si avvalgono di apposite campagne di comunicazione che incoraggiano i consumatori a parlare di un particolare prodotto o servizio ed agevolano lo scambio di informazioni attorno ad esso. Si parla a tal proposito di marketing del passaparola (*Word of Mouth*, Marketing o WOMM), che può essere definito come "uno sforzo compiuto da un'organizzazione per influenzare il modo in cui i consumatori creano e/o distribuiscono le informazioni rilevanti dal punto di vista del marketing ad altri consumatori". Internet e il potenziale comunicativo dei social network hanno accresciuto in maniera esponenziale il potenziale di tale forma di marketing. Si parla infatti oggi di *online Word-of-mouth* (eWom), che presenta numerosi vantaggi:

- la rapida e ampia circolazione delle informazioni attraverso blog
- discussioni fra gente comune su forum e social network
- il fatto che esse rimangano disponibili in eterno e accessibili tramite una semplice ricerca tramite un motore di ricerca una volta indicizzate
- la possibilità per le aziende di monitorarne gli effetti delle azioni di WOM marketing

La sociologia ha studiato ampiamente l'importanza della WOM, in particolare dalle teorie dell'influenza selettiva sviluppatesi fra gli anni quaranta e cinquanta del ventesimo secolo. Esse raccolgono un vasto ed eterogeneo insieme di teorie fondate sul paradigma cognitivo generale della psicologia, ossia che l'influenza di un soggetto su un organismo determina risposte che sono proporzionate alle differenze esistenti fra gli individui. Sono tutte accomunate da una forte attenzione all'analisi del rapporto fra comportamento individuale e comportamenti di gruppo attivati dai mezzi di comunicazione di massa. Nel nostro caso è importante ricordare la teoria delle relazioni sociali e in particolare la teoria del *two-step flow of*

communication. Nel 1955 Paul Lazarsfeld ed Elihu Katz pubblicarono Personal Influence: the Part Played by People in the Flow of Mass Communication. È qui che elaborarono la ormai famosa teoria del *two step flow of communication.* I due studiosi affermavano che non esiste un flusso unitario di informazioni che si muove dai media ai destinatari finali. Il flusso comunicativo segue un percorso composto da due fasi: la prima dai media agli opinion leader, la seconda dagli opinion leader al gruppo sociale di riferimento. L'opinion leader attua una mediazione, egli a sua volta influenza attraverso canali interpersonali gli individui meno esposti ai media. La teoria introduce due concetti molto interessanti: il concetto di gruppo sociale e la nozione di *opinion leader*. Ma cosa è un *opinion leader*? È un "individuo con più o meno ampio seguito di pubblico che ha la capacità di influenzare le opinioni e gli atteggiamenti degli altri e che, dunque, può avere un ruolo determinante nella diffusione di un certo modello di comportamento o di un particolare bene di consumo". È un membro del gruppo sociale più disponibile all'esposizione dei media e più competente nell'uso degli stessi. Oggi il termine viene molto usato nel marketing e in ambito pubblicitario. Indica "quelle persone che, in virtù della loro capacità di esercitare una determinata influenza nei confronti dell'opinione pubblica, costituiscono per le imprese un target prioritario cui indirizzare messaggi pubblicitari, al fine di accelerarne l'accettazione presso un pubblico più vasto". La teoria del *two-step flow of communication* considera quindi i contatti personali come più in grado di influenzare efficacemente il gruppo sociale di riferimento rispetto ai soli media. Detto in altre parole: il passaparola è più potente di qualsiasi messaggio mediale.

Le ricerche di Lazarsfeld, Berelson, Gaudet e in seguito lo studio congiunto di Katz e Lazarsfeld considerano il ruolo dei gruppi sociali e delle relazioni interpersonali nella fruizione mediale fondamentali, tanto da portare ad una influenza selettiva nella fruizione dei mass media: l'*audience* appare dotata di una capacità selettiva che le permette di selezionare i materiali informativi che riceve in maniera netta, molto di più rispetto a quanto ipotizzato dai comportamentisti. "Se la gente tende a esporsi soprattutto alle comunicazioni di massa secondo i propri atteggiamenti e i propri interessi e a evitare altri contenuti e se, per di più, tende a dimenticare questi altri contenuti appena se li trova davanti agli occhi e se, infine, tende a travisarli anche quando li ricorda, allora è chiaro che la comunicazione di massa molto probabilmente non ne cambierà il punto di vista. È di gran lunga molto più probabile anzi che essa rafforzerà le opinioni preesistenti". La teoria del *two step flow* continuò ad influenzare i sociologi per anni. Ecco un altro estratto molto interessante: "nacque una ricca letteratura da cui risultava che le relazioni sociali informali erano importantissimi fattori intervenienti che determinavano il modo in cui le persone selezionavano il contenuto dei media, lo interpretavano e agivano di conseguenza. Così, la teoria delle relazioni sociali andò ad arricchire ulteriormente le conoscenze delle dinamiche e dei fattori alla base della selettività esercitata dai pubblici nella loro risposta alle comunicazioni di massa".

Le ricerche di Lazarsfeld, Berelson, Gaudet e in seguito lo studio congiunto di Katz e Lazarsfeld considerano il ruolo dei gruppi sociali e delle relazioni interpersonali nella fruizione mediale fondamentali, tanto da portare ad una influenza selettiva nella fruizione dei mass media: l'*audience* appare dotata di una capacità selettiva che le permette di selezionare i materiali informativi che riceve in maniera netta, molto di più rispetto a quanto ipotizzato dai comportamentisti. "Se la gente tende a esporsi soprattutto alle comunicazioni di massa secondo

i propri atteggiamenti e i propri interessi e a evitare altri contenuti e se, per di più, tende a dimenticare questi altri contenuti appena se li trova davanti agli occhi e se, infine, tende a travisarli anche quando li ricorda, allora è chiaro che la comunicazione di massa molto probabilmente non ne cambierà il punto di vista. È di gran lunga molto più probabile anzi che essa rafforzerà le opinioni preesistenti". La teoria del two step *flow* continuò ad influenzare i sociologi per anni. Ecco un altro estratto molto interessante: "nacque una ricca letteratura da cui risultava che le relazioni sociali informali erano importantissimi fattori intervenienti che determinavano il modo in cui le persone selezionavano il contenuto dei media, lo interpretavano e agivano di conseguenza. Così, la teoria delle relazioni sociali andò ad arricchire ulteriormente le conoscenze delle dinamiche e dei fattori alla base della selettività esercitata dai pubblici nella loro risposta alle comunicazioni di massa".

