

Se i dati diventano big

Viola Bachini

Secondo Fosca Giannotti, dirigente di ricerca al Cnr, “I Big Data sono il nuovo microscopio che rende misurabile la società”. E come la scoperta di ogni nuovo microscopio o telescopio nel passato, i Big Data stanno spingendo verso una nuova scienza dei dati: il social mining, in grado di misurare e, in prospettiva, prevedere crisi economiche, epidemie e pandemie, diffusione di opinioni, distribuzione delle risorse economiche o energetiche, bisogni di mobilità

Qualcuno li usa per prevedere il traffico o il meteo. I sociologi li raccolgono con interesse perché possono rivelare i sentimenti della collettività rispetto a un evento: come stanno reagendo, per esempio, i parigini all’attentato dello scorso autunno? Queste opportunità e molte altre arrivano dai Big Data, raccolte di numeri così estese in volume, velocità e varietà da richiedere tecnologie e metodi di analisi specifici per elaborarli. Si tratta in gran parte dei dati che lasciamo in rete, per esempio quando postiamo una foto sui social o mandiamo un messaggio a un amico. Ci troviamo in un momento storico senza precedenti per la ricerca, perché gli scienziati non avevano mai avuto un numero così grande di dati a disposizione. Inoltre, il vantaggio enorme rispetto al passato è che tutti questi dati sono disponibili in formato digitale e quindi facilmente utilizzabili. La disciplina è relativamente nuova. I primi tentativi di estrarre senso dai dati risalgono agli anni ‘80, ma per gli esperti del settore l’inizio dell’era dei Big Data coincide con la pubblicazione nel 2010 del Dremel, il paper scientifico firmato Google che spiega come compiere ricerche su una mole enorme di dati. Si dice che chiunque lavori nell’ambito lo abbia letto almeno una volta. Da quei primi tentativi di analisi i dati digitali sono aumentati vertiginosamente: secondo una stima dell’Università della California, nel 2000 solo un quarto di tutta l’informazione presente nel mondo era registrato su supporto digitale, mentre solo 13 anni più tardi si era arrivati addirittura al 98%. Questo perché ormai tutte le dimensioni della nostra vita sono in qualche modo tracciate. Sui social parliamo di quello che ci piace, tramite il gps dello smartphone lasciamo traccia del

nostro passaggio e le recensioni che lasciamo agli alberghi possono dire molto sul nostro stile di vita. Ogni volta che ci registriamo su un sito per un servizio diamo il nostro consenso all’utilizzo dei dati che raccontano come e quando abbiamo utilizzato quel servizio. Quando compriamo un oggetto online i dati finiscono in un grande database insieme a quelli di tutti gli altri clienti, permettendo così di studiare le nostre preferenze per il prossimo acquisto.

Infine, diciamo molto anche sulle persone che frequentiamo. Sono i contatti telefonici, le mail e le reti di amicizie online a raccontare la nostra vita sociale. Difficile, quasi impossibile, sfuggire a questo scenario. Sono sempre di più gli italiani che possiedono uno smartphone, 7 su 10 secondo l’Istat. Secondo la stessa indagine, pur di riuscire ad acquistare l’ultimo modello siamo disposti a eliminare altre voci di spesa, come il parrucchiere. Così come non riusciremmo a rinunciare al commercio online o altri preziosi servizi delle rete.

LA PRIVACY E IL MONOPOLIO

In un contesto del genere i rischi sulla violazione della privacy e sfruttamento dei dati per fini non etici sono evidenti. Oggi gli aspetti su cui sono chiamati a lavorare giuristi ed esperti di etica ruotano intorno al concetto di privacy, trasparenza e contrasto del monopolio. Quando apriamo una *app* dello smartphone o il browser del computer di casa siamo in buona compagnia: sono circa 2,5 milioni le persone nel mondo che accedono a internet, producendo informazioni per circa 2,5 quintilioni di byte. E ogni volta che lo facciamo, in modo più o meno consapevole, lascia-



mo indizi che dicono qualcosa di noi, spesso anche su aspetti privati. Per questo una delle prime questioni da affrontare riguarda l'informazione degli utenti, che dovrebbero sapere chi utilizza i dati e come. Il fondatore di Facebook, Mark Zuckerberg, nel 2010 disse che la privacy era morta. Quello che però i californiani con la felpa spesso si dimenticano di dire è che c'è una grande asimmetria dell'informazione: ci sono molti dati nelle mani di poche persone. Chi controlla i social network sa tutto o quasi di noi ma raramente è disposto a mettere queste informazioni a disposizione di chi fa ricerca per migliorare la qualità della nostra vita.

