



**MaCSIS**

Università degli Studi di Milano-Bicocca

Centro Interuniversitario MaCSIS

**MaCSIS Working Paper Series**

**LA COMUNICAZIONE DELLA SCIENZA MEDIATA DAI  
SOCIAL MEDIA**

Elisa Bellistri

**Working Paper n.1/2019**

# **UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA**

Dipartimento di Sociologia e Ricerca Sociale

Master Universitario di I Livello

Master in Comunicazione della Scienza  
e dell'Innovazione Sostenibile



## **La comunicazione della scienza mediata dai social media**

Relatore: Dott. Eugenio SANTORO

Tesi di Master di: Elisa BELLISTRI

Matr. N. 849686

Anno Accademico 2018/2019

## Indice

- Sommario ..... pag. 3
- Capitolo 1 – Introduzione ..... pag. 5
- Capitolo 2 – Materiali e metodi ..... pag. 7
- Capitolo 3 – Risultati ..... pag. 12
- Capitolo 4 – Discussione e Conclusioni ..... pag. 23
- Bibliografia ..... pag. 27
- Sitografia ..... pag. 29

## Sommario

La rivoluzione del Web 2.0 viene ricordata per l'enorme diffusione dei social network, grazie ai quali gli utenti del web non sono stati più solo spettatori, ma sono potuti diventare protagonisti della rete, creando i propri contenuti e condividendoli con il proprio pubblico.

Questa rivoluzione non ha coinvolto solo le relazioni personali, ma anche il mondo del lavoro ha potuto sfruttare le potenzialità della rete.

In particolare, nell'ambito della ricerca, in cui da sempre la creazione di contatti e di reti di competenze ha rappresentato un valore aggiunto, i social network hanno dimostrato di giocare un ruolo importante. Accanto ai social network di uso comune (Facebook, Twitter, Youtube), sono sorte altre piattaforme espressamente dedicate al mondo della ricerca (ResearchGate, Academia, Mendeley), ma che conservano gli stessi meccanismi di interazione, come la possibilità di commentare e condividere un post, di indicare il proprio apprezzamento (*Like*), di seguire un contatto (*Follower*) o di stringere una mutua relazione che permette l'accesso a ulteriori contenuti (*Friendship*).

In questa Tesi abbiamo analizzato un caso particolare di ricerca, la ricerca italiana in ambito sanitario, che si svolge negli Istituti di Ricerca e Cura a Carattere Scientifico (IRCCS). L'obiettivo è analizzare il comportamento degli IRCCS sui social, per avere una panoramica sulla diffusione dei social in ambito di ricerca italiana e sul grado di interazione raggiunto, sia con altri istituti di ricerca, sia con il proprio personale di ricerca e anche con il cittadino comune.

Dopo una breve panoramica sull'uso delle diverse tipologie di social, ci siamo concentrati sulle interazioni e le conversazioni con Twitter, un social di uso comune ma che ha precise caratteristiche di target di utenti e una larga diffusione in particolari settori, tra cui quello della ricerca scientifica.

La struttura della Tesi si compone quindi di:

- un Capitolo di **Introduzione**, dove viene riepilogata la letteratura scientifica in merito all'uso dei social network in ambito di ricerca e vengono descritte le ipotesi di lavoro

con le quali si sono analizzate le conversazioni e le relazioni degli account Twitter dei diversi IRCCS;

- un Capitolo di **Materiali e Metodi**, che descrive in dettaglio la metodologia utilizzata per l'estrazione e l'analisi dei dati da Twitter, sia in termini di contenuti che di relazioni tra gli utenti;
- un Capitolo di **Risultati**, dove vengono rappresentati attraverso l'uso di grafici e di altri strumenti visuali i risultati sia quantitativi che qualitativi del tipo di utilizzo che viene fatto di Twitter da parte degli IRCCS;
- un Capitolo di **Discussione e Conclusioni**, dove si commentano i risultati ottenuti, descrivendo sia le potenzialità sia le limitazioni del tipo di analisi utilizzato, ma anche discutendo come lo strumento dei social network possa essere non solo un canale di comunicazione e divulgazione della ricerca, ma anche un modo per sviluppare e migliorare la ricerca stessa.

# Capitolo 1

## Introduzione

La nascita dei **social network** e dei **social media**, nei primi anni 2000, sembrava in un primo tempo che relegasse queste piattaforme a una sola funzione *social*, cioè di relazioni tra le persone, di svago [boyd, danah m. and Ellison, N. B. (2007)]. Ben presto tuttavia è risultato chiaro come la componente *network*, la rete di relazioni possibili via Internet, potesse portare enormi benefici in ambito lavorativo e, nello specifico, nell'ambito della ricerca scientifica. Nel contesto scientifico, i social media permettono una comunicazione bidirezionale, superando il vecchio modello di comunicazione monodirezionale, dove il messaggio è trasferito da un singolo punto a una platea selezionata [Collins, K., Shiffman, D. and Rock, J. (2016)]. La diffusione dei social media in ambito scientifico ha quindi cominciato a crescere in modo esponenziale, sia sui canali social generici che su quelli specializzati. Parlando di social network espressamente dedicati al mondo scientifico, risulta intuitivo il vantaggio che possono dare: la possibilità di stringere contatti con persone dislocate ovunque nel mondo aumenta lo scambio di idee e favorisce le collaborazioni.

Ma l'interazione tra gruppi o istituti di ricerca non necessariamente deve avvenire sui canali dedicati al solo mondo scientifico. L'interazione tra ricercatori, o tra ricercatori e pubblico generico, che avviene su piattaforme social come Twitter, Facebook o LinkedIn, porta con sé ulteriori vantaggi: l'uso dei social media aiuta a comunicare la scienza, adeguando il registro ad un pubblico di volta in volta specifico [Van Eperen, L. and Marincola, F. M. (2011)]; da un altro punto di vista, la presenza dei ricercatori sui social colma un vuoto che viene troppo spesso riempito da divulgatori di fake news [Vosoughi, S., Roy, D. and Aral, S. (2018)].