Un recente studio di Forrester ha analizzato quanto i consumatori abbiano fiducia negli *influencer* (ha preso in considerazione blogger, opinionisti e celebrità) ed è risultato che solamente il 18% ha fiducia in loro. Uno studio condotto dalla Nielsen ha invece dimostrato che la fiducia dei consumatori nei *brand advocate* ha un tasso del 92%, che è lo stesso livello di fiducia che avrebbero in un amico o in un parente. Un *influencer* è definito tramite dimensione della sua audience (numero di follower su Twitter, numero di persone iscritte al suo blog, follower sul suo canale youtube). Un brand advocate è invece definito tramite la probabilità che raccomandi un prodotto. Passando alle motivazioni che guidano i due: l'*influencer* è interessato solamente a far aumentare la sua audience, il brand advocate è interessato ad aiutare i suoi amici. Gli *influencer* rimarranno fedeli per poco tempo, i brand advocate rimarranno fedeli a lungo. Un *influencer* non è necessariamente guidato da una passione sincera, un brand advocate sì. Un *influencer* solitamente ha bisogno di incentivi economici, un *brand advocate* no.

Molto spesso si tende a confondere *audienc*e con *influence*. Avere un ampio numero di persone che ci segue non implica che noi siamo influenti, significa che abbiamo una audience ampia (ben pochi *influencer* sono in grado di guidare i comportamenti di masse di persone). Un altro problema è che molto spesso gli *influencer* hanno una propria agenda: maggiore è la loro fama maggiore è la difficoltà nell'attirare la loro attenzione per far promuovere il tuo prodotto (ciò spesso implica incentivi economici sostanziosi). Il *brand advocate* ha invece una *marketing force* sostenibile. Desiderano engagement nei confronti del tuo marchio e quindi, al contrario degli *influencer,* non aspettano altro che supportarti, promuoverti, difenderti anche nel lungo periodo.

Questa è una analisi presentata a luglio 2017 da Audiweb .

"La *total digital audience* rappresenta il consumo totale del mezzo, offrendo informazioni sulla reach totale (utenti unici al netto delle sovrapposizioni tra i device rilevati), le pagine viste (per quanto riguarda la fruizione via browser) e il tempo speso online. La *total digital audience* è la dimensione più completa del sistema di misurazione messo a punto da Audiweb e disponibile a partire dai dati di gennaio 2014."

Nel mese di luglio 2017, stando alle statistiche di Audiweb, sono stati circa 32 milioni gli italiani dai 2 anni in su che hanno navigato sia da mobile (smartphone e/o tablet) che da PC, collegandosi complessivamente per 55 ore e 32 minuti. I dati mostrano che il 65,7% degli italiani maggiorenni, ossia 28,8 milioni di abitanti, ha

navigato da mobile (smartphone e/o tablet), dedicando alla navigazione in mobilità circa 49 ore e mezza. Gli italiani che hanno navigato anche o solo da computer hanno invece trascorso solo 14 ore totali. Nel giorno medio la *total digital audience* ha raggiunto 24,4 milioni di italiani, online per una durata di 2 ore e 20 minuti tramite i device rilevati.

La fruizione quotidiana dell'online è quindi ormai principalmente spostata sul mobile (smartphone e/o tablet), con 21,8 milioni di utenti fra i 18 e i 74 anni online da questi device. Una quota significativa, 14,4 milioni, ha addirittura navigato esclusivamente in mobilità. La fruizione di internet da PC raggiunge valori inferiori nel giorno medio, con 9,9 milioni di italiani di età superiore ai 2 anni (che diventano 9,5 milioni quando si considerano quelli di età compresa fra i 18 e i 74 anni) che accedono dai device "fissi" per poco più di un'ora.

Analisi più dettagliate sul tempo speso online attraverso i device rilevati, mostrano che nel mese di luglio 2017 gli utenti maggiorenni hanno dedicato ben l'81% del tempo totale online alla navigazione tramite mobile (smartphone e/o tablet) e solamente il 19% alla navigazione da computer. Device diversi portano a stili di fruizione diversi. Stili di fruizione portano a dover generare tipi di contenuti diversi per cogliere l'attenzione dell'utente. Le donne fanno un uso maggiore di internet, privilegiando i dispositivi mobili. Dedicano all'online da mobile 2 ore e 19 minuti nel giorno medio, mentre gli uomini gli dedicano 1 ora e 54 minuti. I 18-24enni raggiungono invece la soglia delle 2 ore e 43 minuti online da mobile, seguiti dai 25-34enni con 2 ore e 20 minuti.

In base ai dati il 92,2% degli utenti online nel mese di luglio 2017 ha navigato tra le applicazioni e servizi dedicati alla ricerca di contenuti e servizi online. L'88,5% degli utenti ha consultato portali generalisti. L'86,6% ha utilizzato servizi e strumenti online, l'85,5% degli utenti ha utilizzato Social Network e l'81,5% ha guardato contenuti video.

Per quanto riguarda le *news* solamente il 61,8% degli utenti ha navigato per cercarle! Se il 100% degli utenti corrisponde al 58,2% della popolazione questo significa che solamente il 35,96% della popolazione si è esposto alle news su internet! Numericamente sono 22.520.988.

Tra gli altri contenuti di interesse emergono le categorie dedicate all'intrattenimento e al tempo libero, come ad esempio i servizi di messaggistica da mobile (sotto-categoria "Cellular/Paging"), con il 78,6 degli utenti online nel mese, i siti di e-commerce ("Mass merchandiser") con il 72,5% degli utenti, mappe e informazioni di viaggio con il 68,7% e le news ("Current event & global news$) con il 61,8% degli utenti.

Passiamo ora all'analisi dei social network. Presenterò prima varie analisi quantitative e poi una analisi qualitativa. Un approccio quantitativo è sempre utile specialmente quando l'analisi riguarda il target potenziale da raggiungere sul canale scelto in una strategia di comunicazione (la quale può essere a fini commerciali o, come abbiamo già visto, a fini politici). Presenterò diverse analisi provenienti da fonti diverse.

Questo è il numero di utenti attivi secondo una analisi di juliusdesign.net. Rispetto agli "utenti registrati", quelli "attivi" sono molto più utili e interessanti: sono infatti quelle persone che utilizzano in modo assiduo la piattaforma Social Media, sono dunque coloro che assiduamente si espongono ai media. Sono dei potenziali *gatekeeper*.

