

LE CONSEGUENZE DI UN TWITT¹

LINDA PAGLI

Dipartimento di Informatica, Università di Pisa

Tutti noi usiamo allegramente i computer senza preoccuparci troppo di quali potrebbero essere le conseguenze. Sia che si acquisti on-line un biglietto aereo o ferroviario, o si scelga l'ultimo romanzo del nostro autore preferito, che si cerchino occasioni di vestiti firmati o si voglia prenotare un albergo di lusso a prezzo super scontato, dobbiamo sempre seguire dei passi obbligati. Ci dobbiamo registrare nel sito di turno, fornendo tutti i nostri dati personali, dobbiamo dare l'autorizzazione al trattamento dei dati mettendo una crocetta in un quadratino piccolo e poco visibile (cosa significhi questa autorizzazione non è chiaro per niente, quello che è chiaro che se non si autorizza non si può andare avanti nel processo di registrazione). In caso di pagamento dobbiamo dare anche i dati della nostra carta di credito o prepagata, per cui dobbiamo fare altre registrazioni e dare autorizzazioni, operazioni anche coperte da parole chiave scelte da noi, spesso così semplici e immediate come la nostra data di nascita, o ancor peggio 123456, che anche un hacker alle prime armi può scoprire in un attimo.

Ma c'è di più: alle enormi memorie di Google, affidiamo tutta la nostra corrispondenza elettronica se usiamo il servizio di posta elettronica Gmail, così comodo per l'utente che non ha più limiti di spazio nella sua casella postale. Gmail si ricorda non solo di tutti gli scambi tra utenti, ma anche dei contenuti delle mail! Con le reti sociali come Facebook e Twitter depositiamo nella rete la nostra storia personale, esprimiamo opinioni, manifestiamo emozioni, punti di vista e inclinazioni politiche, vizi e virtù personali. Gli accessi che facciamo ai siti internet sono anch'essi tutti registrati e indicano ulteriori curiosità e interessi degli individui. Insomma una valanga di informazioni, detta in gergo "Big Data" che opportunamente aggregate, elaborate, filtrate, incrociate e visualizzate per mezzo di algoritmi sofisticati e in breve tempo può rivelare informazioni implicite, relazioni inaspettate e tante altre informazioni interessanti. L'informazione ricavata viene usata per scopi di vario tipo, alcuni già chiari, come il marketing, altri meno e della cui portata e conseguenze non siamo del tutto consapevoli. È notizia recentissima il progetto già attuato dell'amministrazione Obama per l'analisi dei "Big Data" a scopi antiterroristici. L'operazione tenuta segreta e svelata da un giornalista inglese del *Guardian*, è stata recepita dall'americano medio, molto sensibile alle proprie libertà personali, come una sorta di schedatura universale per il controllo capillare dell'individuo. La *Casa Bianca* si difende dicendo che si stanno studiando le reti di relazioni tra

¹ Lezione tenuta il 14 novembre 2012 presso l'ISIS Follonica, (GR) e collegata al seminario tenuto a Pisa, presso La Limonaia di Palazzo Ruschi, il 6 novembre 2013.

individui alla ricerca di quelle sospette, non il contenuto delle singole e-mail, e che se si vuole combattere il terrorismo forse qualche sacrificio a livello di privacy va anche fatto; ciò nonostante l'operazione ha avuto un forte impatto emotivo e provocato una sensazione di Grande Fratello che osserva tutto. Mi domando perché tanto stupore, quando si affida tutta la nostra privacy alla rete.

Bisogna inoltre considerare che gli interessi economici in ballo sono enormi e per dirla con Andrew Keen, scrittore esperto di problemi legati al web: "I dati personali sono il nuovo petrolio".

A livello personale è importante acquisire consapevolezza nell'uso della rete, sapere che ogni nostra azione sul web, anche la più insignificante viene registrata e potrà essere ricordata negli anni a venire, che tutte le informazioni che affidiamo alla rete direttamente o indirettamente potranno essere utilizzate. Significa essere coscienti che quando si fa un accesso a un sito proibito, c'è una polizia altamente specializzata, in Italia la polizia postale, che lo può vedere e perseguire. Anche l'operazione più semplice come un *twitt* può avere conseguenze, bisogna tener presente che in contatto con la rete non ci troviamo in una stanza isolata, ma in una pubblica piazza colma di persone.