Facendo riferimento al caso Burioni, ormai divenuto un vero e proprio metodo di comunicazione sui social, efficace o meno (ma non è questo il punto), sicuramente bisogna riconoscere il suo tentativo di fare da controaltare alle voci che diffondono timori ingiustificati sui vaccini o, in generale, screditano la medicina. Le percentuali dei *no-vax* in Italia sono esigue (la copertura vaccinale in età pediatrica copre il 94.5% dei bambini, fonte: Ministero della Salute <https://tinyurl.com/yxy75u55>), eppure la loro presenza sui social sembra notevolmente maggioritaria. Nel caso specifico della ricerca in ambito sanitario,

promossa in Italia dagli Istituti di Ricerca e Cura a Carattere Scientifico (IRCCS), la divulgazione a un pubblico generico e il contrasto alle fake news diventano obiettivi ancora più rilevanti. La divulgazione della ricerca in ambito medico-scientifico può diventare anche uno strumento di sensibilizzazione e di prevenzione su specifici temi sanitari o promuovere stili di vita salutari [Santoro, E. et al. (2015)].

La rivoluzione del web 2.0 non è stata solo la diffusione dei social media e social network, ma di un'un'intera concezione della proprietà intellettuale e dello scambio di idee. Con la diffusione di Internet, si sono diffuse le riviste scientifiche open source, mutuando uno schema che diversi anni era già utilizzato dagli sviluppatori di software. La relazione tra i social media, open science e big data scientifici ormai è strettissima [Voytek, B. (2017)].

I social media servono quindi a promuovere la propria ricerca, ma questo risultato si ottiene solo facendo un uso attivo dei social, che riesca a catturare in modo efficace il pubblico, generico e specialista. L'obiettivo di questo progetto è valutare quanto in Italia, nello specifico caso degli Istituti di Ricerca e Cura a Carattere Scientifico, si utilizzano i social, muovendosi quindi nella direzione suggerita dall'evoluzione della società e della comunità scientifica.

## Capitolo 2

### Materiali e Metodi

#### 2.1 Social Network e Big Data

Una piattaforma come Twitter non racchiude solo dei contenuti (i post), ma anche un enorme insieme di dati impliciti che aumentano il contenuto informativo della base di dati. La frequenza con cui vengono pubblicati i post, il numero di follower, il tipo di post pubblicati o il numero di like, commenti e retweet ricevuti contribuiscono ad arricchire il profilo utente. Questo insieme di **dati strutturati e non strutturati**, nell'ordine delle centinaia di migliaia di singoli elementi, prende il nome di Big Data. La crescita esponenziale dei big data ha richiesto prima l'uso di tecniche di classificazione basate sulla semantica, il web 3.0, poi l'uso di tecniche di classificazione basate su algoritmi di apprendimento automatico (*machine learning*), il web 4.0. Internet non è più solo uno strumento interattivo, ma è diventato una fonte continua di dati e informazioni.

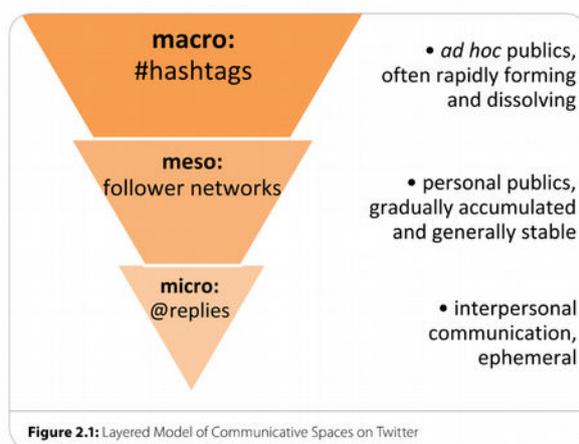
Anche l'analisi dell'uso di Twitter da parte degli IRCCS costituisce un sistema complesso da cui estrarre nuova informazione. Se infatti il numero di account ufficiali Twitter degli IRCCS non supera le poche decine, il numero di follower, di post pubblicati e di interazioni su ogni post accrescono in modo esponenziale la base di dati.

Sono stati quindi scelti dei parametri per restringere il campo di estrazione dei dati, senza alterare la significatività del risultato.

In particolare si è scelto di:

- considerare per ogni IRCCS l'**account ufficiale**;
- limitare l'estrazione delle conversazioni agli **ultimi 50 post** pubblicati. Analisi di tipo diverso sono state eseguite in tempi diversi, non necessariamente sulla stessa base di post. Le due principali estrazioni di post sono state realizzate al 29 settembre 2019 e al 30 novembre 2019. Trattandosi di una base di dati a campione e di analisi diverse su due insiemi di post, la differenza del campione non pregiudica la validità del risultato;

- valutare l'*engagement* dei post attraverso i parametri di **numero di like** e **numero di retweet**.
- costruire un network delle relazioni tra gli IRCCS attraverso la relazione di **follower** (Twitter identifica anche il livello reciproco, quello di friend - gli account seguiti dall'account preso in esame. - Questo tipo di interazione non è stata considerata nel progetto);
- costruire un grafo delle relazioni attive tra gli IRCCS e i loro follower attraverso l'azione di **retweet**.



Le diverse azioni possibili sui social sono state descritte da Bruns, Axel & Moe, Hallvard (2013) in diversi livelli di comunicazione, secondo uno schema rappresentato in figura 2.1.

*Figura 2.1. I tre livelli di comunicazione su Twitter.*

## 2.2 Le API di Twitter

L'estrazione dei dati per ogni account IRCCS è stata realizzata in modo automatico attraverso le **API** (*Application Programming Interface*) messe a disposizione da Twitter. Una API è una procedura software che permette di accedere alle funzionalità di altre applicazioni, disaggregando i dati e quindi potendo accedere al dato grezzo per ulteriori elaborazioni.

Twitter controlla l'accesso alle API mediante alcune limitazioni.

Anzitutto, per poter accedere alle funzionalità avanzate è necessario registrarsi come **sviluppatori** e creare una App dedicata all'uso, collegandosi all'indirizzo <https://developer.twitter.com>. La registrazione come sviluppatori e la creazione dell'app consentono di ricevere un codice univoco di identificazione (**App ID**) e un **token**, ossia una chiave segreta di autenticazione con cui stabilire la connessione tra il proprio sistema e l'applicativo di Twitter. Nel caso di questo progetto, è stata creata una app chiamata IRCCS (Fig. 2.2).



Figura 2.2. La App creata per accedere alle API.

Una volta ottenuta la chiave di autenticazione, bisogna scegliere l'applicativo con cui realizzare la connessione e l'estrazione dei dati.