FACEBOOK	30 Milioni Utenti Attivi	via Vincos
YOUTUBE	24 Milioni Utenti Attivi	via YouTube
TWITTER	6.9 Milioni Utenti Attivi	via Wired Italia
TUMBLR	2.5 Milioni Utenti Attivi	via Yhaoo
SNAPCHAT	2 Milioni Utenti Attivi	via Wired
LINKEDIN	8 Milioni Utenti Attivi	via La Stampa
INSTAGRAM	14 Milioni Utenti Attivi	via Wired Italia
GOOGLE PLUS	7.3 Milioni Utenti Attivi	via GlobalWebIndex
PINTEREST	4.7 Milioni Utenti Attivi	via PinterestItaly

Figura 1 social network in Italia

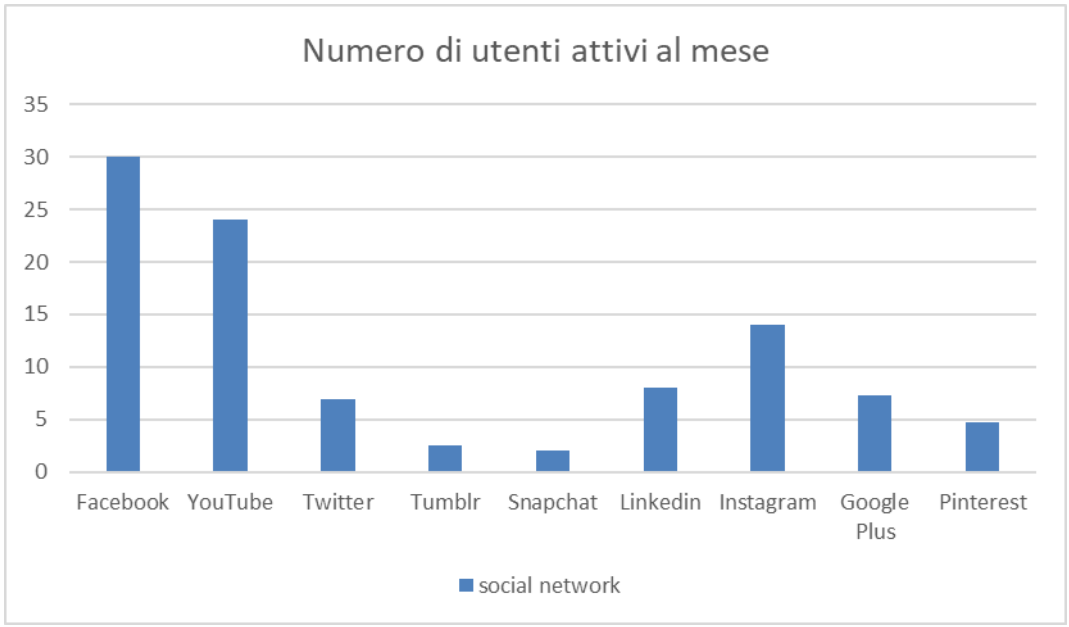


Figura 2numero di utenti attivi al mese in Italia

Analizziamo ora il report Digital in 2017 nato dalla collaborazione tra We Are Social e Hootsuite.

Il tasso di penetrazione per quanto riguarda il numero di utenti internet è più alto rispetto a quello fornito da audiweb, 66% contro 58,2%. Per quanto riguarda il numero di utenti attivi sui *social media* invece le percentuali sono simili. Anche le percentuali riguardanti i dispositivi mobili sono simili. Possiamo quindi dire con sicurezza che gli italiani si connettono sempre di più e sempre di più da dispositivi mobili.

Qui notiamo il calo significativo nel traffico generato da PC e nell'aumento vertiginoso del traffico generato da dispositivi mobili. Come detto in precedenza, dispositivi diversi portano a stili di fruizione diversi che portano a favorire tipi di contenuti e formati diversi.

Il 13° Rapporto Censis-Ucsi sulla comunicazione pubblicato nel 2016 ci fornisce altri dati importantissimi. Secondo il rapporto bel il 73,7% degli italiani sul web, il livello di penetrazione è quindi superiore rispetto a quello stimato da Hootsuite. "Social network e piattaforme online indispensabili nella nostra vita quotidiana. Facebook è il social network più popolare: è usato dal 56,2% degli italiani (il 44,3% nel 2013), raggiunge l'89,4% di utenza tra i giovani under 30 e il 72,8% tra le persone più istruite, diplomate e laureate. L'utenza di YouTube è passata dal 38,7% del 2013 al 46,8% del 2016 (fino al 73,9% tra i giovani). Instagram è salito dal 4,3% di utenti del 2013 al 16,8% del 2016 (e il 39,6% dei giovani). E WhatsApp ha conosciuto un vero e proprio boom: nel 2016 è usato dal 61,3% degli italiani (l'89,4% dei giovani)." Utilissima è l'analisi fatta riguardante il rapporto tra nuovi media e sfiducia nei confronti della classe dirigente: "I media digitali tra élite e popolo. Le ultime tendenze indicano che gli strumenti della disintermediazione digitale si stanno infilando come cunei nel solco di divaricazione scavato tra élite e popolo, prestandosi all'opera di decostruzione delle diverse forme di autorità costituite, fino a sfociare nelle mutevoli forme del populismo che si stanno diffondendo rapidamente in Italia e in Occidente. Si tratta di una sfiducia nelle classi dirigenti al potere e in istituzioni di lunga durata che oggi si salda alla fede nel potenziale di emancipazione delle comunità attribuito ai processi di disintermediazione resi possibili dalla rete. Si sta così radicando un nuovo mito fondativo della cultura web: la convinzione che il lifelogging, i dispositivi di self-tracking e i servizi di social networking potranno fornire risposte ai bisogni della collettività più efficaci, veloci, trasparenti ed economiche di quanto finora sia stato fatto." Importante è anche l'analisi riguardante il rapporto fra anziani e social media: "La frattura generazionale: giovani e anziani sempre più lontani. Le distanze tra i consumi mediatici giovanili e quelli degli anziani continuano ad essere rilevantissime. Tra i giovani under 30 la quota di utenti della rete arriva al 95,9%, mentre è ferma al 31,3% tra gli over 65 anni. L'89,4% dei primi usa telefoni smartphone, ma lo fa solo il 16,2% dei secondi. L'89,3% dei giovani è iscritto a Facebook, contro appena il 16,3% degli anziani. Il 73,9% dei giovani usa YouTube, come fa solo l'11,2% degli ultrasessantacinquenni. Oltre la metà dei giovani (il 54,7%) consulta i siti web di informazione, contro appena un anziano su dieci (il 13,8%). Il 37,3% dei primi ascolta la radio attraverso il telefono cellulare, mentre lo fa solo l'1,2% dei secondi. E se un giovane su tre (il 36,3%) ha già un tablet, solo il 7,7% degli anziani lo usa. Su Twitter poi c'è un quarto dei giovani (il 24%) e un marginale 1,7% degli over 65."