Capire o anche solo intuire il funzionamento di un motore di ricerca o di un servizio web può aiutare. Analizziamo quindi come funziona *Twitter*, che è la rete sociale che ha avuto più successo di recente, ma il discorso potrebbe essere esteso a molti altri servizi di rete.



Figure 1. *Twittering Machine*, Paul Klee, 1922 – Museum of Modern Art (MoMA) New York

Twitter permette lo scambio di messaggi brevi, di dimensione standard che possono essere inviati anche da un semplice cellulare in forma di sms. Possiede inoltre un'architettura completamente aperta (*open source*), cioè non proprietaria in accordo alla filosofia di partecipazione e condivisione di tanti prodotti software (come Unix e Linux per esempio) tanto cara alla rete. Questo permette a utenti che siano anche programmatori esperti di accedere e migliorare i programmi di base e poter anche sviluppare nuove applicazioni da usare insieme a Twitter. A ogni utente che si registra viene assegnata una pagina personale, che può essere osservata da seguaci (*follower*) cioè altri utenti interessati a lui, che a sua volta può decidere di seguire le pagine di altri utenti (*following*). Nella pagina vengono mostrati i twitt in arrivo più recenti, sia quelli degli utenti che si seguono sia quelli che arrivano a tutti.

Ogni utente può inviare i twitt ai suoi seguaci oppure renderli pubblici, o ancora rimbalzare un twitt di rilievo (*re-twitt*). Per la sua semplicità e immediatezza il mezzo si è imposto molto velocemente come strumento utilissimo alla diffusione rapida di notizie. È stato così efficace che nel caso di alcuni eventi improvvisi come il disastroso terremoto in Giappone del 2011, ha permesso di coordinare i soccorsi e salvare persone. Le notizie portate da twitter hanno definitivamente stabilito un nuovo modo di fare informazione, che è diventato corale, composto da testimonianze dirette e immediate quindi cariche di emozioni, un vero e proprio termometro degli eventi. Il suo successo è decretato dai numeri della sua diffusione: 200 milioni di utenti di cui 2.5 solo in Italia per una media di 350 milioni di twitt al giorno. Dal 2009 funziona anche come motore di ricerca: le parole di interesse contenute nei messaggi vengono precedute da un simbolo speciale (#) per poter essere cercate in tutti i twitt. Si formano così gruppi d'interesse su argomenti specifici e le relazioni di vicinanza e gli scambi tra utenti privilegiano interessi e passioni comuni, piuttosto che l'amicizia, la parentela o vissuto comune, come in altre reti tipo *Facebook*. Twitter si discosta dalle altre reti perché riunisce persone con uguali interessi, è meno privata e riflette più il presente che la storia personale.

Twitter è diventato sinonimo della primavera araba perché attraverso i tweet dei manifestanti è stato possibile vivere con loro i cortei e le proteste, ascoltare da vicino le emozioni di orrore, di rivolta e il coraggio di tanti giovani nel rivendicare i propri diritti.

Con Twitter comunicano Obama e Medvedev, e così hanno mandato in pensione la famosa linea rossa, linea telefonica riservata i capi di stato americano e russo, dove si scambiavano messaggi crittografati durante la guerra fredda. Un twitt è arrivato da un astronauta della Nasa dallo spazio e la foto di un'auto in fiamme durante la corsa di Daytona è stata inviata in un twitt dal pilota dell'auto che la seguiva. Curiosità che indicano impieghi sempre nuovi e inaspettati della rete di informazione. La *sentiment analysis*², è il procedimento con cui si analizzano i twitter scambiati sulla rete per capire

2 <http://www.ccs.neu.edu/home/amislove/twittermood/>

le sensazioni di campioni di persone; si è studiato per esempio come cambia dell'umore degli americani nelle diverse ore del giorno per scoprire che l'umore tende a peggiorare nelle ultime ore di lavoro e tra mezzanotte e le tre probabilmente a causa della solitudine e dell'insonnia di chi manda messaggi. Non sono grandi scoperte ovviamente, ma pur sempre interessanti perché estrapolate direttamente dai messaggi.