Per gli specifici scopi di questo progetto, è stato scelto l'ambiente di analisi statistica **R Project**, un software gratuito, rilasciato sotto licenza GNU. Per R è stato sviluppato il pacchetto **Rtweet 0.7.0**, una libreria di funzioni per l'accesso ai dati Twitter. Il pacchetto `rtweet` è disponibile con licenza GNU e liberamente scaricabile all'indirizzo <https://rtweet.info/>.

Riporto in sintesi le principali funzioni utilizzate e la loro descrizione.

- `lookup_users()` Restituisce gli attributi di un account, quali la data di registrazione, il numero di post, il numero di account seguiti e di account da cui si è seguiti.
- `get_timeline()` Restituisce i post dell'account specificato e gli attributi corrispondenti.
- `get_follower()` Restituisce gli identificativi (IDs), i nomi e gli attributi degli account che seguono l'utenza specificata.

Molti dati Twitter sono ottenibili in modo diretto, ma per alcune tipologie (per esempio gli account che hanno messo like a un determinato post) si può risalire al dato solo con un algoritmo ricorsivo indiretto. Twitter impone inoltre delle limitazioni quantitative e temporali di accesso al database. Ogni funzione può restituire un numero massimo di risultati (per esempio la funzione `get_follower()` restituisce al massimo 75000 IDs per volta) e può essere eseguita un numero limitato di volte al minuto.

Queste limitazioni rendono indispensabile l'uso di una procedura software in grado di effettuare autonomamente le interrogazioni al database e archiviare su file in locale i dati di volta in volta raccolti.

### 2.3 *Analisi statistica, analisi delle conversazioni e analisi dei grafi*

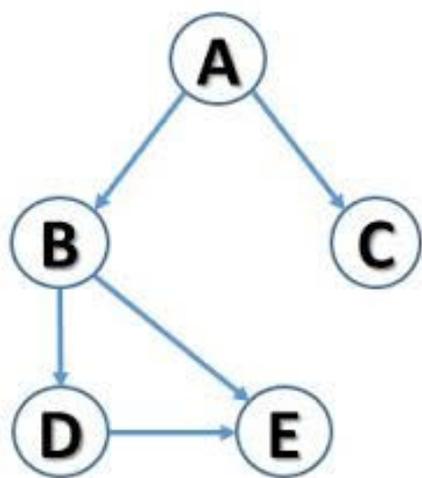
I dati raccolti in maniera disaggregata sono stati analizzati con a) **metodi statistici quantitativi**, b) con metodologie di **analisi semantica** del testo e c) attraverso gli strumenti di analisi della **teoria dei grafi**.

Per le analisi di tipo quantitativo (frequenza delle occorrenze, media, varianza) sono state utilizzate le funzioni native del pacchetto R Project.

Per le analisi del testo si è fatto uso del pacchetto **Text Mining TM 0.7-6**, sviluppato da Ingo Feinerer e integrato in R. Le funzioni di questo pacchetto permettono l'analisi sintattica dei post, l'eliminazione delle parole non rilevanti (per esempio preposizioni o articoli), la suddivisione del testo in unità elementari di significato. Con il testo dei post così riorganizzato, è possibile fare analisi di frequenza delle parole utilizzate e analisi semantica, per esempio attribuendo un valore alla conversazione (positiva, negativa, neutra) o inserendola in una classificazione.

Infine, per utilizzare gli strumenti di analisi della teoria dei grafi, è stato utilizzato il software Cytoscape ([www.cytoscape.org](http://www.cytoscape.org)). Caricando su cytoscape i dati di relazione tra gli account (account IRCCS e account dei follower), il software visualizza in forma di grafo (strumento matematico dove gli oggetti sono nodi e le relazioni tra gli oggetti sono archi, Fig. 2.3) la base di dati.

Il software, oltre alla visualizzazione, permette anche analisi avanzate. Sono stati utilizzati gli strumenti di network analysis offerti dal software, attraverso i quali calcolare gli indici che caratterizzano la rete. In particolare è stato valutato l'indice di **inbound/outbound degree**, che misura in numero di archi in ingresso/in uscita da un nodo e viene usato per caratterizzare la topologia della rete [Leicht, E. A. and Newman, M. E. J. (2008)].



*Figura 2.3. Un esempio di grafo.*

## Capitolo 3

### Risultati

#### 3.1 La diffusione dei social network negli IRCCS

Gli IRCCS in Italia sono 51 (fonte: Ministero della Sanità <https://tinyurl.com/ru52859>, dati aggiornati al 4 novembre 2019), sia a carattere pubblico che privato, distribuiti nelle diverse regioni italiane. Il primo obiettivo del progetto è stato quantificare la presenza sui social media degli IRCCS, considerando i principali social media di uso comune (Facebook, Twitter, Instagram, LinkedIn e Youtube) e raggruppando sotto la colonna Altri l'eventuale presenza di alcuni centri in social media diversi, il cui campione era troppo poco numeroso per essere esplicitato (a titolo di esempio: Vimeo, Google Photo, Flickr).