La politica sui social network in Italia

Trovo utile analizzare il numero di *like* e *follower* dei politici su Facebook e Twitter.

Sembrerebbe che utenti con ideologie politiche diverse preferiscono piattaforme diverse, con la sinistra che favorisce decisamente Twitter e la destra che favorisce Facebook. Il movimento 5 stelle ha un elettorato estremamente eterogeneo, per questo in entrambi i casi ha un ampio numero di *follower* e di *like*.

In generale Twitter sembra una piattaforma più orientata a sinistra. Guardiamo i 20 account italiani più seguiti su Twitter nel 2015 e poi nel 2016.

Come politici abbiamo solamente Renzi e Grillo e come giornale solamente la Repubblica. Guardiamo cosa succede nel 2016.

Renzi è ora addirittura al settimo posto mentre Grillo è al quattordicesimo. È impressionante il numero di follower di Renzi su Twitter alla luce di quanto tale *social* sia meno popolare di Facebook in Italia.

Passiamo ora ad una analisi di tipo qualitativo. Nel 2017 Blogmeter, una società italiana che si occupa di *social media intelligence,* utilizzando un campione di 1501 residenti italiani di età compresa fra i 15 e i 64 anni, ha tentato di scoprire "perché gli italiani usano i social media e quali sono i loro impieghi nella vita di tutti i giorni". Che relazione hanno i social media con le relazioni personali, con gli acquisti, con l'informazione? A chi crediamo? A chi dedichiamo più tempo?

Analizzando le modalità con cui vengono utilizzati i vari canali lo studio fa una importante distinzione fra social di cittadinanza e social funzionali. "Della prima categoria fanno parte quei social che usiamo tutti i giorni, anche più volte al giorno, e più volte a settimana, che in un certo senso definiscono la nostra identità online" ha spiegato Alberto Stracuzzi, *customer intelligence director* di BlogMeter. "Facebook è il maggiore rappresentate: ben l'84% degli intervistati ha dichiarato di utilizzarlo più volte al giorno; gli altri sono YouTube, Instagram e Whatsapp".

Per *social* funzionali invece si intendono quei canali che vengono utilizzati per soddisfare un bisogno o un interesse specifico. I principali sono Google Plus, Twitter e Linkedin, che rispettivamente il 40%, il 35% e il 31% dei 1501 intervistati afferma di usare saltuariamente. C'è anche TripAdvisor, consultato per scegliere ristoranti o locali. Questo diverso approccio influenza anche l'atteggiamento e il posizionamento delle aziende sui social. "Stare su un social di cittadinanza è faticoso, con investimenti, per avere una presenza continuativa, con il rischio anche di essere asfissiante. Al contrario su un social funzionale come TripAdvisor, l'importante è saper rispondere alle domande che un utente può porre connettendosi una volta a settimana".

Il 6-7% dice di non poter più fare a meno dei social e il 4% degli intervistati pensa che sia inevitabile iscriversi. Tuttavia stando alla ricerca gli italiani si fanno problemi a cancellarsi da quelli che non apprezzano. Il social più abbandonato in assoluto è Tinder, con ben 3,5 italiani su 10 che hanno dichiarato di essersi iscritti e poi cancellati. Seguono Snapchat, con il 25%, Pinterest e Twitter, con il 10%.

Con l’aumentare dell’età diminuisce il numero di social a cui si è iscritti: nella fascia di età compresa tra i 18 e i 34 anni, la media di social e servizi di messaggistica posseduti è superiore a sette. Dopo i 45 anni, tuttavia, scende a tre canali.

Instagram e YouTube sono i canali su cui gli utenti più giovane, quelli nella fascia di età compresa tra i 15 e i 17 anni, dichiarano di passare più tempo. All’aumentare dell’età subentrano poi Facebook (18-24) e, dagli over 35 anni in su, anche tv e giornali.

Ma cosa spinge ad utilizzare i social? Tra le motivazioni la più gettonata è la curiosità e l’interesse (21%), seguita poi dal desiderio di creazione di relazioni nuove e personali (17%), mentre il 14% afferma di utilizzarli per svago o piacere. Quali sono le ragioni che spingono ad usare un social piuttosto di un altro? Facebook è il più versatile, il più adatto a rispondere a quasi tutte le esigenze (fatta eccezione forse per le ricerche di lavoro). TripAdvisor è utile per leggere recensioni, YouTube per informarsi, mentre per seguire brand e personaggi celebri gli intervistati preferiscono Instragram.

Canali di comunicazione più tradizionali come la televisione e i magazine continuano a mantenere una forte credibilità anche tra gli utenti del web che ritengono poco affidabili Facebook, YouTube e i blog. “Un dato questo che messo anche in relazione al tema delle *fake news*, dimostra come gli utenti se hanno bisogno di credibilità si rivolgono ad altre fonti”. È quindi un errore considerare gli utenti dei social dei “creduloni. Il problema non sorge quando una news circola sui social, ma quando a rilanciarla sono le testate ritenute credibili”.

Quando invece si tratta di fare compere online i canali digitali – tra i siti di ecommerce e quelli di recensioni – tornano ad essere ritenuti attendibili.

Nell’ultima parte della ricerca viene dato anche spazio a *celebrities* e *influencer*. Cantanti, giornalisti e scrittori sono i personaggi di cui ci si fida di più, anche se i più seguiti restano musicisti e personaggi televisivi (33%). Tra i giornalisti popolari sui social abbiamo: Beppe Severgnini, Alberto Angela, Giordano Bruno Guerri e Selvaggia Lucarelli

Dall’analisi, emerge anche che il rapporto con gli *influencer* è però complesso e sfaccettato: se fan-base e credibilità sono aspetti non sempre correlati, età e numero di *influencer* seguiti sì. I giovani sembrano seguire infatti un numero maggiore di personaggi appartenenti a categorie diverse, mentre invecchiando si diventa più selettivi.

Cluster analysis della campagna referendaria su Twitter

Abbiamo utilizzato la *cluster analysis* per analizzare gli utenti che su Twitter hanno preso parte alla campagna elettorale relativa al terzo referendum costituzionale nella storia della Repubblica Italiana, che ha avuto luogo il 4 dicembre 2016. La maggioranza dei votanti respinse il testo di legge costituzionale della cosiddetta riforma Renzi-Boschi, approvato in via definitiva dalla Camera il 12 aprile 2016 e recante modifiche alla parte seconda della Costituzione.