Eppure, assicurano gli esperti, aprire i dati porterebbe grandi vantaggi, sia economici che da un punto di vista della partecipazione attiva dei cittadini a problemi che li riguardano da vicino, come la pianificazione di bonifiche ambientali o la messa in sicurezza sismica delle scuole. Oggi esistono profonde



Chi controlla l'informazione su tutto su di noi ma è restio a condividerne i contenuti

differenze tra la legislazione europea e quella quella statunitense. La prima è molto più rigida, a tal punto da costringere qualche mese fa la Corte di Giustizia dell'Unione Europea a invalidare l'accordo "Safe Harbor", che regolamentava il passaggio di dati dai server europei verso i centri di elaborazione americani. Tradizionalmente in Europa la privacy non è ritenuta un diritto negoziabile, mentre negli Stati Uniti si tende di più ad accettare di scambiare informazioni riguardanti la sfera privata per ottenere un servizio gratuito. Un po' quello che succede quando utilizziamo l'account di posta elettronica senza pagare un centesimo, ma di fatto permettendo a chi lo gestisce di capire che lavoro facciamo e chi frequentiamo.

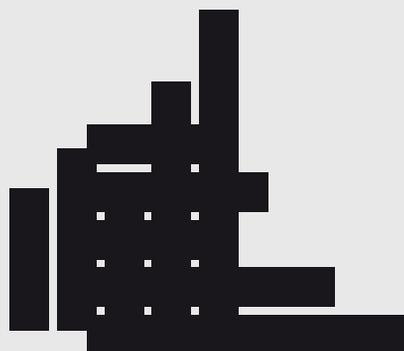
L'ETICA DEL DATA ANALYST

Ogni minuto vengono inviate 200 milioni di e-mail,

caricate 60 ore di video ed effettuate più di 9 milioni di telefonate. Dino Pedreschi, professore di informatica all'Università di Pisa, le chiama "briciole digitali". Per il ricercatore la chiave di una gestione etica dei dati risiede anche nell'identità del Pollicino che raccoglie queste briciole. «Finché i dati rimarranno nelle mani di pochi saranno utilizzati per fini commerciali o di marketing».

Costruire un'alternativa che sfrutti il potenziale racchiuso nei dati per fini socialmente utile è possibile, e a Pisa è già realtà. Risale a pochi mesi fa l'inaugurazione del *So Big Data Lab*, la prima infrastruttura di ricerca tutta dedicata all'analisi dei Big Data. Il laboratorio è aperto a tutti coloro che hanno in mente un progetto di studio o ricerca che parta dai Big Data e che abbia un'utilità pubblica. La struttura metterà a disposizione attrezzature e competenze, permettendo ai cittadini di riappropriarsi delle briciole che lasciano in rete. Grazie ai dati sul traffico, per esempio, gli automobilisti potranno ricevere consigli sulle strade da percorrere. E le foto delle vacanze potranno aiutare altri turisti a pianificare il proprio percorso di visita. «La figura del *data analyst* è nuova e ancora in evoluzione. Sarebbe fondamentale formare persone in grado di districarsi anche tra etica e diritto», riflette Pedreschi. Anche per questo ha deciso di inserire corsi appositi all'interno del master in *Big Data analytics e social mining*, che dirige. In uno scenario di crisi, con l'economia che fatica a riprendere e un mercato del lavoro con sempre meno domanda, i Big Data sembrerebbero offrire grandi possibilità anche in senso occupazionale. E le potenzialità sono ancora tutte da scoprire, visto che il 60% della popolazione mondiale ancora non ha accesso a internet.