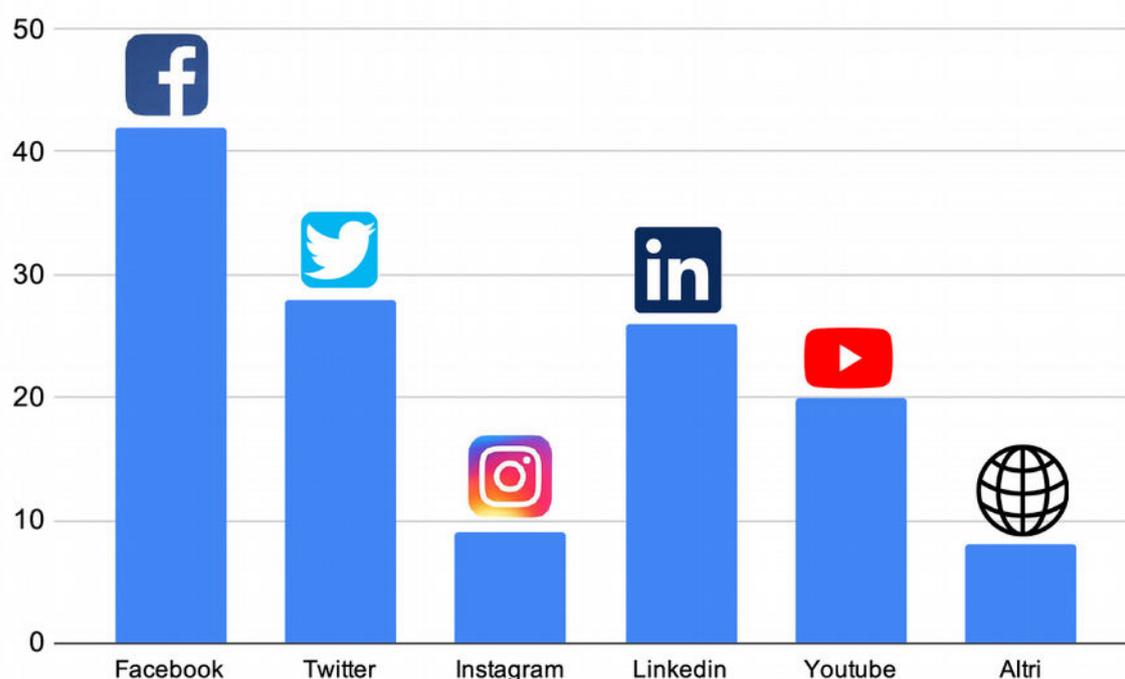


Figura 3.1. L'uso dei diversi canali social da parte degli IRCCS

Il social media più utilizzato risulta Facebook, con una presenza di oltre 40 account ufficiali di IRCCS, seguito da Twitter, con una presenza di 28 account IRCCS. A livelli

comparabili è anche il numero di IRCCS che ha un account LinkedIn, seguito poi da Youtube e infine da Instagram (Fig. 3.1).

Se distinguiamo l'uso dei social media tra gli IRCCS di diritto pubblico e quelli di diritto privato, pur con una lieve differenza, nei tre dei principali social analizzati (Facebook, Twitter e LinkedIn) sono sempre gli IRCCS di diritto pubblico ad avere un numero maggiore di presenza sui social in termini percentuali (Fig. 3.2a).

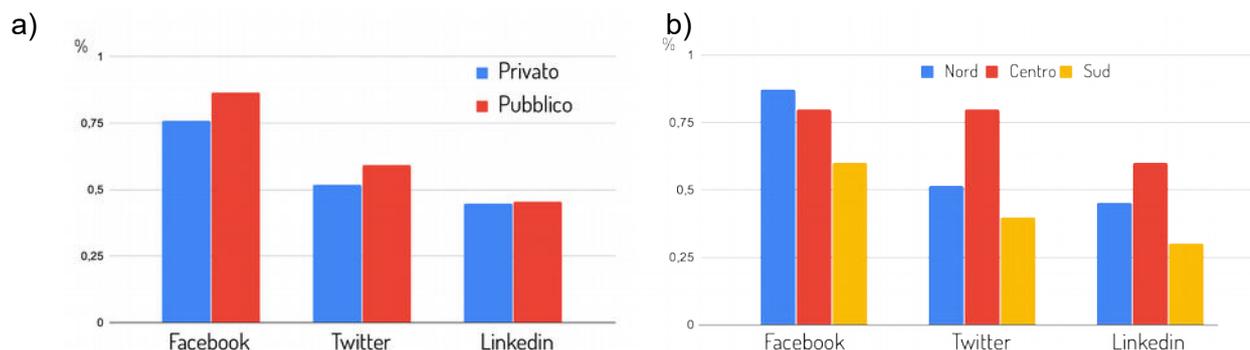


Figura 3.2. La distribuzione dell'uso dei social nelle categorie pubblico e privato (a). Suddivisione geografica (b).

Nella suddivisione per regione (Fig. 3.2b), la tendenza è quella di una minore diffusione social da nord a sud, con l'unica eccezione della diffusione al centro, che ha un valore di diffusione dei 3 social molto simile (Roma accentra la gran parte dei valori).

Per gli scopi di questo progetto, lo studio dell'interazione social tra gli IRCCS, raggiungibile solo con un'estrazione e aggregazione automatica dei dati, i due social candidati all'analisi risultavano chiaramente Facebook e Twitter. La scelta tuttavia è ricaduta su Twitter per due motivi: in prima istanza per la diffusione che ha Twitter in ambito di ricerca, risultando quindi più adeguato a studiare le interazioni tra gli istituti di ricerca piuttosto che con il cittadino comune, secondariamente poichè l'accesso alle API presentava meno restrizioni. Facebook recentemente ha introdotto un sistema di autenticazione all'uso delle API molto vincolante e non idoneo agli scopi di questo progetto.

Dei 28 istituti di ricerca che hanno un account Twitter, non tutti compariranno nei risultati a seguire, in quanto alcuni di questi account presentano un'attività scarsa o nulla, restituendo di conseguenza valori nulli in determinate estrazioni di dati.

### 3.2 L'analisi degli account Twitter

Ogni account Twitter dei 28 IRCCS selezionati è stato caratterizzato con un insieme di attributi: data di iscrizione, numero di follower, frequenza di pubblicazione dei post. I singoli post sono stati poi caratterizzati dagli attributi: numero di like, numero di retweet. L'insieme di questi attributi serve a dare una stima complessiva dell'**engagement** creato dal singolo account.