Il campione è composto da 97 *tweet* riconducibili a 90 utenti. L'arco temporale della raccolta dati va dal 29 al 5 dicembre. Nel periodo considerato sono stati scaricati i tweet contenenti le due *keyword* "referendum" e "costituzionale". Dai dati ottenuti sono stati selezionati i *tweet* contenenti *hashtag* caratterizzati in senso "partisan" ("iovotosi, iovotono, bastaunsi, iodicono). Il nostro intento è suddividere questi potenziali *influencer* e *brand advocate* in gruppi il più possibile omogenei al loro interno. Le variabili considerate sono per ciascun utente sono: produttività (media), numero (medio) di follower, numero di amici, se il *tweet* era un *retweet* o meno, il giorno di generazione del *tweet*, il fatto che fosse favorevole o contrario alla riforma.

La nostra analisi è stata divisa in due fasi: nella prima fase abbiamo fatto una analisi di cluster gerarchica, nella seconda fase abbiamo fatto una analisi non gerarchica.

	Conteggio di tweet_id_str	Media di produttività	Media di user_followers_count	Media di user_friends_count	Max di tweet_retweeted_01
Media	1,08	121,07	3149,77	1992,64	0,30
Errore standard	0,03	50,88	1163,91	729,46	0,05
Mediana	1	19	397	511,5	0
Moda	1	1	27	207	0
Deviazione standarc	0,31	482,71	11041,81	6920,23	0,46
Curtosi	19,90	78,55	36,12	76,49	-1,24
Asimmetria	4,31	8,62	5,93	8,47	0,89
Intervallo	2	4502	73526	64406	1
Minimo	1	1	0	3	0
Massimo	3	4503	73526	64409	1
Conteggio	90	90	90	90	90

Le statistiche descrittive delle variabili (in tabella) sono parzialmente in linea con quanto scoperto da studi precedenti. Secondo l'articolo di Jon Bruner pubblicato per O'Reily Radar il 18 dicembre 2013, su un campione casuale di 400,000 utenti l'account mediano ha un singolo follower (prendendo in considerazione gli account che si sono loggati almeno una volta al mese). Se invece prendiamo in considerazione gli account che hanno postato almeno una volta in un mese l'account mediano ha 61 follower. Un account con 1000 follower si trova già nel 96esimo percentile. Il 76% segue più persone di quante poi seguano loro. Come possiamo vedere anche nel nostro caso la maggioranza degli utenti segue più persone di quante poi seguano indietro, tuttavia l'account mediano ha molti più follower e amici di quello dello studio di Jon Bruner (il nostro campione è però molto più piccolo).

Sulla matrice delle distanze tra le unità (i 90 utenti), calcolata a partire dai valori standardizzati delle variabili, è stato applicato il metodo di Ward che ha generato il seguente dendrogramma.

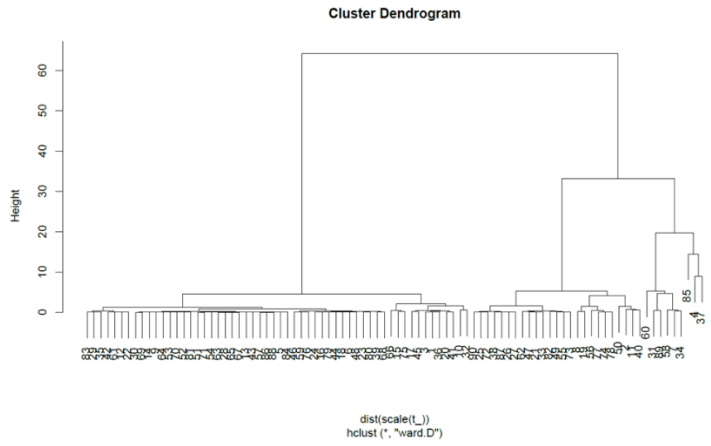


Figura 3 dendrogramma, metodo di Ward

Per determinare il numero di gruppi (operazione necessaria per la seconda fase della nostra analisi) useremo NbClust, un pacchetto di R per determinare il numero rilevante di cluster in un data set attraverso l'uso di ben 30 indici di *cluster validity*. Utilizzeremo l'indice Silhouette, la cui formula può essere riscritta come segue:

$$s_i = \frac{b_i - a_i}{max(a_i, b_i)}$$

a_i è la distanza media fra l'unità *i* e le altre unità all'interno dello stesso cluster dell'unità *i*. Il valore b_i è invece la distanza media tra l'unità *i* e le unità del più vicino degli altri cluster.

In base all'indice Silhouette il numero di cluster ottimale è 7, poiché con 7 cluster abbiamo un valore pari a 0,8132. Con 8 cluster avremmo avuto un valore pari a 0,7414 e con 6 un valore pari a 0,7993. Possiamo chiaramente vedere che anche 6 è un numero accettabile, tuttavia 7 lo migliora. Con 8 invece abbiamo un calo notevole, quindi non va preso in considerazione. Raccogliendo i dati di Nbclust ecco dunque la migliore partizione:

1	2	3	4	5	6	7
56	25	1	5	1	1	1

Applichiamo in ultima analisi l'algoritmo delle k-medie, utilizzando come dati di input la matrice dei dati standardizzata e un numero iniziale di centri pari a 7.

Riportiamo dunque i centri ottenuti:

	numero_tweet	produttivita	media_user_followers_count	media_user_friends_count	max_tweet_retweeted01
1	-0.2523361	0.169441683	-0.039840838	0.17168176	-0.65100655
2	3.5327060	0.002278821	0.008012976	0.15431836	0.07233406
3	-0.2523361	-0.203160341	6.373611739	0.07635049	-0.65100655
4	-0.2523361	-0.207515046	-0.243737408	-0.21331193	-0.65100655
5	-0.2523361	0.008894307	-0.010608061	-0.01935731	1.51901528
6	-0.2523361	-0.128580746	6.332857562	9.01940580	-0.65100655
7	-0.2523361	9.077855929	-0.267326776	-0.28722124	-0.65100655

```
Within cluster sum of squares by cluster:
[1]  1.7161930 16.9561815  0.0000000  0.7300259 10.1183414  0.0000000  0.0000000
 (between_SS / total_SS =  93.4 %)
```

Riportiamo la devianza interna di ciascun gruppo e l'indice R^2 pari a 93,4%

Analizziamo però da vicino tale risultato.