Twitter è dunque uno strumento migliorato dalla collaborazione di tanti, che permette di fare informazione dal basso, di grande aiuto in situazioni di emergenza, che funziona anche come motore di ricerca e tutto questo ci viene fornito gratuitamente. Ma come fa Twitter a guadagnare se il servizio è gratuito? La filosofia è un po' quella di *Google*, ci sono i messaggi sponsorizzati che vengono pubblicati a pagamento. La natura dei messaggi pubblicitari è indicata chiaramente e la loro veste non è invasiva. Ci sono stati grandi investitori che ci hanno creduto molto che ci hanno impegnato i loro capitali. Come faranno a guadagnarci?

Avere i dati a disposizione significa capire l'orientamento del pubblico e seguire i movimenti degli individui, una sorta di enorme intercettazione telefonica globale. Avere un'informazione prima degli altri può essere molto utile in vari campi. Immaginate un politico che intercetta gusti e tendenze dei giovani e se ne vuol fare portavoce, potrà anche usare lo stesso linguaggio. S'investirà più o meno su un prodotto di qualsiasi genere, un cantante, un film, un oggetto se si viene a sapere che piace in modo particolare. Queste sono le applicazioni più immediate che vengono subito in mente, prevedere le estreme conseguenze di questo meccanismo è ancora qualcosa che sfugge.

Un altro aspetto da considerare è l'analisi velocissima di insiemi di dati di enormi dimensioni. Che si possa fare lo sappiamo già attraverso l'esperienza diretta con i motori di ricerca. Facciamo una richiesta specificando una o più parole chiave e in un attimo abbiamo la risposta, che il più delle volte è più che soddisfacente. Com'è possibile? Per spiegarlo nel dettaglio dovremmo scrivere un trattato d'informatica. Contentiamoci di fornire alcune intuizioni sui complessi meccanismi che regolano il funzionamento dei motori di ricerca e di altri servizi web, rimandando i lettori interessati a testi più approfonditi.

Una delle chiavi del successo è un'enorme disponibilità di potenza di calcolo, unita a un uso massiccio del calcolo parallelo. Con calcolo parallelo (o distribuito) si indica un approccio alla risoluzione dei problemi ove insiemi di computer eterogenei, invece che un singolo computer, si associano per risolvere problemi su dati di grandi dimensioni e si accordano per stabilire strategie e modalità di comunicazione. La capacità intrinseca di risolvere problemi non cambia se a lavorare è un computer solitario o un gruppo di essi, quello che cambia radicalmente è la dimensione dei problemi che possono essere risolti. Se poi i problemi considerati godono di particolari caratteristiche, si prestano molto bene a essere risolti parallelamente, al che corrisponde un risparmio notevole in tempo di elaborazione. Ci sono invece dei problemi per cui la soluzione parallela con più computer non aiuta e a volte è addirittura controproducente.

La soluzione collettiva può essere dunque una mossa vincente o meno, a seconda delle situazioni e tenendo che nel calcolo parallelo sorgono problemi nuovi, a volte

complicati. È per esempio più difficile definire un algoritmo, e anche solo capirne l'esecuzione può risultare molto complicato, come seguire la conversazione animata di tante persone che parlano contemporaneamente.

Se i computer devono accedere a risorse comuni in maniera esclusiva, per esempio una porzione di memoria su cui devono poter scrivere, si dovranno stabilire delle precedenze e se queste si chiudono ciclicamente si possono presentare problemi di *deadlock* che fanno andare il sistema in un situazione di stallo. Ovviamente queste situazioni devono essere previste a priori ma non è così facile prevederle se le regole di priorità sono molte. Altre difficoltà possono insorgere perché i computer generalmente hanno solo una conoscenza parziale della rete, l'andamento del calcolo dipende dai ritardi sulla rete e va previsto per ogni eventualità e un algoritmo in genere corrisponde a numerose esecuzioni distinte.