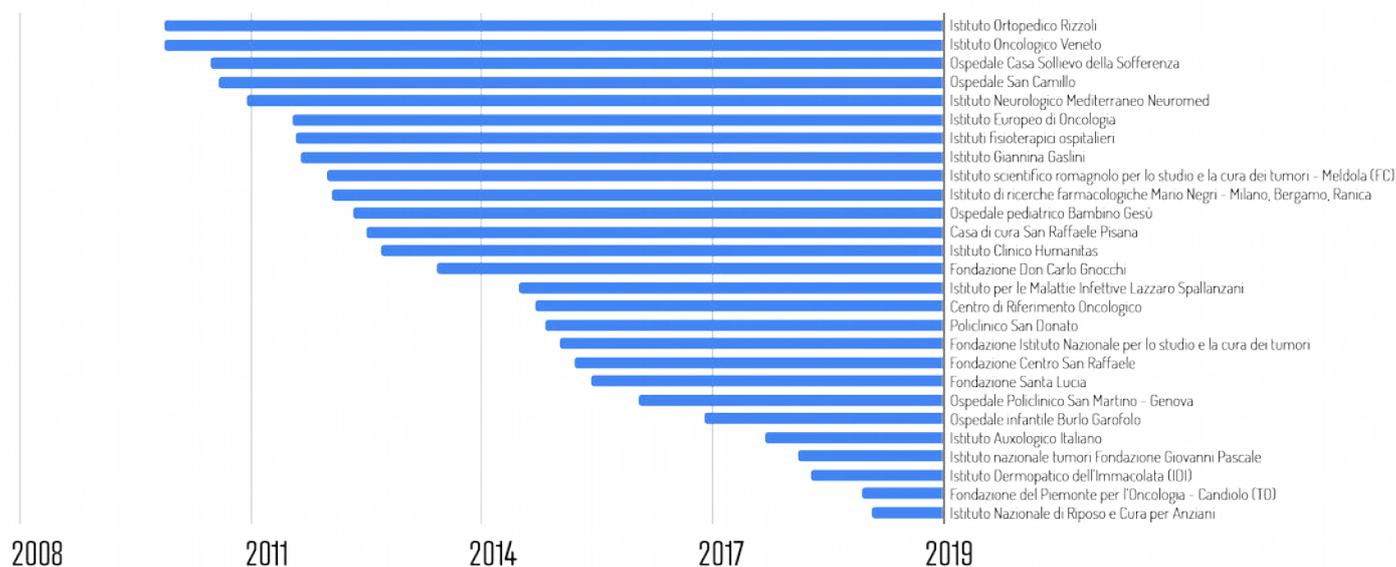


Figura 3.3. La data di iscrizione a Twitter per ogni IRCCS.

La data di iscrizione per ogni account è rappresentata dal diagramma in Fig. 3.3. Twitter è nato nel 2006, ma la diffusione pubblica di larga scala è stata nel periodo 2007-2009, in concomitanza con i primi account di alcuni IRCCS. La distribuzione delle date di registrazione è tuttavia molto disomogenea, con alcuni esempi la cui registrazione risale al 2018. La sola data di registrazione non è tuttavia sufficiente a spiegare il maggiore o minore pubblico raggiunto dai diversi istituti. Nel diagramma di Fig. 3.4 la dimensione dei cerchi è proporzionale al numero di post di ciascun IRCCS, mentre la posizione lungo l'asse delle y indica il numero di follower. L'ascissa invece ordina gli elementi secondo il tempo trascorso tra un post e l'altro degli ultimi 50 post presi in esame.

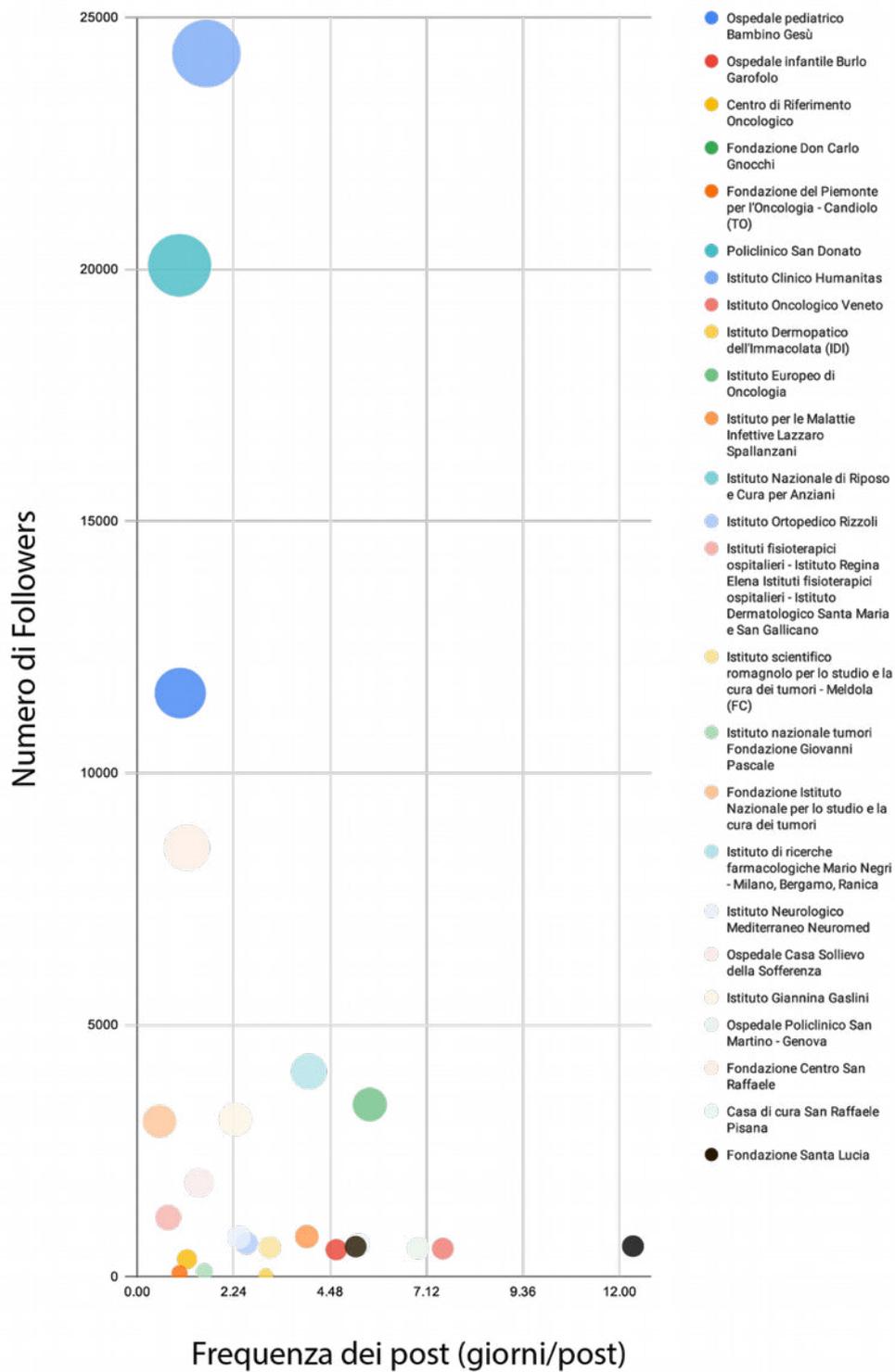


Figura 3.4. La relazione tra la frequenza dei post e il numero dei follower. La dimensione dei punti rappresenta il numero complessivo di post.

La concomitanza di un alto numero di follower in presenza di una frequenza elevata di post lascia supporre che mantenere un livello di attività costante sui social è una condizione necessaria ma non sufficiente a raggiungere un vasto pubblico.

