

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea Triennale in Informatica

**FENOMENI DI DIFFUSIONE DELLE
NOTIZIE IN SITI DI MICROBLOGGING:
MODELLI DI ANALISI
E STUDIO DI CASO**

Tesi di Laurea in Basi di Dati e Sistemi Informativi

Relatore:
Chiar.mo Prof.
Danilo Montesi

Correlatore:
Dott. Matteo Magnani

Presentata da:
Barbara Iadarola

Sessione III
Anno Accademico 2009/2010

Introduzione

L'informazione è alla base della conoscenza umana. Senza, non si potrebbe sapere nulla di ciò che esiste, di ciò che è stato o di quello che potrebbe accadere. Ogni giorno si assimilano moltissime informazioni, che vengono registrate nella propria memoria per essere riutilizzate all'occorrenza. Ne esistono di vari generi, ma il loro insieme va a formare quella che è la cultura, educazione, tradizione e storia dell'individuo. Per questo motivo è importante la loro diffusione e salvaguardia, impedendone la perdita che costerebbe la dipartita di una parte di sé, del proprio passato o del proprio futuro.

Al giorno d'oggi le informazioni possono essere acquisite tramite persone, libri, riviste, giornali, la televisione, il Web. I canali di trasmissione sono molti, alcuni più efficaci di altri. Tra questi, internet è diventato un potente strumento di comunicazione, il quale consente l'interazione tra chi naviga nel Web (ossia gli utenti) e una partecipazione attiva alla diffusione di informazioni. Nello specifico, esistono siti (chiamati di microblogging) in cui sono gli stessi utenti a decidere se un'informazione possa essere o meno inserita nella propria pagina personale. In questo caso, si è di fronte a una nuova "gestione dell'informazione", che può variare da utente a utente e può defluire in catene di propagazione (percorsi che compiono i dati e le notizie tra i navigatori del Web) dai risvolti spesso incerti. Ma esiste un modello che possa spiegare l'avanzata delle informazioni tra gli utenti? Se fosse possibile capirne la dinamica, si potrebbe venire a conoscenza di quali sono le informazioni più soggette a propagazione, gli utenti che più ne influenzano i percorsi, quante persone ne vengono a conoscenza o il tempo per cui resta

attiva un'informazione, descrivendone una sorta di ciclo di vita.

E' possibile nel mondo reale trovare delle caratteristiche ricorrenti in queste propagazioni, in modo da poter sviluppare un metodo universale per acquisirne e analizzarne le dinamiche? I siti di microblogging non seguono regole precise, perciò si va incontro a un insieme apparentemente casuale di informazioni che necessitano una chiave di lettura [1] [2]. Quest'ultima è proprio quella che si è cercata, con la speranza di poter sfruttare i risultati ottenuti nell'ipotesi di una futura gestione dell'informazione più consapevole.

L'obiettivo della tesi è quello di identificare un modello che mostri con chiarezza quali sono i passaggi da affrontare nella ricerca di una logica di fondo nella gestione delle informazioni in rete.

Il modello (fig. 1), che si pensa possa essere riutilizzato o rielaborato in base alle proprie necessità, prevede come primo passo la scelta della notizia da studiare. In questo caso di studio si osserva infatti la gestione di un'informazione su un fatto accaduto, ovvero una notizia. Si cercano quindi tutti i messaggi relativi all'evento in un sito di microblogging, acquisendone i dati. Questi vengono poi adattati alle esigenze di chi ne usufruisce, cominciando con un'operazione di pulizia che terminerà con un dataset pronto per essere analizzato. Dall'insieme ricavato si potranno osservare le molteplici attività degli utenti, estraendone vari tipi di reti che potranno indicare i percorsi dell'informazione, studiati da diversi punti di vista. Si giungerà infine a considerazioni più generali, alla ricerca di una chiave di lettura chiara e di cui si possa in futuro usufruire.

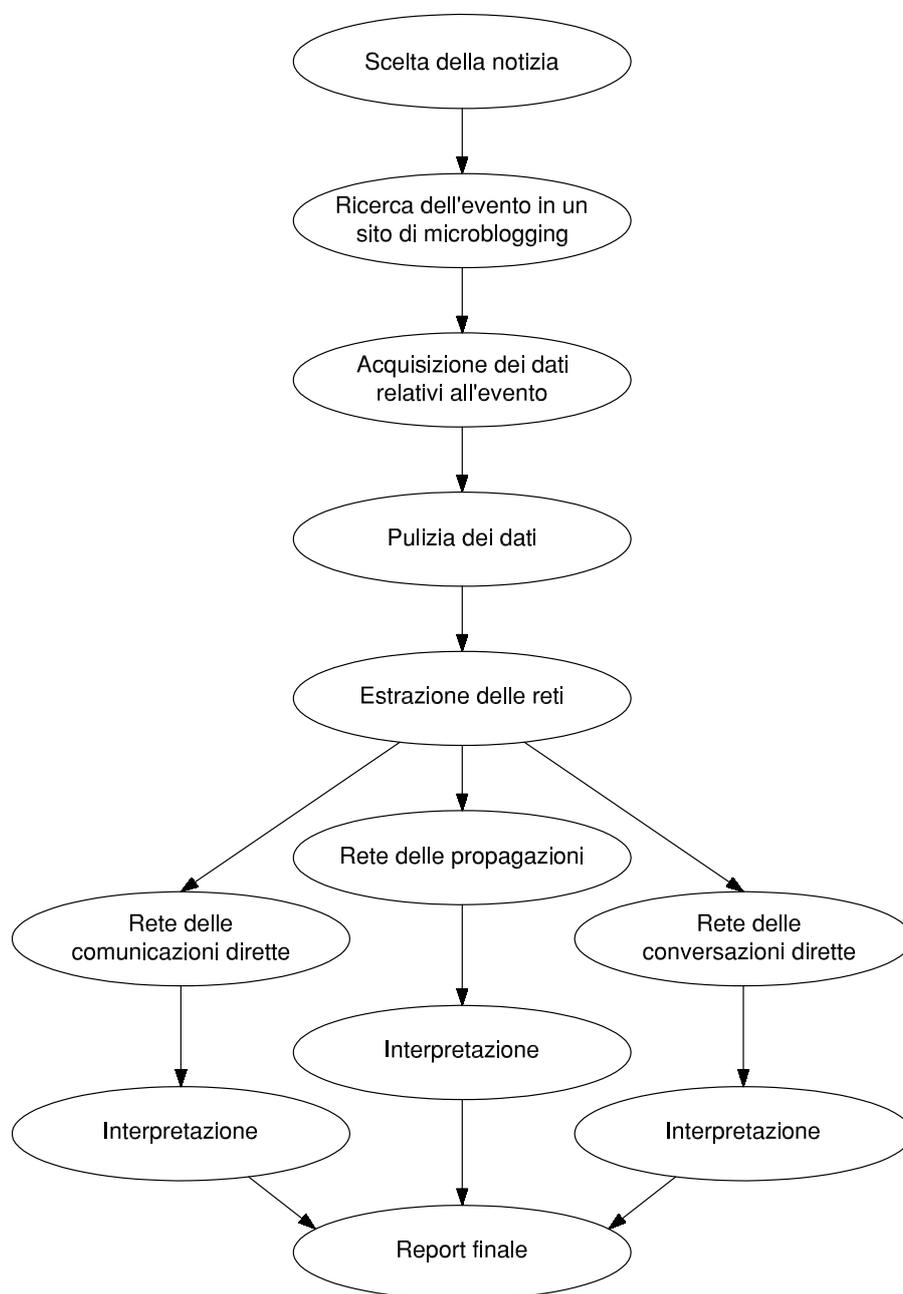


Figura 1: Il modello adottato

I siti di microblogging che sono stati utilizzati per questa ricerca sono FriendFeed e Twitter. La notizia scelta è la liberazione dei 33 minatori cileni, intrappolati nella miniera crollata il 5 agosto 2010 a San José, in Cile. Gli strumenti informatici adottati per la parte iniziale di pulizia dei dati ed estrazione delle reti sono stati rispettivamente PostgreSQL e un programma ideato a partire da un modello di propagazione della notizia, scritto in C++ (e poi automatizzato con uno script). Per la visualizzazione delle reti estratte si sono usati due programmi di modellazione di grafi, Gephi e Graphviz.

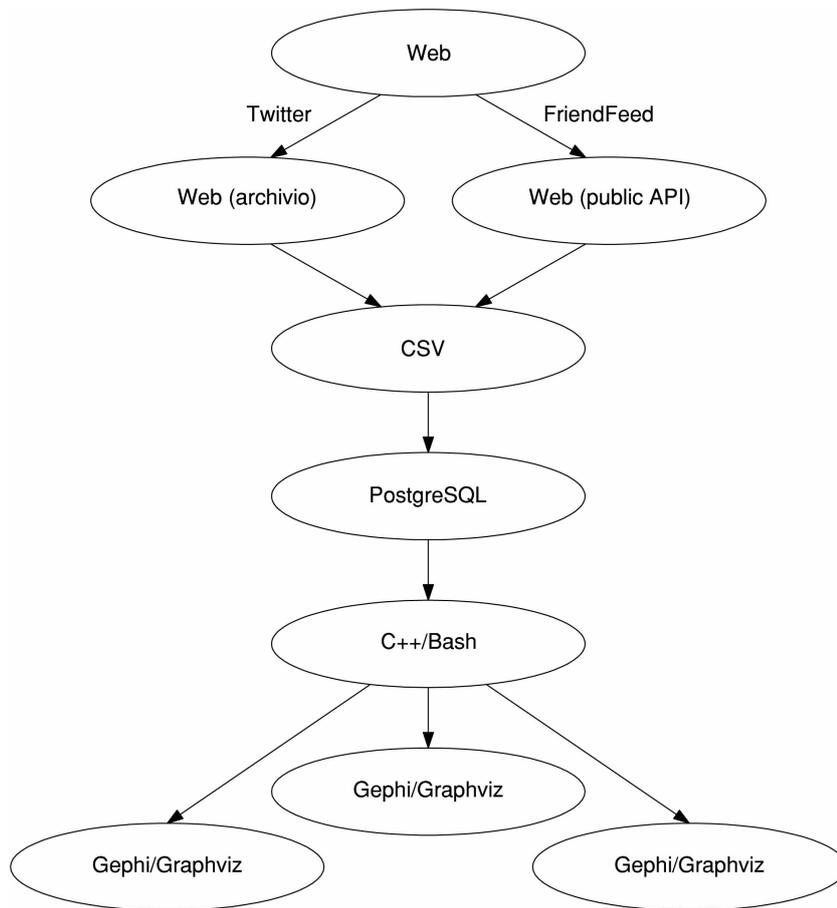


Figura 2: Gli strumenti utilizzati

Il lavoro è stato svolto in collaborazione con il Dott. Matteo Magnani del Dipartimento di Informatica di Bologna, e Luca Rossi del Dipartimento di Scienze della Comunicazione di Urbino [6]. In particolare, essi si sono occupati dell'acquisizione dei dati dai due siti di microblogging e delle interpretazioni finali delle reti estratte [3].

Nei prossimi capitoli verranno approfonditi i componenti di questo modello, indicando per ognuno di essi come si sono utilizzati gli strumenti che ne hanno permesso lo sviluppo (fig. 2).

Il primo capitolo presenta i criteri di scelta della notizia più adatta a questo tipo di osservazioni.

Il secondo capitolo mostra le dinamiche di un sito di microblogging, con la relativa acquisizione dei dati.

Nel terzo capitolo viene spiegato come agire prima di utilizzare i dati, facilitandone la successiva analisi.

Il quarto capitolo enuncia alcune tecniche di estrazioni di reti per l'osservazione della notizia.

A concludere, il quinto capitolo espone alcune interpretazioni a seguito delle analisi svolte.