Una soluzione collettiva di un problema è normalmente più veloce di una soluzione elaborata dal singolo computer quindi può essere necessaria in caso debba essere terminata entro una certa scadenza temporale; la soluzione collettiva può essere conveniente se si può facilmente decomporre il lavoro da fare tra computer diversi e ricombinare altrettanto facilmente le soluzioni ottenute, può essere invece inutile o dannosa nel caso che il tempo di coordinamento tra computer e della raccolta dei risultati sia maggiore del guadagno dovuto al parallelismo. Si possono trovare tantissimi problemi "veri" che si adattano a una qualsiasi delle situazioni precedenti.

L'uso del parallelismo si è rivelato vincente proprio per i motori di ricerca e per servizi web che analizzano dati. I motori di ricerca e anche Twitter fanno un uso aggressivo del parallelismo. Quando si articola una domanda, la risposta appare sul video in modo praticamente istantaneo. In questo brevissimo intervallo di tempo le parole chiave sono state cercate sulle pagine se non di tutto almeno di una parte cospicua del web. Le nostre parole chiave contemporaneamente a tutte le altre viaggiano nella rete a milioni al secondo.

Varie copie dell'informazione contenuta nel web è registrata nelle memorie private dei motori che costruiscono anche i dizionari di parole chiave. Le memorie private, dette *cluster*, sono costituite da grandi insiemi di computer, normalmente semplici PC, e dislocate in varie parti del mondo. Una richiesta viene inviata al cluster più vicino oppure a quello meno affollato. Anche se i dati complessivi hanno dimensioni enormi essi possono essere decomposti in piccoli pezzi di ugual dimensione e la richiesta distribuita in tantissime copie a tutti quanti. La ricerca è un algoritmo da ripetere identico su tutti i pezzi contemporaneamente. Le ricerche procedono indipendenti una dall'altra e non devono comunicare dati parziali. Il risultato finale è semplicemente una concatenazione dei risultati ottenuti da ciascun pezzo. Siamo nella situazione ideale per il calcolo parallelo, i benefici sono ottimali e i risultati si vedono!

Esiste dunque un'enorme mole di dati accumulati dall'interazione tra la rete e i suoi utenti. L'analisi di questi dati, apparentemente impossibile viste le dimensioni, può essere realizzata in parte in modo relativamente semplice e veloce con l'uso del paralleli-

smo. Il web ci fornisce tanti bei servizi apparentemente gratuiti, ma questi ci prelevano tutti i nostri dati personali. I nostri dati personali hanno sicuramente un grande valore sul mercato, ma forniscono anche strumenti molto potenti di controllo sull'individuo. Chiediamoci se è uno scambio equo.

Vogliamo concludere però con una buona notizia: Twitter si distingue dalle precedenti reti sociali e anche dai motori di ricerca perché mette a disposizione, oltre ai suoi programmi, anche parte dei suoi dati. Una piccola percentuale (l'1%), ma sempre una grandissima quantità di twitt è resa pubblica a chiunque la voglia "ascoltare"; basta avere un account su Twitter, accedere a un determinato indirizzo web³ per scaricare un flusso continuo di twitt. I twitt portano con sé informazioni preziose come per esempio il numero di seguaci di chi l'ha emesso, utili per stabilire la rilevanza del twitt stesso. Si può utilizzare il programma *open source* Map-Reduce⁴ che permette di gestire problemi radicalmente paralleli come quelli visti sopra. Tutto il lavoro di smistamento dei dati ai vari computer, la replicazione delle domande, l'organizzazione e il bilanciamento del carico, è realizzato automaticamente dal programma stesso senza che chi lo usa si debba preoccupare minimamente di gestire il parallelismo. Scrivere un programma che usa questa tecnica per analizzare dati è relativamente semplice e prescinde dalla complicazione di dover esplicitamente dirigere le operazioni parallele. Con un po' di disponibilità di rete dedicata, che si può tranquillamente affittare a poco prezzo, possiamo analizzare i twitt come meglio crediamo, per fare affari, ricerca o politica, per scoprire nessi o prevedere eventi e magari inventarsi un lavoro.

Intanto, in via precauzionale, cerchiamo di lasciare nel web il minor numero di tracce possibile...⁵

3 <https://stream.twitter.com/1/statuses/sample.json>

4 Rilasciata da Google.

5 Vedi anche L. Pagli: "Tu twitti, loro guadagnano" *X la tangente*, n.36 dicembre 2012.