Il numero di follower tuttavia è solo un primo indizio della popolarità di account e quindi dell'efficacia di un'attività sui social media. Per misurare il grado di attività dei follower siamo ricorsi a due altri parametri: il numero di like e di retweet ai post. In tabelle 3.1 sono riassunti i valori medi per ciascun istituto di ricerca, sempre in relazione agli ultimi 50 post pubblicati alla data del 30 novembre 2019.

	Like	Retweet
Ospedale pediatrico Bambino Gesù	17.42	18.22
Ospedale infantile Burlo Garofolo	2.12	1.06
Centro di Riferimento Oncologico	1.44	2.24
Fondazione Don Carlo Gnocchi	3.12	1.46
Fondazione del Piemonte per l'Oncologia - Candiolo (TO)	NA	NA
Policlinico San Donato	6.30	1.14
Istituto Clinico Humanitas	2.06	1.04
Istituto Oncologico Veneto	1.10	1.32
Istituto Dermopatico dell'Immacolata (IDI)	0.10	0.04
Istituto Europeo di Oncologia	6.02	3.34
Istituto per le Malattie Infettive Lazzaro Spallanzani	5.48	5.08
Istituto Ortopedico Rizzoli	0.54	8.30
Istituti fisioterapici ospitalieri - Istituto Regina Elena Istituti fisioterapici ospitalieri - Istituto Dermatologico Santa Maria e San Gallicano	4.00	3.14
Istituto scientifico romagnolo per lo studio e la cura dei tumori - Meldola (FC)	2.58	1.30
Istituto nazionale tumori Fondazione Giovanni Pascale	1.44	0.46
Fondazione Istituto Nazionale per lo studio e la cura dei tumori	2.36	4.20
Istituto di ricerche farmacologiche Mario Negri - Milano, Bergamo, Ranica	2.08	7.34
Istituto Neurologico Mediterraneo Neuromed	2.14	1.22
Ospedale Casa Sollievo della Sofferenza	1.44	1.16
Istituto Giannina Gaslini	3.20	4.18
Ospedale Policlinico San Martino - Genova	8.06	3.50
Fondazione Centro San Raffaele	6.24	2.38
Casa di cura San Raffaele Pisana	1.00	0.48

Fondazione Santa Lucia	4.32	1.58
Ospedale San Camillo	0.32	0.34

Tabella 3.1. Numero medio di post e di retweet per ciascun IRCCS



Il numero medio di post è un primo indice di valutazione. Nei casi che si discostano però dal valore medio atteso, come per l'ospedale di ricerca Bambino Gesù, è stata fatta una verifica del tipo di post pubblicati. Il risultato trovato è che alcuni di questi post sono stati fatti in concomitanza di un evento con personaggi dello spettacolo. Il target è quindi molto più ampio e permette di raggiungere valori di uno o due ordini di grandezza superiori a quelli medi. Un esempio di post con oltre 300 like è mostrato in Fig. 3.5.

Figura 3.5. Post in occasione dell'evento *Una serata di stelle*

### 3.3 L'analisi delle conversazioni attraverso Twitter

Gli strumenti di analisi delle conversazioni permettono di classificare i post in base alla semantica delle frasi, ma anche trovare nuova informazione che va oltre la semantica. Un esempio è lo studio della frequenza delle parole in una ampia base di dati. Nel caso in esame, gli ultimi 50 post, a cui è consentita una lunghezza massima di 280 caratteri, rappresentano una quantità di parole sufficiente a una analisi di frequenza. Per ogni istituto è stato stilato un elenco di parole usate in ordine decrescente di frequenza, rappresentato poi in forma grafica come negli esempi di Fig. 3.6 e Fig. 3.7, che sono due casi rappresentativi di parole maggiormente usate. Con un procedimento simile si possono analizzare gli hashtag maggiormente usati. L'analisi degli hashtag ha il vantaggio di presentare dei dati già significativi per quello che riguarda l'argomento su cui il post vuole richiamare l'attenzione. Lo strumento dell'hashtag infatti è costruito appositamente per dare una sintesi delle parole







esempio, la frequenza degli hashtag per l'IRCCS Bambino Gesù). Rilevante il caso di Humanitas che, nei 50 post considerati, non ha mai usato gli hashtag.

### 3.4 Lo studio delle interazioni tra gli IRCCS usando Twitter

Da ultimo, dopo l'analisi dell'engagement e l'analisi delle conversazioni, è stata fatta l'analisi delle interazioni. Per ogni IRCCS sono stati considerati gli IDs degli account che hanno fatto il retweet di almeno un post. Tra questi ci possono essere altri IRCCS, oppure account che seguono e interagiscono con più di un IRCCS. Questo insieme di relazioni, espresso sotto forma di grafo (Fig. 3.11), permette di visualizzare alcuni risultati importanti.

- 1) Il software Cytoscape riorganizza in modo automatico la posizione dei nodi per similitudine di relazione. Questo permette di visualizzare in modo grafico la distribuzione delle relazioni. Nel caso degli IRCCS la configurazione di small world (rif) per ogni istituto è fortemente accentuata. Alcuni IRCCS addirittura non presentano nessuna relazione condivisa con l'esterno, e sono rappresentati come sottografi isolati nel margine in basso a sinistra della figura.
- 2) Il parametro di **inbound/outbound** degree quantifica il numero di relazioni dirette a uno stesso nodo o provenienti da uno stesso nodo. Nel caso in esame, è un indice di quanto uno stesso account interagisca ripetutamente con uno degli IRCCS. La maggior parte degli account interagiscono in modo occasionale, riferendosi all'insieme considerato di 50 post. Pochi account hanno un'interazione costante.
- 3) Nella configurazione small world pochi nodi fanno da ponte (*bridge*) tra un account e l'altro. Da un'analisi a campione di questi nodi risulta che sono ricercatori o che si tratta di account correlati a livello aziendale.



## Capitolo 4

### Discussione e Conclusioni

#### *4.1 Diffusione dei social media nella ricerca sanitaria*

Gli istituti di ricerca esaminati in questo progetto rappresentano una situazione intermedia tra la struttura sanitaria e il centro di ricerca tradizionale. Nel valutare la diffusione dell'uso dei social media, bisogna sia tenere in considerazione i dati presenti in letteratura per le ASL e le strutture sanitarie [Terry, M. (2009)], sia i dati di diffusione dei social media nella ricerca [Van Eperen, L. and Marincola, F. M. (2011)].