Indice

Introduzione	i
1 Scelta della notizia	1
1.1 Prime ricerche	1
1.2 Crollo della miniera di San José, in Cile	3
1.3 Scelta dell'intervallo temporale	5
2 Ricerca dell'evento in siti di microblogging	7
2.1 FriendFeed	7
2.2 Twitter	8
2.3 Acquisizione dei dati relativi all'evento	9
3 Pulizia dei dati	11
3.1 La creazione di tabelle	11
3.1.1 Ricerca delle parole chiave	14
3.1.2 Numero di post al giorno	14
3.1.3 Visualizzazione delle conversazioni	15
4 Estrazione delle reti	17
4.1 Rete delle comunicazioni dirette	20
4.2 Rete delle propagazioni	23
4.2.1 Modello di propagazione	24
4.2.2 Rappresentazione degli utenti	24
4.2.3 La creazione della rete	26

4.2.4	Propagazioni	27
4.3	Rete delle conversazioni dirette	33
5	Interpretazioni	35
5.1	Rete delle comunicazioni dirette	35
5.2	Rete delle propagazioni	37
5.3	Rete delle conversazioni dirette	37
5.4	Esposti nel tempo	38
	Conclusioni	41
	Bibliografia	45

Elenco delle figure

1	Il modello adottato	iii
2	Gli strumenti utilizzati	iv
1.1	Le 10 maggiori sorgenti delle entry	2
1.2	Miniera di San José	4
1.3	Numero di post al giorno in FriendFeed	6
2.1	Il logo di FriendFeed	7
2.2	Il logo di Twitter	8
3.1	Esempio di conversazione su FriendFeed	16
3.2	Esempio di rete dei contatti di FriendFeed	16
4.1	Esempio di rete	19
4.2	Rete delle comunicazioni dirette in Twitter	21
4.3	Rete delle comunicazioni dirette in FriendFeed	22
4.4	Esempio di propagazione	23
4.5	Stato degli utenti durante la propagazione di notizie	24
4.6	Esposti nel tempo in FriendFeed	30
4.7	Esposti nel tempo in Twitter	31
4.8	Percorsi di propagazione in FriendFeed	31
4.9	Percorsi di propagazione in Twitter	32
4.10	5 conversazioni più lunghe in FriendFeed	34
4.11	5 conversazioni più lunghe in Twitter	34

Capitolo 1

Scelta della notizia

Individuare la tipologia di informazione che l'utente maggiormente prende in considerazione è fondamentale per una corretta osservazione della loro gestione in rete. Notizie improvvise e inaspettate, che coinvolgano una sola persona (di solito famosa, o scomparsa) o molteplici, richiamano a sé l'attenzione di molti, e sono quelle che hanno maggiore probabilità di rimanere attive per più tempo. La ricerca si è orientata quindi su notizie di stampo sia nazionale che internazionale di un certo impatto (principalmente emotivo), come la morte di un personaggio famoso, o una catastrofe naturale.

1.1 Prime ricerche

Per questo lavoro, si è utilizzato un insieme di dati già pronto, procurato dal sito di FriendFeed. La scelta della notizia doveva ricadere, cioè, in un periodo prestabilito, più esattamente dal 1 Agosto 2010 al 17 Ottobre 2010. Inizialmente il dataset era composto dai messaggi fino al 30 Settembre 2010 con 12,5 milioni di entry, 3,7 milioni di commenti, 800000 like e 28 milioni di relazioni tra utenti (nella rete dei contatti) [4]. È stato interessante notare quali fossero le sorgenti da cui provenivano i messaggi, essendo FriendFeed un aggregatore di aggiornamenti. La percentuale è stata ricavata in modo semi-automatico e precedentemente l'inizio della ricerca, mostrando informazioni

sui siti maggiormente utilizzati nel Web (fig. 1.1).

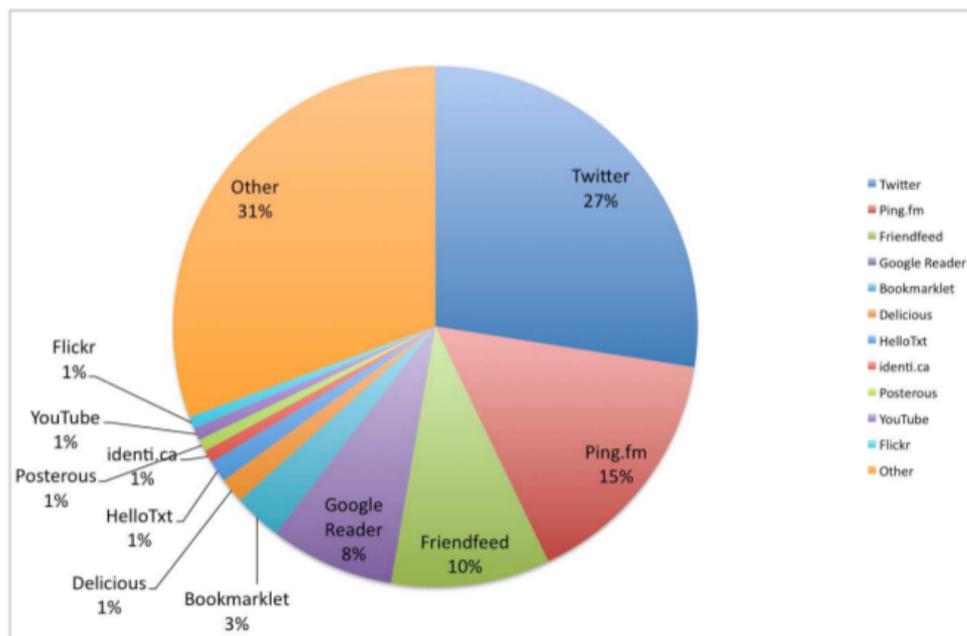


Figura 1.1: Le 10 maggiori sorgenti delle entry. Fonte: [4]

Il dataset è stato successivamente ampliato, includendo i messaggi fino al 17 Ottobre. Esso conteneva 3,3 milioni di entry, 900000 commenti e 200000 like, ed è sull'insieme totale che si sono svolte le ricerche delle notizie più interessanti.

Agli inizi di Agosto 2010, la British Petroleum ha annunciato il successo dell'operazione "Static Kill" che consisteva nel richiudere a 106 giorni dall'esplosione della piattaforma Deepwater Horizon il pozzo di petrolio sottomarino che ha originato la marea nera nel Golfo del Messico [7]. Si ricordi che l'incidente, avvenuto il 20 aprile 2010, è stato dichiarato come la più grave catastrofe ambientale nella storia degli USA, con gravi conseguenze sia per le popolazioni lì residenti che per la flora e la fauna locali. Questa notizia tuttavia non ha ricevuto l'attenzione di molti utenti, i quali si sono limitati tra l'altro a pubblicare l'informazione on line senza praticamente interagire

gli uni con gli altri. Dopo la raccolta dei dati e la successiva analisi si è deciso quindi di accantonarla.

Un'altro evento di rilievo preso inizialmente in considerazione è stato quello della morte dell'ex presidente della Repubblica, Francesco Cossiga, avvenuta il 17 Agosto [8]. Un numero abbastanza consistente di utenti (italiani) ha propagato la notizia, commentando con opinioni proprie l'avvenimento, e lasciando intendere la presenza di due correnti di pensiero totalmente opposte sulla sua moralità. Si è però deciso di cercare un evento che coinvolgesse non solo il popolo italiano ma anche quello mondiale.

La morte di un personaggio famoso come l'attore Tony Curtis (avvenuta il 29 Settembre 2010 [9]) si è rivelata migliore delle due precedenti, poiché il numero dei messaggi era notevolmente maggiore, ma non quanto quelli relativi a un altro evento avvenuto agli inizi di Agosto, descritto nella prossima sezione.

1.2 Crollo della miniera di San José, in Cile

Il 5 Agosto 2010 a San José, in Cile, è crollata una miniera sotterranea lasciando intrappolati 33 minatori a circa 700 metri di profondità [10]. L'evento ha scatenato un forte interesse mediatico, innescando anche le proteste di chi non si trovava d'accordo nel pubblicizzare un avvenimento così drammatico. La vicenda si sviluppa nell'arco di poco più di due mesi: dopo l'incidente, il primo tentativo di soccorso viene interrotto il 7 agosto, a causa di un crollo secondario che blocca il condotto di ventilazione, scelto come via d'accesso per i soccorritori (fig. 1.2). I lavori proseguono dal giorno successivo, perforando vari canali per localizzare i minatori, ma senza successo. Il 22 Agosto, grazie a un contatto tra i minatori e i tecnici in superficie, si scopre che i minatori sono vivi: hanno occupato un rifugio e le gallerie circostanti [11]. Sono quindi stati inviati ai 33 uomini beni di prima necessità, e le operazioni hanno proseguito con maggiore ottimismo fino a quando una delle tre trivellatrici impegnate nei nuovi scavi (di un tunnel largo a sufficienza per

estrarre e far passare i minatori) si rompe su una roccia dura, col rischio di compromettere i lavori finora effettuati [12]. Nei giorni successivi si utilizzano nuove strutture per le trivellazioni, e una volta creato il canale è bastato allargarlo a sufficienza per permettere all'ascensore col compito di estrarre i minatori di passare [13].

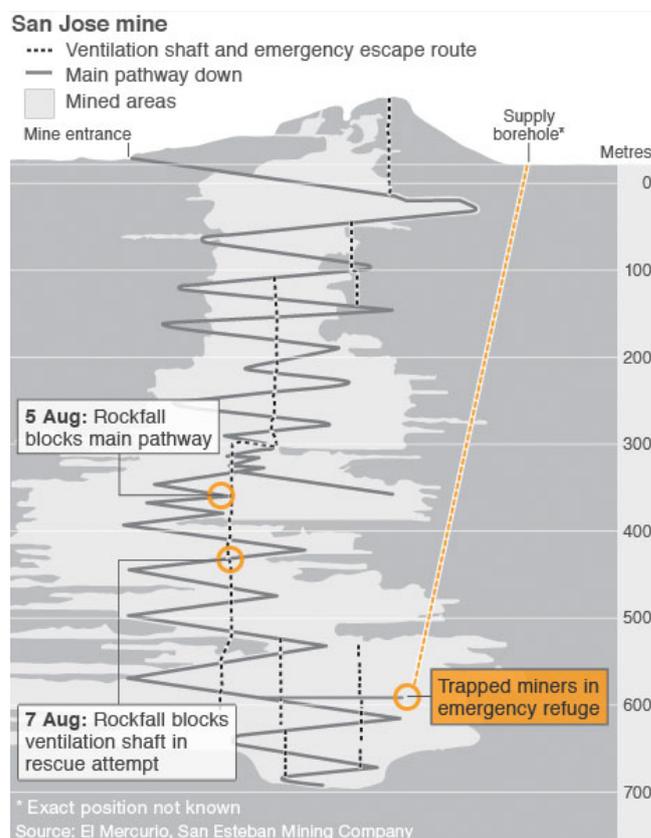


Figura 1.2: Miniera di San José.

Fonte: <http://www.ilpost.it/files/2010/08/Miniera-San-Jose.jpg>

1.3 Scelta dell'intervallo temporale

Da un avvenimento ampio come quello considerato è emersa una suddivisione in eventi consecutivi minori che hanno destato più o meno attenzione da parte degli utenti della rete.

La liberazione dei 33 ha visto la partecipazione di numerosissimi giornalisti, amici e famigliari, ma in particolare dei soccorritori che per 66 volte hanno percorso i 622 metri con la capsula di salvataggio [14]. Con velocità di 1 metro al secondo, essa ha riportato in superficie ogni minatore, in 15-20 minuti a volta. Passati undici minuti dalla mezzanotte locale, viene salvato il primo minatore cileno, tornato in superficie dopo oltre due mesi di prigionia. A seguire, per la durata complessiva di 22 ore, tutti gli altri, 31 cileni e 1 boliviano. Finisce quindi a lieto fine l'operazione ribattezzata "San Lorenzo", come il patrono dei minatori.