Utilizzeremo la cosiddetta *average silhouette widht*. Essa può assumere, come già detto, un valore compreso fra -1 e +1. Un valore negativo non è desiderabile, poiché ciò corrisponde al caso in cui $\boldsymbol{a_i}$, la distanza media nei confronti dei punti nel cluster, è superiore a $\boldsymbol{b_i}$, la minima distanza media nei confronti dei punti in un altro cluster. Vogliamo che il coefficiente sia positivo ($\boldsymbol{a_i < b_i}$) e per $\boldsymbol{a_i}$ vogliamo che esso sia il più possibile vicino a 0 poiché il coefficiente assume il suo valore massimo,1, quando $\boldsymbol{a_i = 0}$. L'*average silhouette coefficient* si calcola semplicemente facendo la media dei *silhouette coefficient* di tutte le unità appartenenti al cluster. Una misura della bontà di un clustering può essere calcolata calcolando l'*average silhouette coefficient* di tutti i punti. Nel grafico vogliamo che la silhouette sia il più larga possibile. Questo ci permette di distinguere un "taglio pulito" rispetto a cluster "deboli" all'interno dello stesso grafico: cluster con una averege silhouette widht più grande sono più pronunciati. Questo è chiarissimo: i cluster 2 (0,41), 4 (0,79), e 5 (0,68) sono enormemente più pronunciati dei cluster 1 (0,003), 3 (0,00), 6 (0,00) e 7 (0,00).

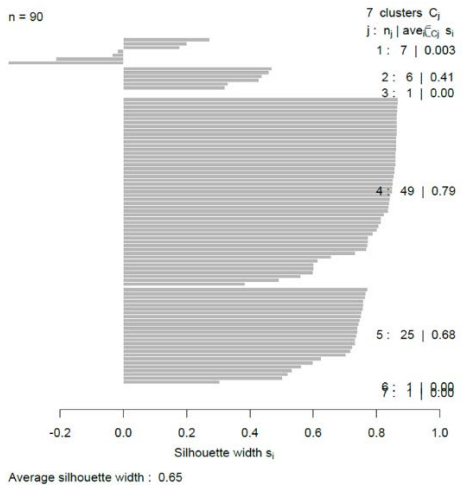


Figura 4 grafico silhouette

Analizziamo più da vicino dunque i cluster analizzando i centroidi e la *silhouette width*.

- CLUSTER 1: Quello che mi è balzato subito all'occhio è che il valore relativo al numero di tweet totali, una volta standardizzato, è enormemente inferiore rispetto alla produttività. Stiamo parlando dunque di persone che si sono mobilitate appositamente per la campagna, con un numero di amici superiore al numero di follower, bassissimo numero di retweet. *Silhouette width* molto bassa.

- CLUSTER 2: Come abbiamo già visto questa cluster ha una mediocre *silhouette widht*. Passando all'analisi dei centroidi anche questo gruppo è interessante: numero di tweet enormemente più alto della produttività. Siamo dunque di fronte a gente che non si è mobilitata appositamente per la campagna. Numero di amici superiore al numero di follower, alto numero di retweet.
- CLUSTER 3: Bassissimo numero di tweet, bassa produttività, altissimo numero di follower, bassissimo numero di amici. Bassissimo numero di retweet. Siamo infatti di fronte ad un politico. Non è la persona comune che poi influenzerà i suoi amici e conoscenti. *Silhouette widht* nulla.

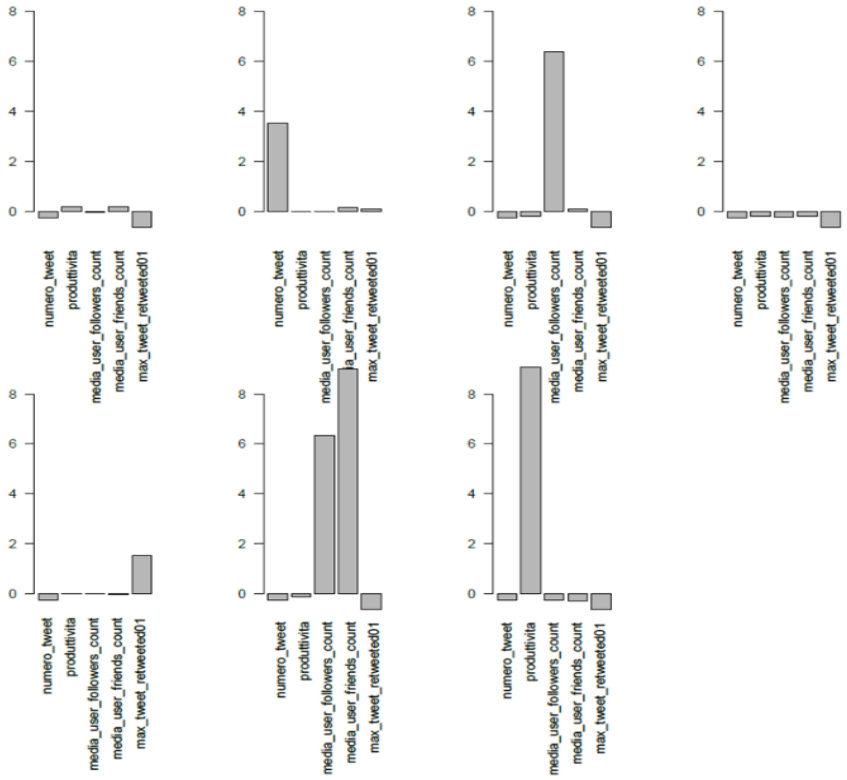


Figura 5 grafico centroidi a barre

- CLUSTER 4: Qui tutti gli indicatori sono bassi. *Silhouette widht* molto alta.
- CLUSTER 5: Cluster decisamente interessante. Numero di tweet inferiore rispetto alla produttività durante la campagna. Follower e amici quasi uguali (e il numero è basso). Altissimo numero di retweet (il più alto fra tutti). Potremmo essere di fronte al gruppo con più influenza nel mondo reale. *Silhouette widht* alta.
- CLUSTER 6: Bassissimo numero di tweet, bassa produttività. Alto numero di follower, numero di amici decisamente superiore. Bassissimo numero di retweet. Silhouette widht nulla.
- CLUSTER 7: Basso numero di tweet, enorme produttività durante la campagna, basso numero di follower, basso numero di amici, basso numero di retweet. Guardando il profilo si legge che l'utente Schiforma ha generato il suo profilo appositamente osteggiare la riforma elettorale. Dubito abbia avuto una influenza pesante nella vita reale. *Silhouette widht* nulla.

Tramite l’analisi degli hashtag siamo riusciti a individuare cosa sostenevano i vari utenti.

CLUSTER 1: quasi esclusivamente favorevole al SI.

CLUSTER 2: esclusivamente favorevole al NO.

CLUSTER 3: esclusivamente favorevole al SI.

CLUSTER 4: metà sì e metà NO.

CLUSTER 5: larga maggioranza SI.

CLUSTER 6: esclusivamente SI.

CLUSTER 7: esclusivamente NO.