Negli ultimi anni sta crescendo l'attenzione sulla correlazione esistente tra le citazioni di un lavoro e gli indici che caratterizzano il livello di diffusione social dell'articolo stesso. Il più diffuso di questi indici è Altmetric (<https://www.altmetric.com/>), composto da diversi parametri (tra cui anche il numero di tweet e di retweet dell'account Twitter collegato al lavoro scientifico). L'utilizzo in modo attivo di un account social non è più quindi solo una vetrina, ma diventa un vero e proprio strumento per l'attività di ricerca [Azer, S. A. and Azer, S. (2019), Coret, M. et al. (2019), Nocera, A. P. et al. (2019)].

La presenza di quattro istituti su cinque in almeno un canale social è un risultato in linea con la sempre maggiore diffusione delle piattaforme social per la ricerca [Dudo, A. (2013)] e con i molteplici usi che un ricercatore può fare dei social per la sua ricerca, dal promuovere il proprio lavoro (Springer Nature Social Media Survey 2017, <https://tinyurl.com/tema29x>), al reclutamento di soggetti sperimentali [Tweet, M. S. et al. (2011)], allo scambio di idee e alla creazione di nuove collaborazioni (Nature, <https://tinyurl.com/v78m6cx>). Allo stesso tempo il risultato si interseca con la sempre maggiore attenzione ai social in ambito sanitario [Sinnenberg, L. et al. (2017)].

Tuttavia un account su una piattaforma social è solo il primo passo di un uso efficiente dello strumento. Il passo successivo è la creazione di una community, data dal numero di follower ma anche dall'interazione che i follower hanno con il centro IRCCS. In questo aspetto alcuni istituti si distanziano dagli altri come numero di follower (Ospedale Bambino Gesù, Policlinico San Donato e Humanitas si distanziano di quasi un ordine di grandezza),

ma in molti casi non si può dire lo stesso quando poi si va a studiare l'interazione con il pubblico.

Un parametro interessante sarebbe valutare in modo dinamico il variare del numero di follower con l'attività di pubblicazioni e con l'engagement prodotto dalle pubblicazioni. In altre parole rispondere alla domanda: quanti e quali post realizzano un engagement sufficiente a creare nuovi follower? Per rispondere a questa domanda serve tuttavia un follow-up continuo e prolungato nel tempo. Per gli scopi di questo progetto si è valutata invece una situazione statica, che certamente fornisce un ritratto solo parziale delle dinamiche social per un IRCCS. Risulta tuttavia evidente che il solo numero di follower non garantisce un elevato engagement dei post. Il numero medio di like e di retweet resta nella maggior parte degli IRCCS un valore basso, nell'ordine delle poche unità. Andando poi ad analizzare i singoli post, ci si accorge che la varianza è elevata e di fatto solo alcuni post ottengono un elevato numero di interazioni. Il dato è particolarmente evidente nel caso dell'IRCCS Bambino Gesù, che ha dei valori medi di like e retweet rispettivamente di 17.42 e 18.22, outlier rispetto alla media degli altri istituti. Analizzando però i singoli post in esame, si nota che i valori medi sono polarizzati da alcuni post dove si promuove la collaborazione tra l'ospedale Bambino Gesù e personaggi famosi dello spettacolo. Argomenti non caratteristici della ricerca sanitaria, che infatti ottengono risultati molto più vicini a quelli del mondo dello spettacolo.

Risulta invece confermata la correlazione tra l'engagement complessivo (numero follower, like, retweet) e la frequenza dei post. Gli IRCCS con un alto numero di follower sono quelli che postano con una frequenza elevata. Non si può invece dire il contrario, ossia che gli IRCCS che postano con frequenza abbiano un elevato numero di follower.

#### *4.2 Analisi delle conversazioni*

Come visto nel precedente paragrafo, un elevato numero di follower e un elevato engagement dei singoli post esula dall'efficacia dei post in tema di diffusione della ricerca, di prevenzione, di divulgazione dei risultati scientifici. Il post con il personaggio televisivo o il cantante famoso genera sicuramente un elevato traffico sull'account, ma quanto contribuisce a quelle che sono realmente le finalità dei canali social di un IRCCS? Il singolo post di tipo mediatico viene diluito in una base di dati delle conversazioni di un elevato numero di post.

In particolare su Twitter, dove i post hanno la lunghezza limitata di 280 caratteri, la lunghezza di ogni post si mantiene costante e non è correlata alla diffusione del post. Valutando la frequenza delle parole, di cui in questa Tesi sono rappresentati alcuni casi esemplificativi, si nota una migliore aderenza a quelli che sono i temi della ricerca scientifica e dell'attenzione alla salute pubblica.

Le parole più usate possono essere correlate a delle patologie (cancer, riabilitazione, tumori, pazienti, ecc.), a degli eventi (notte dei ricercatori, eventi, settembre, ecc.) o alla promozione di risultati ottenuti all'interno dell'istituto (scopri, leggi, guarda, ecc.). In due istituti abbiamo trovato il risultato curioso in cui la parola più usata è un carattere UNICODE che rappresenta una mano che indica: . La seconda parola più usata in questi due casi è “scopri”. Inequivocabilmente la maggior parte dei post sfruttano il meccanismo del **call-to-action**, cioè l'invito a interagire eseguendo un'azione (nello specifico l'azione di click).

#### *4.3 Le interazioni tra IRCCS e follower degli IRCCS*

Il risultato del grafo delle interazioni tra IRCCS e utenti attivi nei confronti dell'IRCCS mostra un primo risultato evidente: la configurazione a small world dei singoli centri. Escludendo i sei centri che non presentano nessuna interazione reciproca, anche nei rimanenti casi è fortemente sbilanciato il numero dei collegamenti tra centri diversi, a favore di una elevata densità di connessioni con il proprio pubblico. Tra le connessioni interne spicca poi un elevato numero dovuto a follower occasionali. Un'analisi non mostrata in questa Tesi ha ripetuto la creazione del diagramma di relazioni, prendendo in considerazione solo i retweet effettuati dai follower di ciascun istituto. Il risultato si rivela ancora più scarno, dimostrando che molte delle interazioni sono in realtà occasionali e uniche.