La scelta è ricaduta su questa parte dell'evento, per non lavorare su un insieme di dati troppo discontinuo ed eterogeneo, quale era quello iniziale. Successive analisi del dataset totale hanno confermato questa soluzione (fig. 1.3), come esposto in uno dei prossimi capitoli.

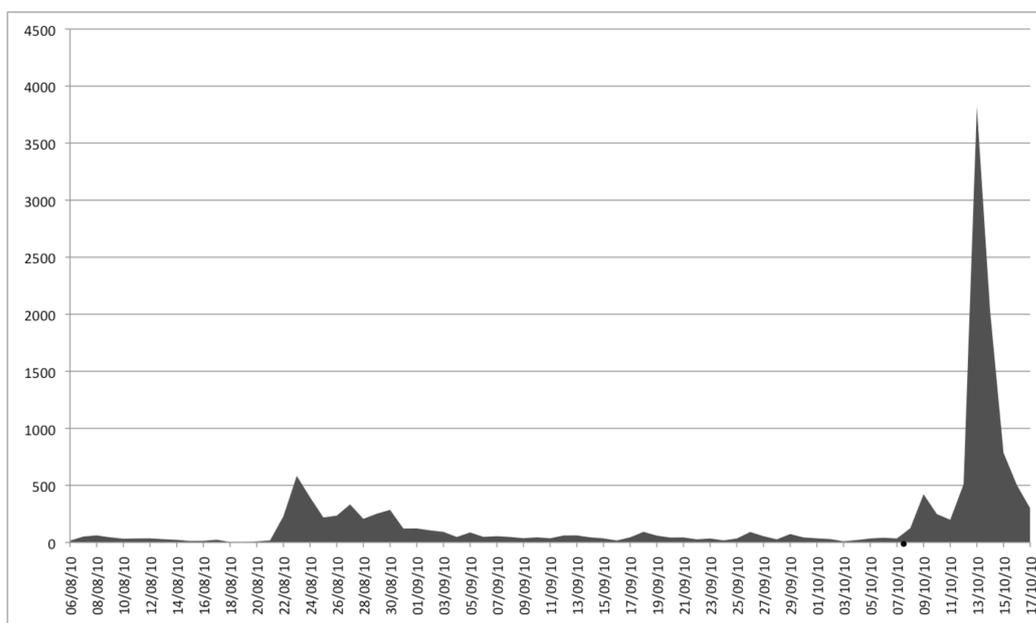


Figura 1.3: Numero di post al giorno in FriendFeed

Capitolo 2

Ricerca dell'evento in siti di microblogging

Il microblogging è una forma di pubblicazione continua di materiale (immagini, video, messaggi, citazioni...) da parte di utenti su servizi di Social Network [15]. Questi contenuti possono essere visualizzati da tutti oppure solamente da chi segue l'utente, ovvero i suoi “follower”. Due dei più famosi siti di microblogging, e quelli che sono stati utilizzati in questa ricerca, sono FriendFeed e Twitter.

2.1 FriendFeed



Figura 2.1: Il logo di FriendFeed

FriendFeed (logo in fig. 2.1) è un servizio Web che consente agli utenti di pubblicare in tempo reale contenuti di varia natura, commentare i post di altri utenti e apprezzarli (ovvero esprimere un “like”) [16]. Gli aggiornamenti provenienti da reti sociali, blog e tumblelog vengono altresì inseriti, creando

un unico flusso di informazioni che rappresenta le attività dell'utente nella rete.

È semplice identificare le discussioni all'interno del sito, poiché ogni post scritto da un utente (o condiviso da questo) è immediatamente seguito dagli eventuali commenti relativi, compresi i like, rendendo interamente leggibile e accessibile la conversazione in corso. Se si commenta il messaggio di qualcuno, per esempio, la conversazione appare visibile ai propri follower, permettendo quindi l'intrusione nella discussione di un terzo, che può anche non conoscere chi originalmente ha scritto il messaggio (ma solo chi l'ha commentato/apprezzato). In questo modo si possono formare catene di comunicazione fra utenti diversi, non necessariamente tutti collegati tra loro.

2.2 Twitter



Figura 2.2: Il logo di Twitter

Twitter (logo in fig. 2.2) funziona in modo leggermente diverso da FriendFeed. La pagina di ogni utente è aggiornabile con brevi messaggi di testo (Twitter deriva da tweet, che significa “cinguettare”) [17]. In esso può comparire un hashtag, ovvero parole o frasi precedute dal simbolo “#”, che dona una visibilità maggiore al post poiché rintracciabile da parte di altri utenti con una ricerca specifica (come se fosse una parola chiave).

È possibile rispondere ai tweet di un utente o retweetare i messaggi, ovvero ripeterne il testo associandogli la sigla RT e il nome utente del messaggio originale. In questo modo, coloro che seguono chi ha riportato il retweet sono esposti a un messaggio che probabilmente non avrebbero potuto leggere. Seguire una conversazione su Twitter è più complicato, poiché le risposte ai

messaggi degli utenti non sono visibili nel loro insieme, ma appaiono singolarmente nella propria pagina (con la sintassi “@nomeutente” ad indicare a chi ci si rivolge).

2.3 Acquisizione dei dati relativi all'evento

L'acquisizione del dataset relativo alla notizia dei minatori si è svolta in due modi differenti, a seconda del sito preso in considerazione.

Per FriendFeed si è usata la sua timeline pubblica, attraverso cui si è monitorata la quasi totalità delle azioni degli utenti nel sito, con una frequenza di circa 200 messaggi ogni 3 secondi (incluso anche i messaggi che apparivano più volte). In un file csv sono stati salvati solamente i messaggi (le entry), in un secondo tutti i commenti e per ultimi tutti gli apprezzamenti (i like). I file finali contenevano dati dal 1 Agosto al 17 Ottobre, ovvero più di due mesi di interazioni tra utenti.

Per il recupero dei dati su Twitter si è utilizzato invece un archivio online (twapperkeeper.com) in cui si è creato un nuovo archivio a partire da un hashtag. La ricerca degli hashtag non è una ricerca precisa né totale: essi sono scelti personalmente dagli utenti, per cui si può scrivere al suo interno quello che si desidera, anche se riguardo l'argomento esistono hashtag già ampiamente utilizzati. In questo caso, sono stati scelti dall'archivio tutti i messaggi contenente l'hashtag “#mineros” inizialmente, poi “#rescatemineiros” dopo aver ristretto l'analisi dei dati alla sola liberazione dei minatori. Anche qui l'archivio dei messaggi è stato salvato su file csv.

Per entrambi i siti si è salvato in un ulteriore file (per piattaforma) l'insieme degli utenti attivi specificando, riga per riga, per ogni utente un suo follower, inserendoli tutti. Questi file rappresentano così la rete dei contatti degli utenti, fondamentale per conoscere i percorsi delle informazioni sui due siti.

Capitolo 3

Pulizia dei dati

I dati acquisiti dal sito di FriendFeed non possono essere immediatamente analizzati: necessitano di una pulizia mirata che ne estragga solo quelli più utili a ciò che si vuole osservare. Nei prossimi paragrafi verranno esposte tecniche di estrazione con diversi strumenti, che permetteranno di avere una visione più ordinata dell'insieme di dati.

Nel caso di Twitter, non sono state necessarie operazioni di pulizia, avendo il dataset interamente dedicato alla liberazione dei minatori (recuperato tramite l'hashtag *#rescatemineros*). È però probabile che non tutti i messaggi relativi alla vicenda siano stati recuperati, rendendo l'insieme dei dati parzialmente incompleto.

3.1 La creazione di tabelle

Per facilitare la lettura dei file salvati si è scelto di utilizzare PostgreSQL, un database relazionale rilasciato con licenza libera. In poche e semplici operazioni si è caricato su un database il file con tutte le entry, ovvero tutti i messaggi, in una tabella. In un'altra sono stati caricati solo i commenti, in una terza i like e per ultima la rete di contatti. Ogni tabella è stata così costruita:

- ENTRY:

```
CREATE TABLE entry (  
  eid character varying,  
  postedby character varying,  
  sourcename character varying,  
  sourceurl character varying,  
  geox character varying,  
  geoy character varying,  
  date timestamp without time zone,  
  body character varying,  
  numimg integer,  
  imgurl character varying,  
  numimgv integer,  
  imgurlv character varying);
```

dove:

`eid` è l'identificatore del messaggio (assegnato dal sito), `postedby` è l'identificatore dell'autore del messaggio, `sourcename` è la sorgente del messaggio, `sourceurl` è l'indirizzo della sorgente del messaggio, `geox` e `geoy` sono campi vuoti, `date` è la data del messaggio (GMT+1), `body` è il testo del messaggio, `numimg` è il numero delle immagini, `imgurl` è l'URL delle immagini, `numimgv` è il numero dei video e `imgurlv` è l'URL dei video.

- COMMENTI:

```
CREATE TABLE comments (  
  postid character varying,  
  eid character varying,  
  postedby character varying,  
  sourcename character varying,  
  sourceurl character varying,
```

```
geox character varying,  
geoy character varying,  
date timestamp without time zone,  
body character varying);
```

dove:

`postid` è l'identificatore del commento (assegnato dal sito), `eid` è l'identificatore del messaggio commentato (quello originale), `postedby` è l'identificatore dell'autore del commento, `sourcename` è la sorgente del messaggio commentato, `sourceurl` è l'indirizzo della sorgente del messaggio, `geox` e `geoy` sono campi vuoti, `date` è la data del commento (GMT+1) e `body` è il testo del commento.

- LIKE:

```
CREATE TABLE likes (  
userid character varying,  
postid character varying,  
date timestamp without time zone);
```

dove:

`userid` è l'identificatore dell'autore del like, `postid` è l'identificatore del messaggio apprezzato (quello originale) e `date` è la data dell'apprezzamento del messaggio (GMT+1).

- RETE:

```
CREATE TABLE rete (  
follower character varying,  
followed character varying);
```

dove:

`follower` è l'identificatore dell'utente che segue l'utente indicato nella colonna `followed` (fig. 3.2).

Una volta create le tabelle si è eliminata ogni entry che ancora non avesse a che fare con la notizia dei minatori liberati, con una ricerca mirata delle giuste parole chiave.

3.1.1 Ricerca delle parole chiave

Al contrario di Twitter, dove l'hashtag ha permesso il recupero dei soli messaggi relativi alla vicenda, per FriendFeed l'insieme di dati è ancora un caotico agglomerato di messaggi che deve essere filtrato. Questo filtro è stato applicato alla tabella delle entry, dove le parole chiave che hanno permesso l'estrazione dei messaggi corretti sono state:

miner(s), bergleute and Chile, minero(s), minatore/i, San José, San José

eliminando poi tutte quelle che contenessero:

China, Coahuila, Indian, Spanish

che non avevano niente a che fare con l'incidente accaduto in Cile. Il processo di pulizia dei messaggi è stato svolto osservando le entry presenti nella tabella, cercando di eliminare il maggior numero possibile di quelle estranee alla notizia. Dato l'alto numero di messaggi c'è comunque un margine di errore di cui bisogna tenere conto.

3.1.2 Numero di post al giorno

Al fine di individuare il periodo in cui la notizia si fosse sviluppata più ampiamente e più in fretta, è stata utilizzata un'interrogazione SQL:

```
select extract(day from date) as day,  
extract(month from date) as month,  
extract(year from date) as year,  
count(*) from tabellaminatori  
group by day, month, year  
order by year, month, day;
```

che contasse per ogni giorno presente nel dataset il numero di messaggi presenti.