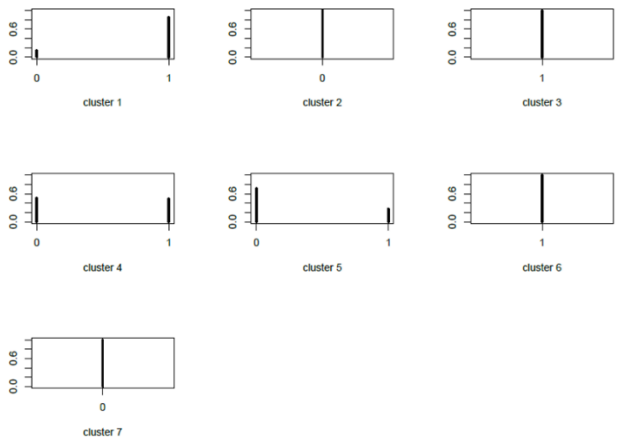


Figura 6 SI o NO al referendum

Abbiamo costruito un *alluvial plot* che mette in relazione gli utenti e i loro retweet.

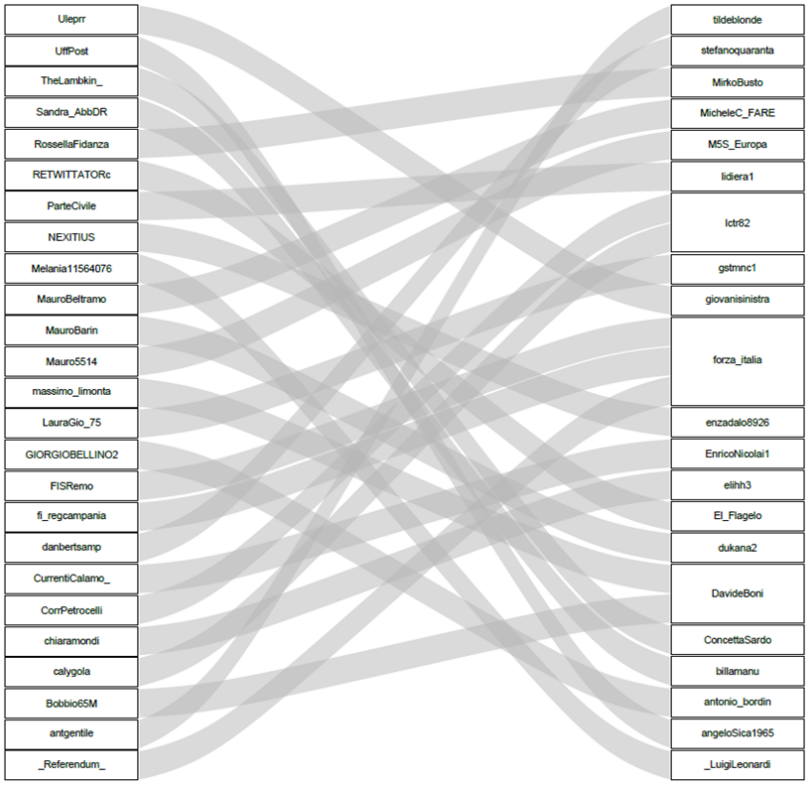


Figura 7 alluvial plot dei retweet

Costruiamo ora due diverse categorie di word cloud.

Word cloud degli hashtag:

- Word cloud generale
- Word cloud del NO
- Word cloud del SI

Possiamo subito vedere che coloro che hanno sostenuto il no fanno un uso molto più ampio degli hashtag.

Word cloud delle parole dei tweet:

- Parole relative al SI
- Parole relative al NO

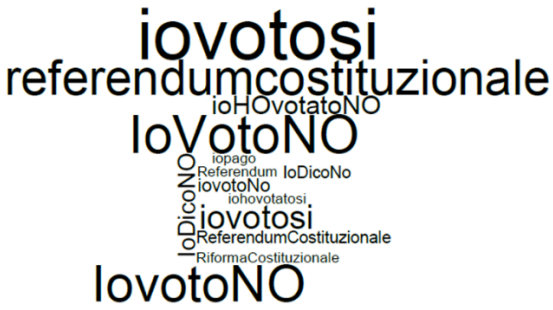


Figura 8 word cloud di tutti gli hashtag

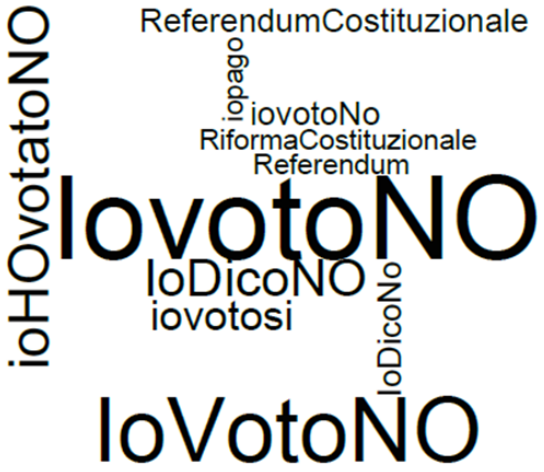


Figura 9 word cloud degli hashtag del NO

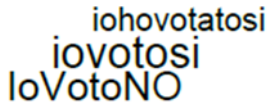


Figura 10 word cloud degli hashtag del SI

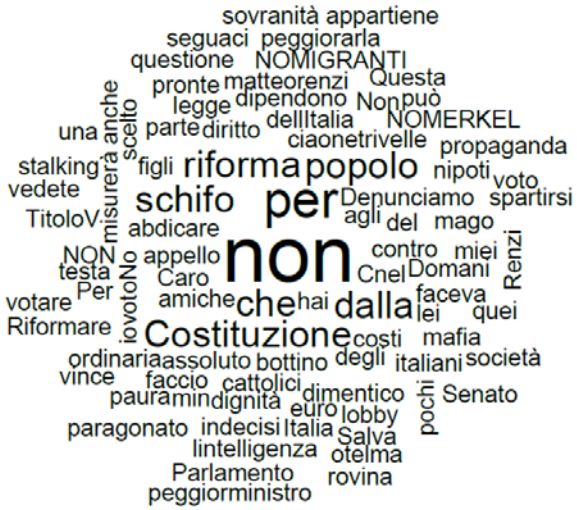


Figura 3 tweet NO

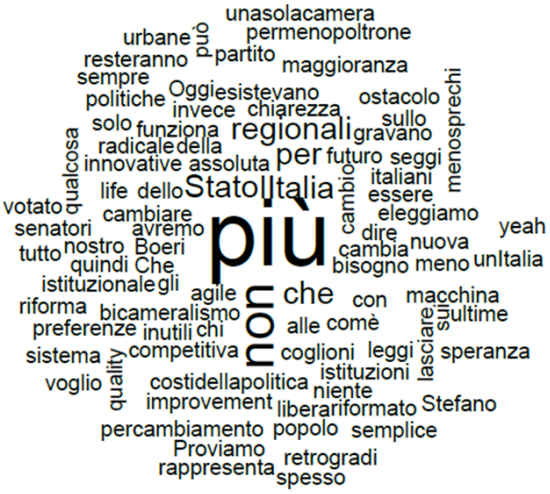


Figura 12 tweet SI

Possiamo vedere come fra coloro che hanno votato NO abbondino riferimenti a messaggi anti UE, mentre tra coloro che hanno votato SI le parole cambiare e cambiamento.