Definito il lavoro più "sexy" dall'economista Hal Ronald Varian sull'*Economist*, quello del *data analyst* sembra un mestiere in grado di aprire posizioni importanti e ben retribuite. Il *data analyst* è a tutti gli effetti un manager, che copre in maniera trasversale tutti i reparti di un'azienda, trasformando i dati in informazioni comprensibili e utili per prendere decisioni. Per chi sa districarsi nella giungla dei dati il futuro appare roseo: secondo uno studio, entro il



2020 il mondo sarà composto da più di 7 miliardi di persone, 10 milioni di aziende e, soprattutto, oltre 30 miliardi di dispositivi elettronici connessi alla rete. Un *data analyst* può aspirare addirittura anche a un lavoro alla Casa Bianca. Lo scorso febbraio Barack Obama ne ha assunto uno, istituendo per la prima volta il ruolo di *chief data scientist* nello staff presidenziale. Il fortunato è Dj Patil, un matematico indo-americano con un passato di analista in diverse *big company* a stelle e strisce. Oppure, rimanendo nel Belpaese, un *data scientist* potrebbe lavorare nel centro di ricerca che sorgerà a Rho, lì dove c'era il sito dell'Expo e dove lo stato italiano sarebbe pronto a investimenti a otto zeri.

ACCESSIBILITÀ E QUALITÀ

La privacy non è l'unico nodo da sciogliere. In questi anni sono in atto dibattiti accesi sui diritti dei cit-

Ogni minuto vengono caricate 60 ore di video, inviate 200 milioni di e-mail ed effettuate più di 9 milioni di telefonate

tadini a essere informati sui dati, e sulla qualità e attendibilità dei dati stessi. La legge italiana garantisce dal 2009 l'accesso ai dati delle pubbliche amministrazioni, tuttavia al momento i giornalisti che lavorano con i numeri, i *datajournalist*, lamentano una grande difficoltà a reperire le informazioni che servono per le loro inchieste. Il nostro Paese, infatti, si trova al venticinquesimo posto della classifica sui dati aperti, cioè facilmente accessibili. In testa alla graduatoria c'è la Gran Bretagna, mentre l'Italia ha ancora molta strada da fare per far conoscere come operano le sue istituzioni pubbliche e a quali spese. Sulla qualità delle ricerche condotte con i Big Data ci sono pochi dubbi: potendo studiare un numero altissimo di persone e variabili di comportamento il potenziale di affidabilità è enorme. La differenza rispetto alle vecchie indagini statistiche riguarda il campione. Quel ristretto gruppo di persone che gli

1234689x-3+
+3-w/32-x

esperti selezionavano con cura per essere rappresentativo di tutta la popolazione, con i Big Data non ha più senso di esistere. Così come non sono più necessarie le complesse estrapolazioni per generalizzare i dati relativi a pochi individui, perché oggi è possibile osservare una gran parte della società. Che i Big Data possano aiutare a fare previsioni lo dimostra il giornalista e guru della statistica Nate Silver, che nel 2008 indovina i risultati delle elezioni americane in 49 stati su 50 e quattro anni più tardi fa l'en-plein. E ci sarebbero i Big Data anche dietro alla vittoria dell'ultimo mondiale di calcio. La Germania si era affidata a telecamere e app in grado di elaborare le tattiche degli avversari e suggerire schemi con cui affrontarle.

Tuttavia qualcuno esorta a non cedere a facili entusiasmi. Viktor Mayer ha scritto un libro dal titolo eloquente: "Big Data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la vostra libertà". Il professore di Oxford ammonisce: i Big Data sono solo uno strumento, servono a informare più che a spiegare, ci aiutano a capire ma possono anche indurci al fraintendimento a seconda di come vengono utilizzati. Questo perché sarà sempre impossibile rappresentare la società nel suo insieme. «Tra non molto il mondo dei Big Data, così come lo conosciamo, apparirà ridicolmente superato, come oggi appaiono i quattro kilobyte di memoria scrivibile di cui disponeva il computer di controllo dell'Apollo 11», riflette Mayer, tirando in ballo persino Platone. Le informazioni che siamo in grado di raccogliere e processare, infatti, visto che rappresentano solo una frazione della complessità, sarebbero come le ombre che si muovevano sulle pareti della caverna del filosofo greco. Non sbagliate, ma incomplete.