L'interrogazione è stata posta alla tabella contenente tutti i messaggi a partire dal crollo della miniera, dopo la ricerca delle parole chiave. Dovendo svolgere le osservazioni in un periodo breve e con un'alta presenza di messaggi, gli oltre due mesi di prigionia erano eccessivi per una corretta analisi. Si è subito notata una suddivisione in più sotto-eventi dell'avvenimento (come mostrato in figura 1.3).

I dati estratti hanno reso visibile un incremento dell'attività nel sito (e quindi maggiore interesse da parte degli utenti) nella parte finale della vicenda, ovvero la liberazione dei 33. Altri picchi, ma di minore entità, sussistono qualche giorno prima della liberazione, periodo in cui cominciano gli annunci dell'operazione di salvataggio, e attorno al 22 agosto 2010, quando si scopre che i minatori sono vivi grazie ad un loro avviso dal sottosuolo. Si è deciso di mantenere il dataset più ricco, scegliendo tutti e soli i messaggi scritti dalle 12 del 12 Ottobre 2010 alle 12 del 14 Ottobre 2010: 48 ore in cui gli utenti sono stati esposti alla notizia della liberazione, la diretta dell'estrazione e il salvataggio finale dei minatori (adattando anche Twitter alla stessa scelta).

3.1.3 Visualizzazione delle conversazioni

Estratta la tabella con le sole entry relative alla liberazione dei minatori, è stato necessario eliminare alcune colonne inutili. Si è deciso di mantenere, ai fini delle analisi, l'id della entry, chi l'ha scritta, a chi è indirizzata (in questo caso sé stesso), la sorgente del messaggio, la data, il tipo (ovvero l'identificatore dell'azione compiuta nel sito, quindi l'inserimento di una entry 'E') e il testo (tenuto solo per un'analisi qualitativa dei discorsi tra gli utenti).

Per una visione omogenea dell'insieme dei dati si sono unite con un join entry, commenti e like in un'unica tabella. L'interrogazione SQL corrispondente è stata eseguita sulla base dell'id delle entry, dato presente in tutte e tre le tabelle.

Si è ricavata quindi una tabella finale con ogni conversazione avvenuta nel sito riguardo la liberazione, ancora da ordinare ma ormai pronta per essere analizzata. L'ordinamento è stato successivo: prima si sono ordinati i messaggi per id delle entry uguali (cioè si sono raggruppate le conversazioni singole, con commenti e like a loro volta in ordine cronologico), poi si sono ordinate le entry per data. In questo modo si è potuto scandire in ordine ogni conversazione formata, come mostrato in figura 3.1.

entryid	postedby	source	name	date	type	recipient	body
character	character	character	timestamp	wikit	character	character	character varying
e/b5cd	blondeir	Tumblr	2010-10-12	E	blondein	domani inizieranno le operazioni di salvataggio dei trentatré minatori cilen	
e/6f4c	olavale	FriendFe	2010-10-12	E	olavale	#speremaben Diret	
e/6f4c	dyloniso	FriendFe	2010-10-12	L	olavale	/	
e/6f4c	dyloniso	FriendFe	2010-10-12	L	olavale	/	
e/6f4c	olavale	FriendFe	2010-10-12	C	olavale	sembra che inizino a far scendere la capsula..sembra ci vorranno 48 ore per	
e/6f4c	olavale	FriendFe	2010-10-12	C	olavale	allora: prima fanno scendere completamente la capsula, poi un soccorritore.	
e/6f4c	olavale	FriendFe	2010-10-12	C	olavale	probabilmente verso le 22.30-23 salirà il primo..	
e/6f4c	olavale	FriendFe	2010-10-12	C	olavale	..é lunga ancora...	
e/6f4c	olavale	FriendFe	2010-10-13	C	olavale	http://cnrchile.cl/v	
e/6f4c	olavale	FriendFe	2010-10-13	C	olavale	Piñera show	
e/6f4c	pollycof	FriendFe	2010-10-13	C	olavale	uh vero dimenticavo... aggiornamenti?	
e/6f4c	olavale	FriendFe	2010-10-13	C	olavale	fra mezz'ora partono.. fino ad ora fuffa (preparativi di preparativi, contro	
e/6f4c	pollycof	FriendFe	2010-10-13	C	olavale	e non avevo capito che vivi lì, qué bien :-)	
e/6f4c	olavale	FriendFe	2010-10-13	C	olavale	siamo col fiato sospeso da 2 mesi.. anche se la cosa é stato sfruttata med	
e/6f4c	olavale	FriendFe	2010-10-13	C	olavale	Piñera benedice la capsula (..mi ricorda qualcuno..)	
e/6f4c	olavale	FriendFe	2010-10-13	C	olavale	gambe e braccia in posizione (la prova 65 metri)	
e/6f4c	olavale	FriendFe	2010-10-13	C	olavale	...	

Figura 3.1: Esempio di conversazione su FriendFeed

	follower	followed
	character	character
	varying	varying
1	lotsemann1	spaceastronautics
2	algarrido	spaceastronautics
3	almitahp	spaceastronautics
4	alwaysbestrts	spaceastronautics
5	imuck	spaceastronautics
6	anarchists	spaceastronautics
7	andrewlm	spaceastronautics
8	angelcakes85	spaceastronautics
9	experienceseeker	spaceastronautics
10	icue	spaceastronautics
11	daferret	spaceastronautics
12	badruddeen	spaceastronautics
13	foreveradog	spaceastronautics
14	brianbrawdy	spaceastronautics
15	burritosound	spaceastronautics
16	paintingdenver	spaceastronautics
17	radiatorabies	spaceastronautics

Figura 3.2: Esempio di rete dei contatti di FriendFeed

Capitolo 4

Estrazione delle reti

L'insieme di dati raccolto può disorientare chi non conosca le dinamiche dei due siti di microblogging, anche dopo le operazioni di pulizia. Per Friend-Feed l'osservazione dei messaggi è più semplice grazie alla tabella che riporta le singole conversazioni. Su Twitter, invece, vengono ampiamente utilizzati sigle, abbreviazioni e messaggi sintetizzati per rimanere dentro i limiti di testo imposti dalla piattaforma, e le conversazioni non sono immediatamente riconoscibili o definibili.

Un esempio può essere il retweet di una risposta, che contiene più di un'informazione rilevante, ma pecca di difficile comprensione:

RT @BrendLou: @Tito_UK saludos; todo
saldrá bien, dios esta con ellos;
#rescatemineros

Esso è contraddistinto dall'utilizzo di una sigla ("RT"), di caratteri speciali ("@" e "#") seguiti da un nome e di un testo. Questi segni sono indici di un utilizzo ben preciso del sito che permettono agli utenti di instaurare vere e proprie conversazioni. La possibilità di riconoscerli dipende dalla conoscenza che si ha della piattaforma e degli usi comuni degli utenti, che con l'evolversi del tempo hanno utilizzato tecniche comunicative sempre più propense al risparmio di tempo e caratteri.

Analizzando l'esempio citato si nota innanzitutto la presenza della sigla tipica del retweet "RT" che lo identifica come tale. Se si prosegue nella lettura si incontra però un'ambiguità nell'uso della "@": subito dopo il RT il simbolo affianca il nome dell'utente di chi ha scritto il messaggio originale (*BrendLou*), mentre di seguito precede il nome dell'utente verso cui ci si rivolge, individuando una risposta (a *Tito_UK*). Non tutti i messaggi che presentano una "@" sono quindi facilmente catalogabili: c'è la possibilità di trovare un retweet, una risposta, o un retweet di una risposta, come in questo caso. Il "#" permette invece di capire l'appartenenza a una certa discussione (qui, la liberazione dei minatori identificata dall'hashtag *#rescatemineros*), ma non è detto sia effettivamente possibile ricostruire un'unica conversazione globale. In rete gli hashtag possono essere diversi, o addirittura non utilizzati, rendendo difficoltosa l'identificazione delle conversazioni su uno stesso ambito.

Questi sono tutti indizi di ciò che accade in rete, estrapolabili da un solo messaggio. Ognuno di essi ha un significato nel percorso dell'informazione, così come il loro insieme. Uno dei primi problemi in questa analisi rappresenta l'identificazione delle giuste tracce da studiare. Una volta che si è scelto cosa osservare, risulterà utile costruire modelli di osservazione per renderne più semplice lo studio.

Un esempio di ciò che è stato appena esposto, e che riporta uno dei possibili utilizzi del sito di Twitter (più complesso da analizzare), è indicato in figura 4.1. Si supponga che l'utente *A* decida di pubblicare una notizia. Gli utenti *B* e *C* che lo seguono ne potrebbero venire a conoscenza, nonostante non sia effettivamente possibile sapere quando e da chi un utente venga esposto da un altro. Una prima osservazione potrebbe essere allora su come si propaga la notizia tra gli utenti, mettendone in evidenza tutti i percorsi. Nel caso in cui *B* rispondesse ad *A*, però, non si avrebbe più a che fare con una propagazione, ma con l'inizio di una conversazione tra i due. Quindi non è più la notizia a viaggiare in rete, ma dei messaggi che possono essere inerenti o meno all'argomento. In questo caso si potrebbe studiare che tipo

di conversazione si crea in questo sito e cioè cosa si dicano realmente gli utenti. F e G comunicano allo stesso modo sulla medesima notizia, ma nei loro messaggi compare un hashtag differente da quello di A , B e C . Ciò comporta la suddivisione della rete in conversazioni diverse su uno stesso argomento. Dunque un'altra osservazione può essere svolta sul modo in cui le conversazioni siano strutturate in rete.

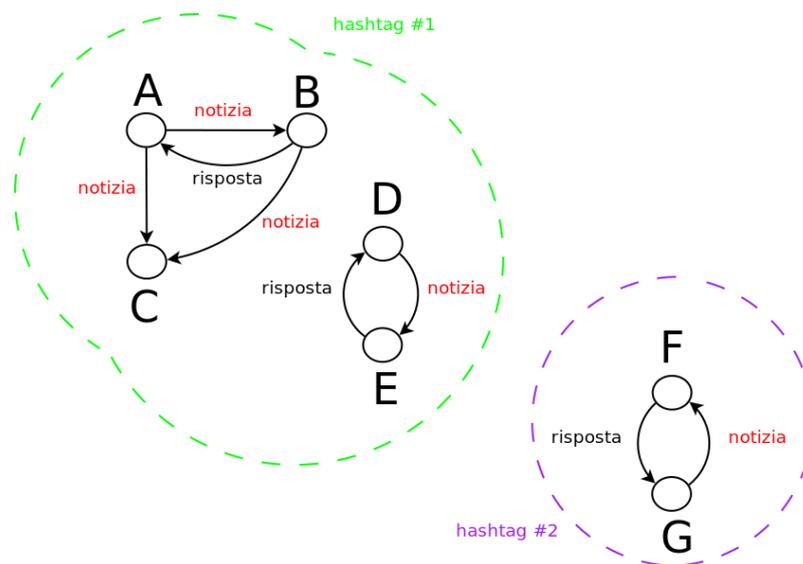


Figura 4.1: Esempio di rete

Queste osservazioni hanno portato alla scelta ed estrazione di 3 tipi di reti per ogni sito di microblogging: una dove i nodi rappresentano gli utenti e gli archi le comunicazioni dirette, una dove gli archi simbolizzano i messaggi di risposta (evidenziando le conversazioni strutturate), e una di utenti dove gli archi disegnano i percorsi di propagazione.

Utilizzando Gephi e Graphviz, due programmi di modellazione di grafi, se ne sono prodotti visualizzazioni diverse, a seconda del programma utilizzato. Gephi permette di creare un grafo partendo da un file che indichi i nodi collegati da un arco. Graphviz, invece, da riga di comando crea un file immagine riferendosi a un file in formato dot, il quale elenca prima tutti i nodi e poi gli archi presenti nel grafo.