Diverso invece il risultato in caso di retweet ripetuti. Spesso l'account che replica più volte l'interazione è anche il nodo di collegamento tra più IRCCS. Il collegamento può avvenire perchè le due strutture collegate hanno anche un legame di tipo organizzativo (ospedali che fanno parte di uno stesso gruppo aziendale) e quindi c'è una figura di social media manager che promuove l'interazione tra i due account, o in modo spontaneo, per esempio attraverso un ricercatore con legami di lavoro con entrambi gli istituti. Nel secondo caso ci si aspetterebbe che, per la naturale propensione del lavoro del ricercatore a fare network, il numero di queste

relazioni-ponte fosse molto maggiore. Seppure limitata nella finestra temporale di post, questa analisi lascia presagire due possibili cause: da una parte l'eventualità che tra i centri IRCCS la mobilità dei ricercatori sia poca, dall'altra la dimostrazione che i singoli ricercatori non fanno divulgazione delle ricerca facendo riferimento ai canali ufficiali dei propri centri.

#### *4.4 Conclusioni*

Questo progetto di Tesi ha voluto indagare il modo in cui i centri IRCCS utilizzano i social media, in particolare Twitter, e con quali risultati. Il lavoro fatto mostra che c'è una sempre maggiore attenzione da parte dei centri di ricerca, nello specifico della ricerca sanitaria, all'uso dei social media e un incremento del numero di account e dei follower che interagiscono. Nel caso degli IRCCS i temi trattati si rivelano un ibrido tra i temi della ricerca in generale e quelli più specifici della salute. Le interazioni sono create dai singoli ricercatori ma anche dagli utenti degli IRCCS in qualità di pazienti o familiari.

Pur nella limitatezza del quadro temporale analizzato (50 post consecutivi a un dato momento dell'anno), il progetto mostra le potenzialità dello studio delle interazioni attraverso i social media.

In particolare, uno degli scopi del progetto è stato quello di costruire una infrastruttura software di analisi, scalabile a casi più complessi e a una maggiore quantità di dati. La struttura logica delle analisi proposte mostra le potenzialità dei big data nel ricavare nuova informazione e nel comprendere le dinamiche di un sistema.

## Bibliografia

Azer, S. A. and Azer, S. (2019) 'Top-cited articles in medical professionalism: a bibliometric analysis versus altmetric scores', *BMJ Open*, 9(7), p. e029433. doi: 10.1136/bmjopen-2019-029433.

Bean, D. M. et al. (2017) 'Network analysis of patient flow in two UK acute care hospitals identifies key sub-networks for A&E performance', *PLoS ONE. Public Library of Science*, 12(10). doi: 10.1371/journal.pone.0185912.

boyd, danah m. and Ellison, N. B. (2007) 'Social Network Sites: Definition, History, and Scholarship', *Journal of Computer-Mediated Communication*, 13(1), pp. 210–230. doi: 10.1111/j.1083-6101.2007.00393.x.

Collins, K., Shiffman, D. and Rock, J. (2016) 'How Are Scientists Using Social Media in the Workplace?', *PLOS ONE*. Edited by S. Goffredo, 11(10), p. e0162680. doi: 10.1371/journal.pone.0162680.

Coret, M. et al. (2019) 'Twitter Activity Is Associated with a Higher Research Citation Index for Academic Thoracic Surgeons', *The Annals of Thoracic Surgery*. doi: 10.1016/j.athoracsur.2019.09.075.

Dudo, A. (2013) 'Toward a Model of Scientists' Public Communication Activity', *Science Communication*, 35(4), pp. 476–501. doi: 10.1177/1075547012460845.

Leicht, E. A. and Newman, M. E. J. (2008) 'Community Structure in Directed Networks', *Physical Review Letters*, 100(11), p. 118703. doi: 10.1103/PhysRevLett.100.118703.

Moe, H. and Bruns, A. (2014) *Structural Layers of Communication on Twitter, Twitter and Society*.

Nocera, A. P. et al. (2019) 'Examining the Correlation Between Altmetric Score and Citations in the Urology Literature', *Urology*, 134, pp. 45–50. doi: 10.1016/j.urology.2019.09.014.

Santoro, E. et al. (2015) ‘Social media and mobile applications in chronic disease prevention and management’, *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.00567.

Sinnenberg, L. et al. (2017) ‘Twitter as a Tool for Health Research: A Systematic Review’, *American Journal of Public Health*, 107(1), pp. e1–e8. doi: 10.2105/AJPH.2016.303512.

Terry, M. (2009) ‘Twittering Healthcare: Social Media and Medicine’, *Telemedicine and e-Health*, 15(6), pp. 507–510. doi: 10.1089/tmj.2009.9955.

Tweet, M. S. et al. (2011) ‘Spontaneous Coronary Artery Dissection: A Disease-Specific, Social Networking Community-Initiated Study’, *Mayo Clinic Proceedings*, 86(9), pp. 845–850. doi: 10.4065/mcp.2011.0312.

Van Eperen, L. and Marincola, F. M. (2011) ‘How scientists use social media to communicate their research’, *Journal of Translational Medicine*, 9(1), p. 199. doi: 10.1186/1479-5876-9-199.

Vosoughi, S., Roy, D. and Aral, S. (2018) ‘The spread of true and false news online’, *Science*, 359(6380), pp. 1146–1151. doi: 10.1126/science.aap9559.

Voytek, B. (2017) ‘Social Media, Open Science, and Data Science Are Inextricably Linked’, *Neuron*. Cell Press, pp. 1219–1222. doi: 10.1016/j.neuron.2017.11.015.

Watts, D. J. and Strogatz, S. H. (1998) ‘Collective dynamics of “small-world” networks’, *Nature*, 393(6684), pp. 440–442. doi: 10.1038/30918.

## Sitografia

Copertura vaccinale autunno 2019 (Ministero della Salute)  
<https://tinyurl.com/yxy75u55>

Twitter per sviluppatori  
<https://developer.twitter.com>

Pacchetto Rtweet  
<https://rtweet.info/>

Software di analisi dei grafi Cytoscape  
[www.cytoscape.org](http://www.cytoscape.org)

Elenco IRCCS (Ministero della Salute)  
<https://tinyurl.com/ru52859>

Altmetrics  
<https://www.altmetric.com/>

Springer Nature Social Media Survey 2017  
<https://tinyurl.com/tema29x>

Online collaborations: scientists and the social media  
<https://tinyurl.com/v78m6cx>

Questo anno di Master è stato un crocevia di stupende opportunità, di persone interessanti, di idee entusiasmanti.

Un ringraziamento particolare al dott. Eugenio Santoro, con il quale ho realizzato questo progetto conclusivo del Master. Grazie per la fiducia avuta nell'organizzare il lavoro, per le opportunità e soprattutto per avermi fatto sentire come una collega di vecchia data tra i ricercatori del suo dipartimento.