In base a quanto detto fino ad ora posso finalmente identificare le caratteristiche del brand advocate politico a cui bisogna mirare su Twitter, la persona "comune" il cui messaggio è in grado di mobilitare amici e parenti:

- Il numero dei follower deve essere pari al numero degli amici, oppure il numero degli amici deve essere pari a circa i 2/3 del numero dei follower.
- Deve avere pochi amici e follower (dobbiamo mirare il più possibile alla persona comune)
- Usa il proprio nome

- Ha scritto pochi messaggi dal momento dell'attivazione del profilo.
- Non scrive troppi messaggi. È facile diventare noiosi su internet.
- Si mobilita solamente durante la campagna (quindi dobbiamo vedere che la media di messaggi scritti durante l'arco temporale della campagna sia superiore rispetto alla media di messaggi scritti sin dall'attivazione del profilo). Questo perché in questa maniera si evita che l'utente risulti "pedante" (se la cerchia di amici e parenti lo reputa "pedante" non leggerà i suoi messaggi con attenzione). L'attenzione è un bene prezioso su internet che va centellinato.
- Retwetta. Può retwettare messaggi di politici, giornali ai suoi amici (attiva così il two step flow of communication). Oppure può saltare questo passo e retweettare direttamente messaggi di amici. Se i due hanno ad esempio una cerchia di amici in comune è più probabile che il messaggio dell'amico venga ascoltato dagli altri amici.
- Usa molti hashtag. Gli hashtag aiutano a dare visibilità ai messaggi (specialmente se si usa un trending hashtag) e aiutano subito a "etichettare" il messaggio (sappiamo alla prima occhiata il punto su cui si concentrerà).
- Usa parole triviali, senza però abusarne. Come già detto su internet c'è una soglia di attenzione bassa. Messaggi brevi e divertenti. Se accompagnati da foto divertenti (ad esempio meme) aiutano ancora meglio a catturare l'attenzione.

Il gruppo che più si avvicina a questo è il gruppo numero 5 (che sostiene in larga maggioranza il no, la scelta che ha vinto).

Bibliografia

CAIAZZO, D., COLAIANNI, A., FEBBRAIO, A., MASI, D., (2009). Buzz marketing nei social media. Come scatenare il passaparola on-line, Fausto Lupetti Editore

SORICE, M., (2009) Sociologia dei mass media. Carocci editore

PACCAGNELLA, L., (2004) Sociologia della Comunicazione. il Mulino

KLAPPER, J. T., (1960). Effects of Mass Communication

DE FLEUR, M. L., BALE-ROKEACH, S., (1989). Theories of Mass Communication

CAIAZZO, D., COLAIANNI, A., FEBBRAIO, A., LISIERO, U., (2009). Buzz marketing nei social media

LASN, K., Culture Jam: The Uncooling of America, (1999). Eagle Brook

HARLOW, S. HARP, D., (2012). Collective action on the Web: A cross-cultural study of social networking sites and online and offline activism in the United States and Latin America. Information, Communication & Society

KARPF, D., (2010). Online political mobilization from the advocacy group's perspective: Looking beyond clicktivism. Policy & Internet

REBER, B. H., KIM, J. K. (2006). How activist groups use websites in media relations: evaluating online press rooms. Journal of Public Relations Research.

Sitografia

http://www.glossariomarketing.it/significato/word-of-mouth/

http://www.unipd.it/ilbo/content/il-mondo-%E2%80%9Cpiccolissimo%E2%80%9D-dei-social-network

http://www.glossariomarketing.it/significato/opinion-leader/

http://www.glossariomarketing.it/significato/public-relations/

http://www.glossariomarketing.it/significato/brand-advocate/

http://www.zuberance.com/downloads/brandAdvocateInsights.pdf

http://www.nielsen.com/us/en/newswire/2012/consumer-trust-in-online-social-and-mobile-advertising-grows.html

https://www.weforum.org/agenda/2016/08/hillary-clinton-or-donald-trump-winning-on-twitter/

http://www.stirista.com/wpcontent/uploads/2016/06/WhosFollowingTrumpAndClinton-1.pdf

https://www.alexa.com/siteinfo/reddit.com

https://www.gruppodigitouch.it/servizi/amplification/social-media-content/

https://www.youtube.com/watch?v=MKH6PAoUuD0

https://www.nytimes.com/2016/11/20/opinion/sunday/reddit-and-the-god-emperor-of-the-internet.html?mcubz=3

https://www.dailydot.com/layer8/donald-trump-inauguration-donations-crowdsourced-journalism-reddit-twitter/

https://www.dariovignali.net/marketing-politico-ed-elettorale/

http://www.repubblica.it/speciali/esteri/presidenziali-usa2016/2016/11/12/news/trump_internet_meme_virali_social_4chan-151826943/

http://thehill.com/homenews/campaign/334897-poll-majority-says-mainstream-media-publishes-fake-news

https://www.tvdigitaldivide.it/2017/09/15/audiweb-32-mln-gli-italiani-online-a-luglio-2017/

http://www.audiweb.it/dati_it/total-digital-audience_it/

http://www.juliusdesign.net/28700/lo-stato-degli-utenti-attivi-e-registrati-sui-social-media-in-italia-e-mondo-2015/

http://www.censis.it/7?shadow_comunicato_stampa=121073

https://www.wired.it/internet/social-network/2016/03/08/italiani-social-media/

http://www.ilsole24ore.com/art/notizie/2017-09-28/su-facebook-testa-testa-grillo-e-salvini-doppiato-renzi-che-si-rifa-twitter-091110.shtml

http://www.ilpost.it/2015/10/09/account-italiani-piu-seguiti-su-twitter/

http://www.ilpost.it/2016/10/10/account-italiani-piu-seguiti-twitter-2/

https://www.youtube.com/watch?v=CTvzvyy3EIk

https://www.oreilly.com/ideas/tweets-loud-and-quiet

https://www.jstatsoft.org/article/view/v061i06/v61i06.pdf