Nei prossimi paragrafi si espongono le tecniche di estrazione di queste reti, seguite da una loro visualizzazione grafica.

4.1 Rete delle comunicazioni dirette

Si è definita rete delle comunicazioni dirette l'insieme dei messaggi esplicitamente indirizzati ad altri utenti. Questo insieme comprendeva per FriendFeed tutti i commenti in risposta ad ogni entry, mentre per Twitter alcuni dei messaggi con un esplicito segno “@” nel corpo del messaggio. Non sono stati inclusi i retweet, poiché non indicano una risposta, ma essi verranno comunque interpretati nel processo di propagazione della notizia. L'obiettivo dell'estrazione delle reti di comunicazione è di evidenziare i flussi di comunicazione esistenti tra gli utenti e di identificare quelli che più comunicano tra loro.

Mentre i commenti di FriendFeed erano già disponibili (divisi per entry) in tabelle dopo la pulizia, per le risposte di Twitter è stato necessario estrarli dal file contenente tutti i messaggi con l'hashtag. Si è utilizzato un programma esterno in Java che cercasse alcuni pattern all'interno del file csv iniziale. Sono stati selezionati i messaggi contenenti il seguente pattern:

```
‘ ‘@(\w*)[:\s\r\n]’ ’
```

ma non

```
‘ ‘RT.?.?.?@(\w*)’ ’
```

```
‘ ‘[Vv][íi]a @(\w*)’ ’
```

```
‘ ‘[\\""]@(\w*)’ ’
```

ovvero tutte le frasi che cominciassero con una “@” seguita o meno da una parola e uno spazio, un ritorno a capo o ritorno carrello (escludendo i retweet, esposti nella prossima sezione). Un esempio ne è:

**@Aby005 buenos dias ! yo seguire
mirando un rato mas... #rescatemineros**

Per entrambe le piattaforme si è poi creato un file che contenesse solamente gli id degli utenti, a coppie, che avessero svolto una comunicazione. Questi file sono stati visualizzati su Gephi, realizzando due grafi che rappresentassero le comunicazioni dirette tra i contatti, uno per FriendFeed e uno per Twitter.

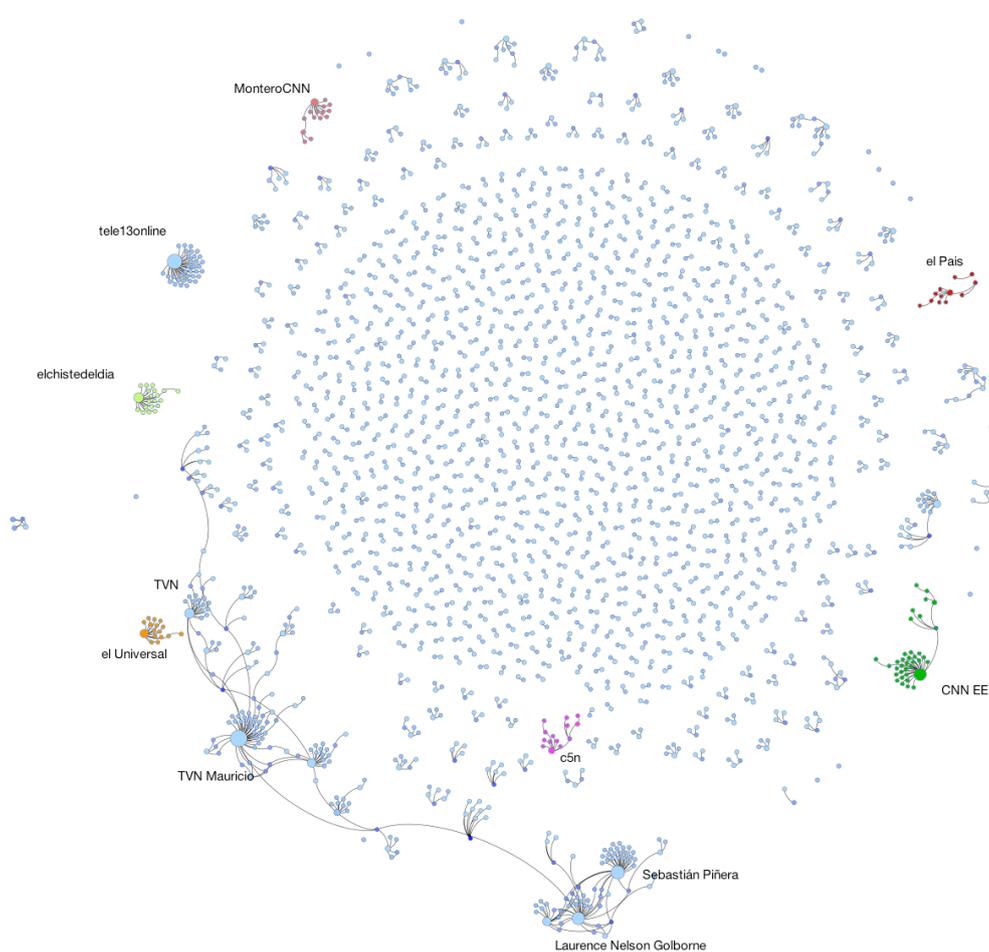


Figura 4.2: Rete delle comunicazioni dirette in Twitter

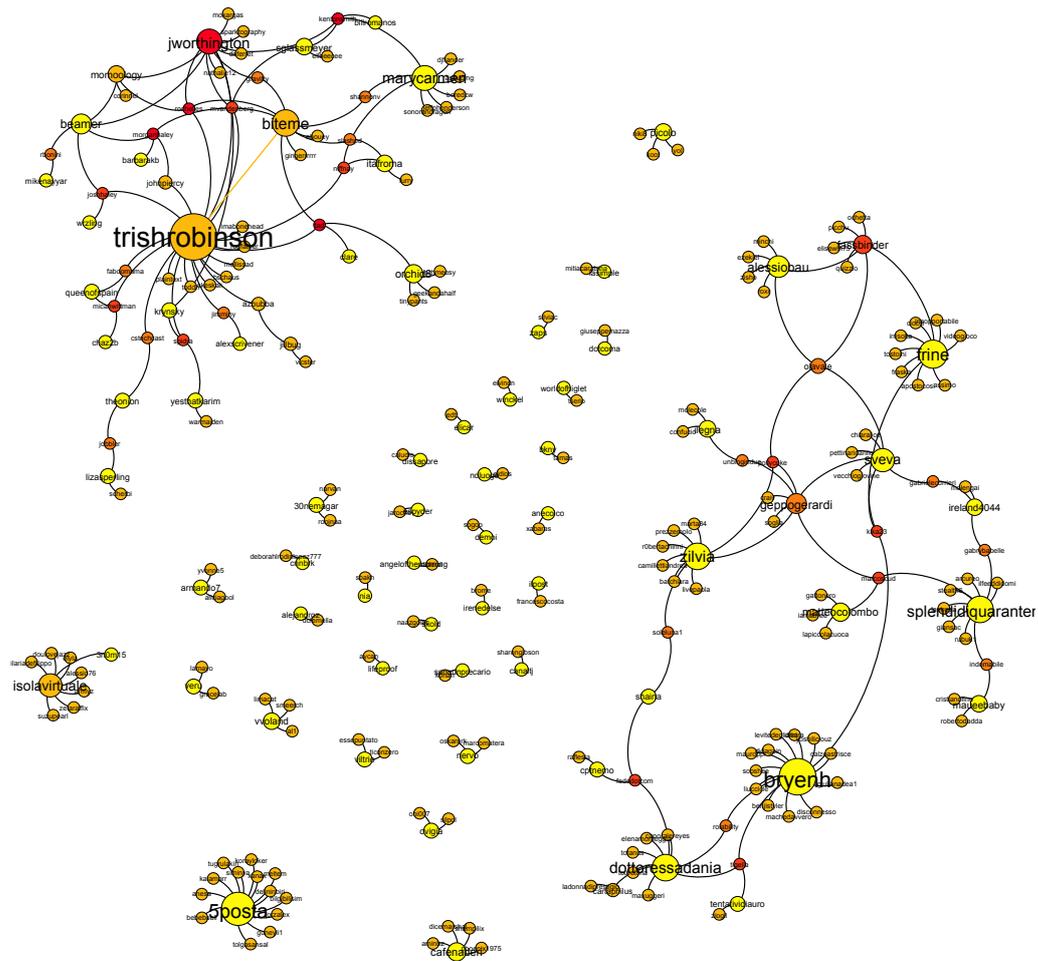


Figura 4.3: Rete delle comunicazioni dirette in FriendFeed

4.2 Rete delle propagazioni

Per analizzare la propagazione della notizia si è creato un programma che a seconda della tipologia di interazione degli utenti con essa ne creasse una rete. Un esempio di interazione è rappresentato dalla figura 4.4.

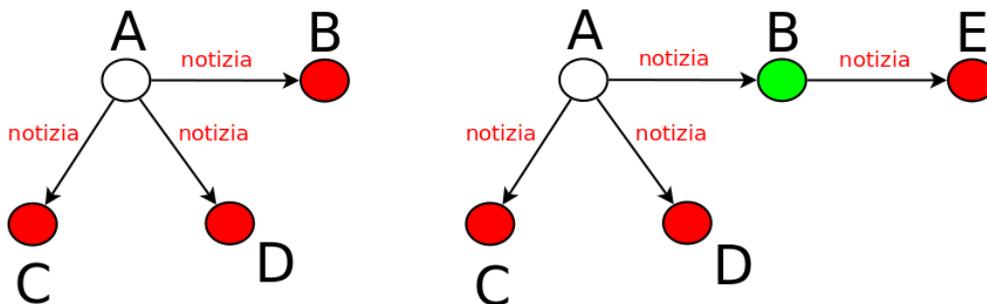


Figura 4.4: Esempio di propagazione

Si supponga che un primo utente (in questo caso A) decida di pubblicare la notizia. I suoi follower B , C e D ne vengono esposti (colorati in rosso nella parte sinistra della figura). Nel caso in cui B dica qualcosa riguardo la notizia dopo averla letta, espone automaticamente i suoi follower diventando un propagatore della notizia (segnato in verde nella parte destra della figura). Esso diventa dunque parte attiva del processo di propagazione, uscendo dallo stato di semplice esposizione (potenziale o effettiva). B è ancora differente da A : non ha inserito per primo la notizia, ma comunque attraverso di lui si sa per certo che sia passata l'informazione (perchè precedentemente esposto).

Sussiste quindi una dinamicità dei comportamenti da tenere in considerazione per seguire i percorsi della notizia. In particolare, per generare la rete si sono dati per assunti alcuni punti, ovvero si è scelto un modello di propagazione. Il modello è stato definito cercando di rappresentare al meglio ogni passaggio dell'informazione tra gli utenti e le loro influenze nei suoi percorsi.

4.2.1 Modello di propagazione

Il modello aiuta a catalogare e figurarsi quali possano essere gli stati degli utenti durante la propagazione della notizia, in modo da costruire con più esattezza il vero percorso dell'informazione.

Si può assumere che gli utenti siano inizialmente inattivi rispetto alla notizia [5]. Essi infatti non hanno ancora compiuto alcuna azione per propagarla nella rete. Nel momento in cui un utente la apprende da un qualche contesto e la inserisce nel sito di microblogging diventa inoculatore ed espone tutti coloro che lo seguono alla notizia stessa. A loro volta, gli utenti esposti possono esporre i propri follower se commentano, apprezzano o retweetano il messaggio del primo utente, propagando così l'informazione (fig. 4.5).

Effettivamente non è possibile conoscere lo stato esatto degli utenti in ogni istante dal momento dell'inserimento della notizia del sito. Per esempio, il fatto che essi siano stati esposti non significa che essi abbiano veramente letto la notizia. È comunque l'approccio logicamente più vicino alla realtà, che verrà rappresentato all'interno del programma di analisi, con strutture appropriate.

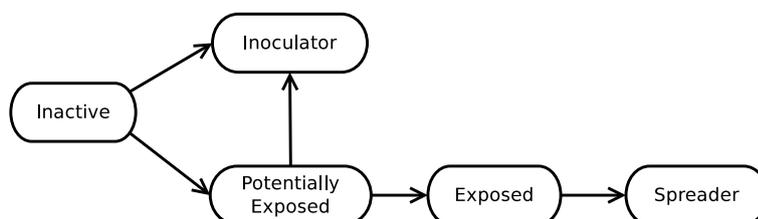


Figura 4.5: Stato degli utenti durante la propagazione di notizie. Fonte: [5]

4.2.2 Rappresentazione degli utenti

Nel programma ideato, si è pensato di creare insiemi e mappe che rappresentassero lo stato degli utenti dal momento di inserimento della notizia nel sito fino alla sua conclusione. Si è scelto di mantenere due insiemi, quello degli utenti attivi e degli inoculatori.

Gli utenti attivi diventano tali nel momento in cui compiono un'azione all'interno del sito (entry, commento, like o retweet). Essi non possono tornare allo stato inattivo, perciò è un insieme che cresce gradualmente fino a che non termina la catena di propagazioni e quindi la notizia non risulta più attiva. Utilizzare un insieme degli inattivi risultava ridondante e inutile poiché inizialmente tutti gli utenti sono inattivi, ma una volta che un utente compie un'azione non potrà più tornare ad esserlo, facendo ormai parte del processo di trasmissione.

Gli inoculatori sono coloro che, inserendo per la prima volta la notizia nel sito, espongono i propri follower ad una nuova informazione.

L'esposizione non è stata presa in considerazione come "potenziale" ma come effettiva: il programma rappresenta come esposti coloro che vengono a sapere della notizia attraverso l'azione di un altro utente nel sito.

Attivi

Gli utenti attivi sono stati rappresentati con un insieme di stringhe, corrispondenti ai loro identificatori (nomi o numeri). Un utente viene inserito nell'insieme quando:

- è il primo a immettere la notizia nel sito attraverso un messaggio, quindi è un inoculatore
- commenta il messaggio di un utente
- apprezza il messaggio di un utente (esprime un "like")
- retweeta il messaggio di un utente
- scrive un nuovo messaggio (entry) una volta esposto da qualcuno all'interno del sito di microblogging

ovvero ogni volta che compie un'azione all'interno del sito.

Inoculatori

Gli inoculatori sono stati rappresentati anch'essi con un insieme di stringhe. Un utente viene inserito nell'insieme quando:

- è il primo a immettere la notizia (non era già stato esposto da nessuno)
- è presente negli esposti ma immette la notizia da un sito esterno (nel caso di FriendFeed, che come si è precisato è un aggregatore di aggiornamenti)

Esposti

Gli esposti sono stati rappresentati con una mappa anziché un insieme:

```
map<string,vector<string> > exposedmap.
```

dove il primo campo rappresenta il nome dell'utente (o il suo ID) e il secondo il vettore contenente tutti coloro che l'hanno esposto.

In questo modo è possibile sapere in ogni momento chi ha esposto ogni utente e in che ordine. L'ordine è necessario per definire in seguito diverse interpretazioni riguardanti chi abbia effettivamente messo tale utente a conoscenza della notizia.

Un utente viene inserito nell'insieme quando:

- segue un utente che compie un'azione sul sito

4.2.3 La creazione della rete

Per potere creare la mappa dei follower è necessario avere una panoramica degli utenti che frequentano il sito, e coloro che essi seguono. A tal fine si sono utilizzati i due file csv contenenti la rete dei contatti di FriendFeed e Twitter. Si è inserito in una mappa:

```
map<string,set<string> > followermap
```

l'insieme dei follower di ogni utente, scandendo i file riga per riga per ricreare la rete. Questa si è successivamente filtrata, recuperando tutti e soli gli utenti che avevano svolto una qualche azione nel sito riguardo la liberazione dei minatori. Così facendo, si è lavorato su un insieme di utenti molto più piccolo, permettendo un'esecuzione più veloce del programma e un risparmio della memoria.

4.2.4 Propagazioni

Un'ulteriore scelta ha riguardato la definizione di propagazione. Una propagazione avviene ogni volta che c'è un percorso certo da un utente ad un altro, ed è rilevabile dalle loro azioni.

Si è considerata propagazione ogni caso in cui:

- un utente commenta, apprezza, o retweeta il messaggio di qualcuno che lo ha esposto: in questo modo si sa per certo che l'utente ha letto il messaggio iniziale
- un utente esposto scrive una nuova entry sul sito: non si sa se abbia effettivamente letto il messaggio che lo ha esposto ma è sicuramente attivo sul sito

Il caso in cui l'utente sia già stato esposto ma introduce la notizia da un sito esterno non viene considerato come propagazione.

A questo punto è chiaro che le esposizioni degli utenti non si basano sul primo livello dei follower, ma dipendono da quante volte viene condivisa la notizia. Esistono vari modi di agire all'interno dei due siti: l'inserimento di un messaggio nuovo (entry), di un commento, l'apprezzamento (like) o il retweet. Il commento e il retweet vengono considerati nello stesso modo, poiché sono entrambi percorsi certi di propagazione dall'utente che ha scritto il messaggio iniziale a chi l'ha commentato (o retweetato) successivamente.

Il programma scritto per creare la rete di propagazioni è strutturato con condizioni successive, in cui vengono controllati i tipi delle azioni svolte dagli

utenti (entry o commenti/like/retweet), esponendo infine i loro follower. Nel caso in cui l'utente risulti già attivo non si compie nessuna azione.

Nei prossimi paragrafi verranno esposte le modalità di approccio all'interno del programma per ogni condizione citata, analizzando i file csv contenenti tutte le entry, i commenti e i like per FriendFeed e tutti i messaggi e i retweet per Twitter, ordinati cronologicamente.

Entry

Se l'utente ha scritto un nuovo messaggio nel sito (una entry) si controlla se è stato esposto da qualcuno.

In caso negativo, esso è un inoculatore e viene inserito nell'insieme corrispondente e in quello degli attivi.

In caso contrario, si controlla la sorgente del messaggio (poiché in FriendFeed possono essere diverse). Se non è FriendFeed, si procede come se fosse un inoculatore (si inserisce quindi l'utente nell'insieme degli attivi e degli inoculatori, cancellandolo dagli esposti). Se invece lo è c'è una propagazione, secondo gli assunti già elencati, quindi si scrive su un file (che verrà analizzato successivamente) il percorso di propagazione (ovvero: l'utente che lo ha esposto per primo e l'utente che ha scritto la entry, con la sua data di inserimento). Si procede poi inserendo l'utente della nuova entry tra gli attivi ed eliminandolo dagli esposti, poiché propagatore.

Commenti (o RT) e Like

Se l'utente ha scritto un commento, espresso un like o retweetato (che è considerato come commento dal punto di vista delle dinamiche di propagazione delle notizie) un messaggio, allora la prima cosa da fare è trovare i suoi follower. Questo è possibile grazie a un metodo del programma, `getFollower(nomeutente)`, che scorre una mappa costituita dai nomi degli utenti e dai corrispondenti insiemi che ne rappresentano i follower. Una volta trovato l'utente che ha eseguito una delle 3 azioni sopracitate, essa restituisce l'insieme di coloro che lo seguono. C'è quindi un percorso di propagazione

tra colui che ha scritto la entry e chi l'ha commentata, e viene scritto il tutto sul file delle propagazioni, seguito dalla data del commento/like/retweet. Si inserisce poi l'utente tra gli attivi e si cancella dagli esposti.

Il recupero dei retweet dal file di Twitter si è effettuato con un programma analogo a quello utilizzato per estrarre le risposte. Questa volta i pattern che si sono cercati sono stati quelli precedentemente eliminati

```
'RT.?.?.?@(\w*)'  
'[Vv][íi]a @(\w*)'  
'[""]@(\w*)'
```

ovvero: la sigla RT seguita da caratteri qualsiasi che apparissero una o più volte e da una risposta, la parola “via” seguita da una risposta o le virgolette davanti a una risposta. Nel primo caso, cioè, si recuperano retweet come:

```
RT @janiceminuet: Super contenta por el  
rescate de los mineros, seguro hoy sera un  
dia lindo! #rescatemineros
```

nel secondo:

```
Fotos oficiales del #rescatemineros con  
licencia CC: http://bit.ly/bvov5A /via  
@jlori
```

mentre nel terzo:

```
"@CNNMex: El minero Pedro Cortez se  
reúne con su compañero y amigo de la  
infancia, Carlos Bugueño  
#rescatemineros // @carlosburgueno6  
funny
```

tutti ampiamente utilizzati.

Esposti e inoculatori nel tempo

Infine, dopo aver controllato ogni tipo di azione eseguibile dall'utente, si espongono i suoi follower, cercandoli con l'opportuno metodo `getFollower(nomeutente)`. Per ognuno di questi, si controlla se già sono attivi; in caso contrario, si aggiorna la loro mappa delle esposizioni, inserendo il nome di chi l'ha esposto nel vettore corrispondente. Come ultima operazione si scrive su un file la data dell'ultima azione avvenuta, il numero degli inoculatori in quell'istante e di seguito la dimensione della mappa degli esposti al momento. L'estrazione della rete delle propagazioni permette infatti di verificare l'andamento con cui gli utenti vengono potenzialmente in contatto con la notizia nel tempo.

Da questi dati si sono ricavati i grafici in figura 4.6 e 4.7 per FriendFeed e Twitter rispettivamente.

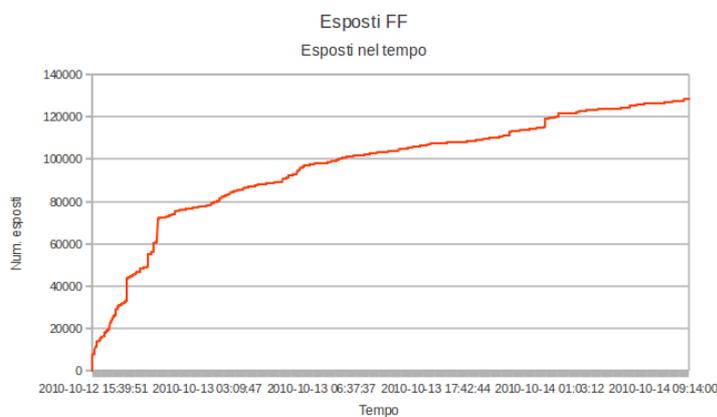


Figura 4.6: Esposti nel tempo in FriendFeed

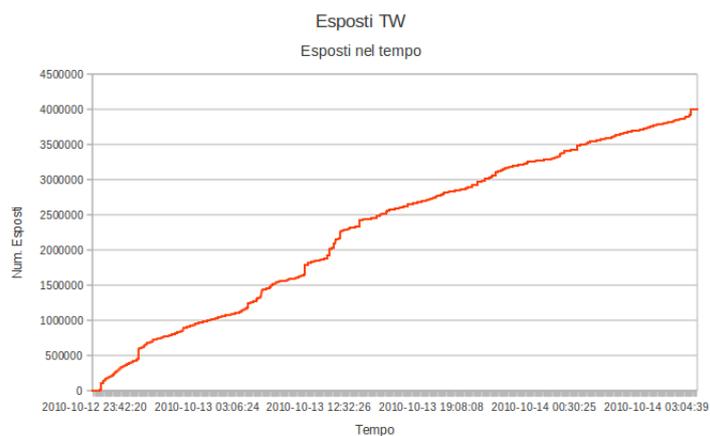


Figura 4.7: Esposti nel tempo in Twitter

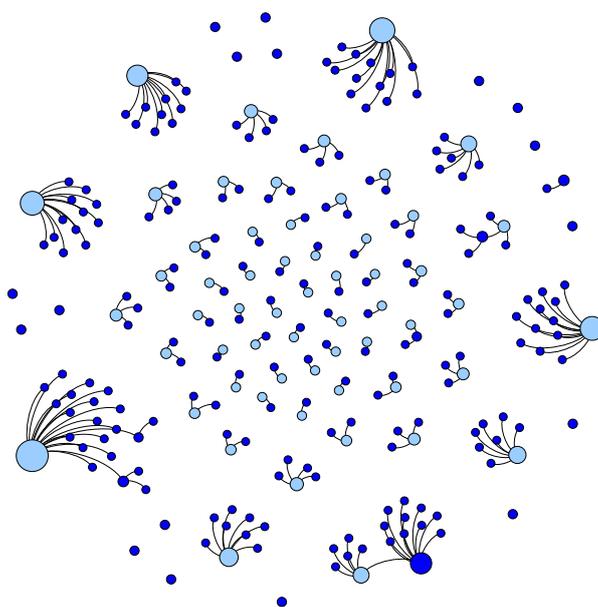


Figura 4.8: Percorsi di propagazione in FriendFeed

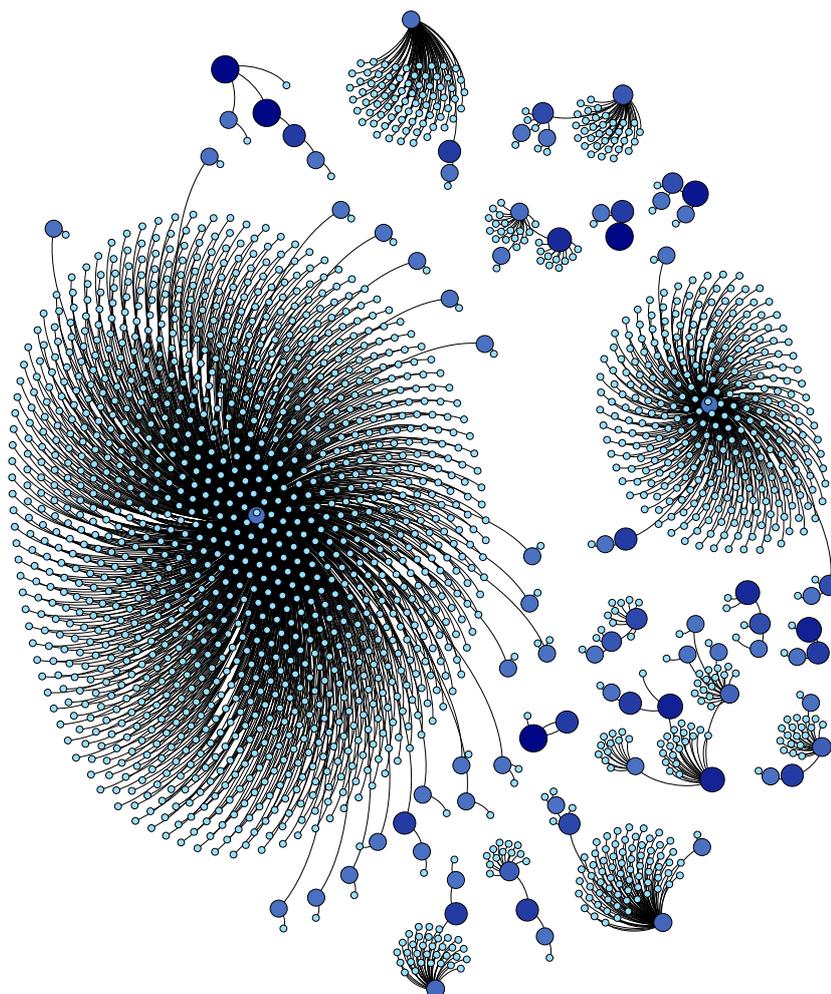


Figura 4.9: Percorsi di propagazione in Twitter

4.3 Rete delle conversazioni dirette

Per poter ricostruire una conversazione su FriendFeed è necessario partire da una delle entry e cercare tutti i commenti relativi. Su Twitter ciò è più complesso, data la struttura del sito, che non permette di visualizzare le risposte nel loro insieme, ma solo nelle pagine personali dei singoli. Solo nel caso di un esplicito segno di risposta (la “@”) è possibile sapere che c’è stata una conversazione tra due individui.

Per studiare le catene di messaggi che possono essere interpretati come conversazioni sono state estratte le 5 più lunghe, per entrambi i siti. Si è utilizzata a questo scopo un’interrogazione SQL

```
select tabellaminatori.eid,  
extract (epoch from date) from tabellaminatori join  
(select eid, count(*) as num from tabellaminatori  
group by eid order by count(*) desc limit 5)  
as topk on topk.eid = tabellaminatori.eid  
order by num desc, eid, date
```

che selezionasse per ogni entry la data delle sue risposte, trasformandole in secondi (dal 1970-01-01 00:00:00-00). Si è contato quindi il numero di commenti/like/risposte(“@”) e si sono ordinate le entry in modo decrescente, prendendo solo le prime 5.

I risultati si sono poi inseriti in un grafico (uno per piattaforma), per visualizzare la durata e l’intensità di ogni conversazione (fig. 4.10 e 4.11). Quella più lunga è rappresentata sulla linea più in basso di ogni grafico, a scalare fino alla più breve delle cinque sulla linea in alto. I punti costituiscono i messaggi che gli utenti si sono inviati, separati dal tempo intercorso tra questi.

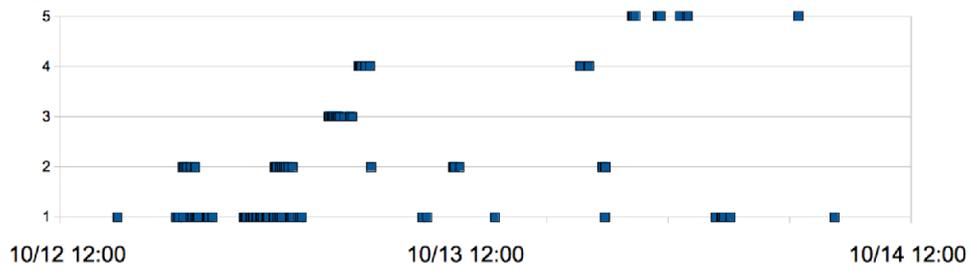


Figura 4.10: 5 conversazioni più lunghe in FriendFeed

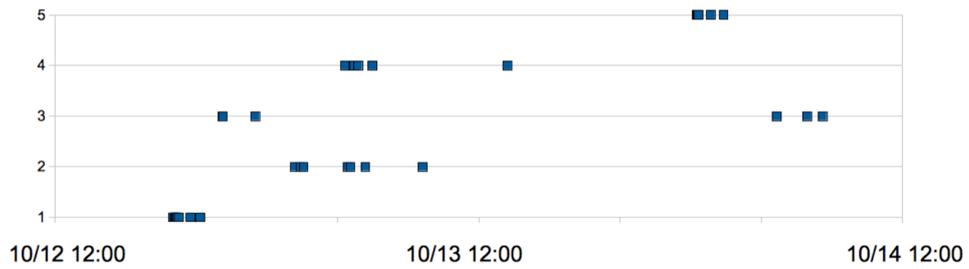


Figura 4.11: 5 conversazioni più lunghe in Twitter

Capitolo 5

Interpretazioni

L'estrazione delle reti necessita di un'interpretazione, per capire a fondo quali percorsi segue nel Web l'informazione. E' stata osservata ogni tipologia di rete, considerando il comportamento degli utenti (l'uso o meno di certe tecniche comunicative o come vengono utilizzati i siti stessi) su entrambe le piattaforme. La presenza di un sociologo in questa fase è stata necessaria per poter interpretare correttamente i dati estratti.

5.1 Rete delle comunicazioni dirette

La visualizzazione della rete delle comunicazioni dirette rivela piccoli gruppi di utenti collegati tra loro tramite archi rappresentanti una comunicazione diretta (i commenti per FriendFeed e le risposte con una “@” per Twitter).

Come mostrato nella figura 4.2 (rete di Twitter), non sono molti i messaggi che si scambiano gli utenti. Nella parte centrale sono indicati quelli che meno hanno interagito con gli altri (un solo arco tra coppie di utenti) mentre nella zona esterna quelli che hanno ricevuto più messaggi di risposta. La figura 4.3 (rete di FriendFeed) è molto simile a quella di Twitter. Anche qui le risposte dirette sono in minor numero rispetto alla quantità di messaggi scritti dagli utenti e le coppie di utenti isolati al centro sono di

quantità ancora più ridotta. Osservando le interazioni si deduce quindi che la tendenza degli utenti a comunicare tra loro è minore rispetto a quella di lasciare messaggi.

Ma chi sono gli utenti che riscuotono maggiore attenzione? Sono emerse caratteristiche differenti dall'analisi dei due siti di microblogging, probabilmente a causa della loro diversa natura. In Twitter gli utenti più commentati sono i canali televisivi e i giornali più importanti del sud America: TVN (la stazione televisiva di stato cilena), El Universal (un importante giornale colombiano), Tele13 (un programma di notizie cileno), C5N (il canale 5 delle notizie argentino), seguiti da el Pais (quotidiano spagnolo) e la CNN spagnola. Allo stesso modo, gli account di giornalisti famosi, come Carlos Montero (CNN spagnola) o Mauricio Bustamante (TVN Mauricio), che probabilmente stavano seguendo dal vivo la notizia, hanno aumentato il livello di comunicazione degli utenti. Questo dimostra come i reporter possano ora interagire con il proprio stesso pubblico, e come gli utenti apprezzino questo tipo di comunicazione senza filtri. Anche il primo ministro cileno Sebastián Piñera ha ricevuto molti messaggi, così come il ministro delle miniere Golborne. Questi, come i giornalisti già citati, rispondono raramente ai messaggi che gli utenti indirizzano loro. Sembra comunque che i navigatori del Web intendano far sentire le proprie opinioni, dimostrando di essere presenti e interessati all'evolversi degli eventi. Al contrario, in FriendFeed non emergono personalità famose, ma conversazioni tra normali utenti.

È interessante anche notare la presenza di una suddivisione geografica della rete. Sul lato destro della figura di FriendFeed si profila la comunità italiana (una delle più grandi), mentre sul sinistro quella americana, entrambe estratte osservando la lingua in cui erano scritti i messaggi degli utenti. Per quanto riguarda Twitter invece si è colorato ogni sottogruppo di utenti appartenente a una certa nazionalità (ricavabile dai gruppi su cui vergevano le discussioni). Il maggiore è sicuramente quello cileno (nella parte in basso a sinistra della figura) che vede la presenza di alcune celebrità del luogo. A seguire, la comunità argentina (colorata di viola e connessa al canale televisivo

C5N), la spagnola (in rosso, connessa al giornale El Pais) e la colombiana (in arancione e connessa al giornale El Universal). Gli altri gruppi sono più difficili da individuare geograficamente, come quello connesso alla CNN spagnola (che è diffusa in tutto il Sud America), colorato in verde.

5.2 Rete delle propagazioni

La visualizzazione della rete delle propagazioni (fig. 4.8 e 4.9) mostra come l'informazione si propaghi da diverse sorgenti. Gli utenti da cui parte la propagazione sono rappresentati dai punti più grossi, e il primo commento/like/retweet in sua direzione crea legami con altri utenti (i punti più piccoli). L'insieme che li collega forma piccole comunità locali. I siti di microblogging non sono dunque ambienti omogenei, ma formati da più sottoinsiemi indipendenti. La conoscenza globale dell'informazione è data quindi da più conoscenze locali.

Sussistono diversi tipi di propagazione in FriendFeed e in Twitter. Il primo sito mostra percorsi di propagazione brevi, mentre le dinamiche di Twitter risultano più complesse. Si hanno in questo caso due tipi di propagazione: in largo e profonda. La prima vede un utente centrale da cui partono molti archi, e rappresenta l'utente che scrive un messaggio che viene retweetato da un alto numero di utenti. La seconda è una propagazione ricorsiva: un utente retweeta il messaggio di qualcuno, a sua volta retweetato da qualcun'altro e così via. Un esempio è la catena di utenti in cima alla figura.

5.3 Rete delle conversazioni dirette

La visualizzazione della rete delle conversazioni dirette mostra la durata delle 5 conversazioni più lunghe in entrambi i siti. I punti blu rappresentano i messaggi, mentre gli spazi indicano il tempo di risposta tra un messaggio e l'altro.

Si osserva che la durata di una conversazione varia in base al sito di microblogging preso in considerazione. Quelle osservate su FriendFeed sono in media più lunghe (cioè sono attive per più tempo) rispetto a quelle di Twitter, probabilmente per la differenza di struttura dei due siti. In particolare, la lunghezza dei messaggi esprimibili da Twitter limita molto lo sviluppo di conversazioni complesse, rendendolo più adatto alla mera propagazione di informazioni. Emerge quindi solo in FriendFeed uno spazio informale in cui poter discutere degli avvenimenti, sentendosi liberi di esprimere la propria opinione, grazie anche al fatto che i commenti sono visualizzati insieme dal sistema.

Ma cosa si sono detti tra loro gli utenti? In FriendFeed il messaggio più commentato ha seguito in diretta l'estrazione, sperando nel lieto fine della vicenda. La seconda entry è già di tipologia differente: l'autore si chiede quale dovrebbe essere la musica che dovrebbe accompagnare il salvataggio dei minatori. La terza, una freddura:

*La sapete quella del minatore cileno che sbaglia a riemergere e si trova al
Marassi in curva serba?*

precede gli altri due messaggi di sostegno ai minatori liberati. Son questi i dialoghi che si sviluppano sul sito, che trovano pochi riscontri nel contesto di Twitter: i messaggi sono più brevi, così come la durata delle conversazioni. Quella più lunga ha comunque un qualche legame con la seconda di FriendFeed: l'autore espone l'idea di Ricardo Arjona (cantante guatemalteco) di dedicare una canzone ai minatori.

5.4 Esposti nel tempo

Il grafico degli esposti nel tempo procede con un andamento apparentemente lineare per Twitter, mentre per FriendFeed mostra un picco iniziale per poi tendere lentamente verso una cifra precisa. Non può essere data una vera e propria interpretazione a questi grafici, poiché il numero di utenti di

FriendFeed è considerevolmente minore rispetto a quelli di Twitter. È comunque da notare che con la seconda azione (6 minuti dopo la prima) nel sito di FriendFeed vengono esposti già un migliaio di follower. Al contrario, Twitter raggiunge il migliaio dopo ben 79 interazioni. Alla terza, FriendFeed raggiunge addirittura i 6000 utenti esposti, cifra che Twitter vedrà solamente dopo 375 azioni sul proprio sito.

Questo indica che inizialmente FriendFeed propaga molto più velocemente la notizia. Il numero degli esposti è però destinato a tendere verso un apparente numero preciso (come intuibile dal grafico), rallentando sempre più velocemente l'esposizione. Il grafico di Twitter non sembra invece raggiungere un punto fisso, superando in poche ore il numero di utenti esposti dell'altro sito e arrivando a contarne un numero consistentemente alto.

Conclusioni

In questo lavoro di tesi si è proposta una metodologia che indicasse le operazioni da compiere per comprendere le modalità di gestione dell'informazione in rete. Un'analisi di questo tipo risulta importante per una lettura più chiara di quello che succede nel Web, strumento utilizzato da persone comuni e da entità di rilievo, come giornalisti e canali televisivi. È un ambiente dove le informazioni viaggiano ogni giorno, esponendo milioni di persone. Questo metodo, applicato un numero cospicuo di volte, ricerca un possibile determinismo negli sviluppi delle interazioni tra utenti riguardo una specifica informazione.

Con questa metodologia si è quindi svolto un primo studio di caso, osservando come la notizia dei minatori cileni liberati ad Agosto 2010 si sia espansa nella rete.

Il modello consta di semplici passi, otto in tutto: la scelta della notizia da studiare, la sua ricerca in uno o più siti di microblogging e la relativa acquisizione dei dati. Dopo un'operazione di pulizia, si estraggono le reti che si vogliono osservare, visualizzandole e ricavandone opportune interpretazioni. Si giungerà infine a considerazioni finali derivanti dall'analisi delle reti estratte. Ciò che si è concluso a seguito delle interpretazioni date alle reti è il seguente.

Per quanto riguarda la rete delle comunicazioni dirette, le attività di comunicazione sono esigue rispetto al numero totale di messaggi scritti. Gli utenti più commentati sono celebrità nel caso di Twitter e persone normali in FriendFeed. Nel primo caso è emerso un tentativo da parte degli utenti

di esprimere pubblicamente le proprie opinioni a figure di rilievo, che però raramente interagiscono con l'utente comune. Nel secondo caso emerge un ambiente più favorevole a conversazioni informali, dove normali utenti scambiano le proprie opinioni senza passare attraverso personalità famose. In entrambe le situazioni è comunque evidente una suddivisione geografica dei gruppi di utenti che comunicano tra loro, connessi in base alla nazionalità.

La rete delle propagazioni denota la presenza di diverse sorgenti da cui l'informazione si propaga contemporaneamente. Queste comunità locali hanno quindi il proprio bagaglio di conoscenza che trasmettono alle comunità più vicine, costruendo una conoscenza globale. Twitter manifesta una propagazione più complessa di quella di FriendFeed, poiché l'utente non solo può propagare verso molti altri utenti (tramite l'uso del RT), ma si possono creare catene di propagazione in profondità quando un utente retweeta un altro, a sua volta retweetato.

La rete delle conversazioni dirette mostra infine come a diversi servizi corrispondano differenti pratiche comunicative. Mentre da Twitter si evince che la difficoltà di instaurare conversazioni nel sito porta gli utenti a pubblicare notizie più che a commentarle con gli altri, FriendFeed mostra un approccio più conversazionale.

I risultati che si sono ottenuti con questo studio di caso non sono ovvi né intuibili a priori. Si è scoperto che gli utenti sono poco inclini alle discussioni (nonostante quello che si pensi riguardo ai siti di microblogging), che l'informazione parte da piccole comunità locali (anziché da un'unica sorgente) e che le conversazioni identificano vere e proprie zone geografiche. La differenza di piattaforma influenza inoltre le propagazioni e la durata delle conversazioni, così come la tipologia. Per questo motivo l'applicazione del modello su diversi siti di microblogging per notizie diverse è utile ai fini di confermare o smentire i risultati ottenuti e giungere a una chiave di lettura.

L'interpretazione delle reti ha comunque permesso una conoscenza più approfondita delle dinamiche della gestione delle informazioni in due delle piattaforme più utilizzate. Grazie al modello creato si è potuto ricostruire il

percorso della notizia in pochi passi, ma le analisi che si possono sviluppare sono molteplici. Al singolo è affidata la scelta degli indizi che l'informazione lascia nel Web, con la speranza che il modello possa costituire un utile contributo per la loro analisi.

Bibliografia

Riferimenti cartacei

- [1] The Economist. A world of connections. A special report on social networking, 2010.
- [2] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*. Cambridge University Press, 1994.
- [3] Matteo Magnani, Luca Rossi, and Barbara Iadarola. Global media events and microblogging sites: a case study. Submitted to the Journal of Computer-Mediated Communication, 2011.
- [4] Matteo Magnani, Danilo Montesi, and Luca Rossi. Information propagation analysis in a social network site. In *International conference on Advances in Social Network Analysis and Mining*, pages 296–300. IEEE Computer Society, Los Alamitos, 2010. isbn: 978-0-7695-4138-9/10.
- [5] Matteo Magnani, Luca Rossi, and Danilo Montesi. Friendfeed Breaking News: death of a public figure. In *Second IEEE International Conference on Social Computing*, pages 528–533, LOS ALAMITOS – USA, 2010. IEEE computer Society.

Riferimenti on line

- [6] SIGSNA.
<http://larica.uniurb.it/sigsna/>. Retrieved: [15-01-2011].
- [7] Il giorno di Static Kill. In mare 5 milioni di barili.
http://www.repubblica.it/ambiente/2010/08/03/news/marea_nera_il_giorno_di_static_kill-6032657/?ref=HREC1-5.
Retrieved: [15-01-2011].
- [8] Addio al Picconatore, è morto Cossiga.
http://www.corriere.it/politica/10_agosto_17/cossiga-morto_18d90b48-a9f3-11df-8b1f-00144f02aabe.shtml.
Retrieved: [15-01-2011].
- [9] Hollywood piange la morte di Tony Curtis eterno seduttore di A qualcuno piace caldo.
http://www.repubblica.it/spettacoli-e-cultura/2010/09/30/news/morto_curtis-7575892/. Retrieved: [15-01-2011].
- [10] Timeline: Trapped Chilean miners.
<http://www.reuters.com/article/2010/08/23/us-chile-miner-accident-timeline-idUSTRE67M0G820100823>.
Retrieved: [16-01-2011].
- [11] Vivi i 33 minatori intrappolati da 17 giorni. Per tirarli fuori ci vorranno quattro mesi.
http://www.repubblica.it/esteri/2010/08/22/news/minatori_cile-6442210/. Retrieved: [16-01-2011].
- [12] Cile: minatori, si blocca il piano B.
http://www.ansa.it/web/notizie/rubriche/topnews/2010/09/12/visualizza_new.html_1783509024.html. Retrieved: [16-01-2011].

- [13] Trivella raggiunge i minatori sottoterra da 42 giorni.
http://www.repubblica.it/esteri/2010/09/17/news/cile-minatori_raggiunti-7177665/?ref=HREC1-7.
Retrieved: [16-01-2011].
- [14] Cile, la liberazione dei minatori.
<http://blog.panorama.it/foto/2010/10/12/cile-la-liberazione-dei-minatori/>. Retrieved: [16-01-2011].
- [15] What is microblogging and where do i get started.
<http://onlinebusiness.about.com/b/2008/07/10/what-is-microblogging-and-where-do-i-get-started.htm>.
Retrieved: [17-01-2011].
- [16] FriendFeed.
<http://friendfeed.com/>. Retrieved: [17-01-2011].
- [17] Twitter.
<http://twitter.com/>. Retrieved: [17-01-2011].

Ringraziamenti

Senza l'aiuto e il sostegno di alcune persone non avrei mai raggiunto questo risultato. Ringrazio in primis i miei genitori, che hanno sempre creduto nelle mie capacità e non mi hanno mai fatto mancare nulla; i professori che con il loro supporto mi hanno aiutato durante il percorso di studi: Fabio Vitali, Andrea Asperti e il mio relatore Danilo Montesi. Ringrazio inoltre Luca Rossi per la disponibilità e l'aiuto offerto in questo lavoro di tesi, e i miei fedelissimi compagni di progetti di cui conserverò un ricordo indelebile (nel vero senso della parola). Al Dott. Matteo Magnani vanno i miei ringraziamenti più sentiti, per tutto l'appoggio, la pazienza e la fiducia riposta in me.