

Dati e social media: quanto siamo prevedibili? E quanto valiamo?



Sì, la vicenda dei dati ceduti, rubati, usati per convincere, manipolare e disinformare sembra una narrazione distopica. La cosa che renderebbe poco credibile l'intera storia, se di narrazione distopica davvero si trattasse, è che tutto quanto ruota attorno alle sbrigative scelte imprenditoriali di un biondino di trentatré anni, che nel 2003 fatica a rimorchiare le ragazze e nel 2017 si trova a capo della quarta maggior società quotata al mondo. Il 97 per cento dei cui ricavi (dato Business Insider) deriva dalla pubblicità.

DA FACEMASH A FACEBOOK. L'altra cosa degna di nota è questa: già quando nel 2003 “inventa” The Facemash, la primissima versione di Facebook, Zuckerberg viene accusato dall'università di Harvard, presso la quale studia, di aver violato la privacy, le norme di copyright, le norme di sicurezza. Per questi motivi rischia di essere espulso, ma alla fine non se ne fa nulla.

PROFILARE E SEGMENTARE. State leggendo il terzo di una serie di articoli sulla questione Facebook-Cambridge Analytica. [Il primo racconta](#) come le informazioni di cui disponiamo orientano le nostre decisioni. [Il secondo dice come](#) due giovanotti di Cambridge abbiano inventato, e si siano lasciati scappare, un metodo che permette di profilare e segmentare il pubblico in rete.

UNIRE ELEMENTI ESISTENTI. La cosa clamorosa – lo ripeto, che male non fa – è che i giovanotti di Cambridge hanno fatto tutto ciò semplicemente mettendo insieme cose che già c'erano: la potenza di calcolo dei computer, i social media e i dati che noi disseminiamo navigando, la ricerca psicografica, un efficace modello della personalità chiamato Big Five.

Se creatività è unire elementi esistenti in combinazioni nuove e utili, come diceva Poincaré (e come umilmente sostiene anche la sottoscritta) non c'è dubbio che si sia trattato di un gesto creativo a suo modo esemplare.

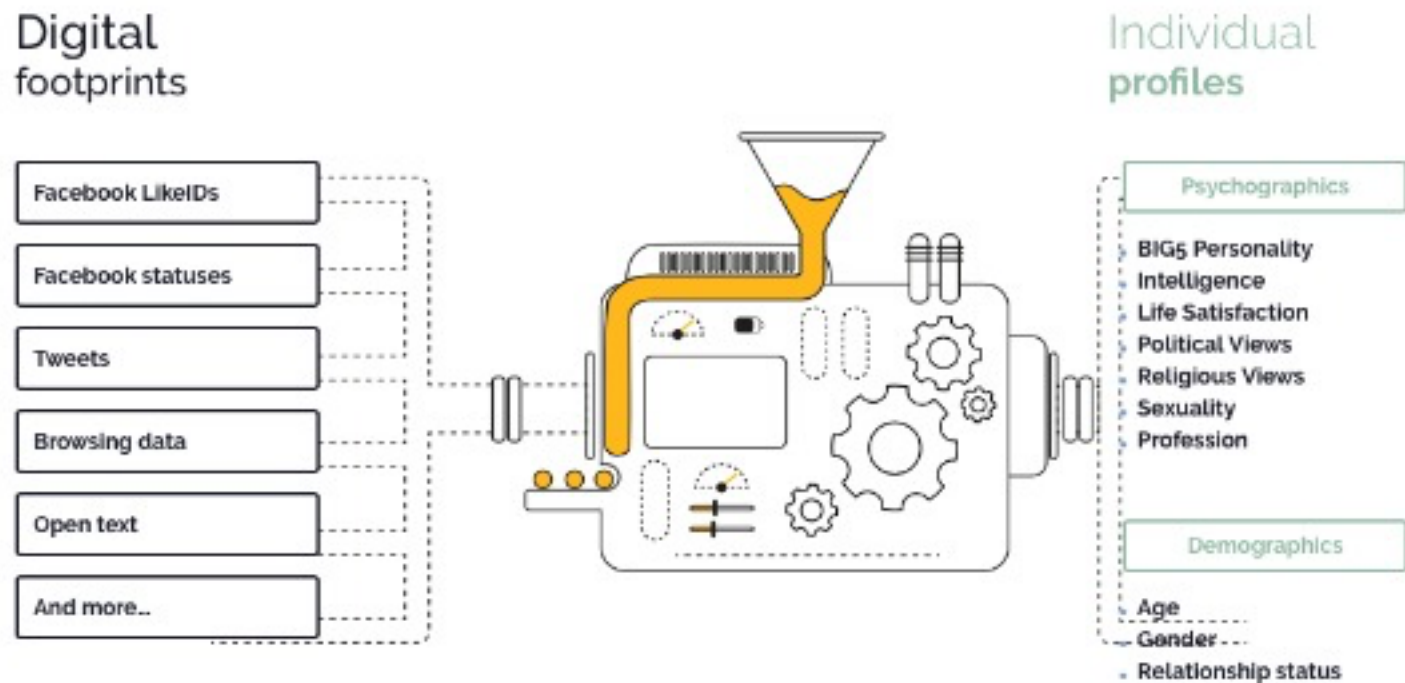
GRANDI NUMERI PER GRANDI RISULTATI. Ma quanto funziona, tutto ciò? Ed è “utile” a chi? Cominciamo dalla prima domanda: sì, funziona. Funziona un po' perché siamo più prevedibili di quanto ci piaccia pensare, molto perché i grandi numeri (i *big data*), la potenza di calcolo e gli algoritmi permettono di andare davvero per il sottile con la profilazione. In sostanza: si possono calcolare molte cose, molto velocemente, in modo molto accurato, a partire da una casistica molto vasta e disponendo di molti ma molti dati su ogni singolo individuo che passi un po' del suo tempo in rete, anche solo navigando.

UNA MANCIATA DI DATI. Ma tutto ciò, ormai, funziona perfino a partire da una manciata di dati. Se volete rendervene conto, potete cimentarvi con il test originale, quello messo a punto dall'università di Cambridge, i cui meccanismi sono stati poi scippati da Cambridge Analytica. Il sito si chiama [applymagicsource](#).

Ne ho sentito parlare diversi mesi fa, all'Internet Festival di Pisa, da [Vesselin Popov](#), il coordinatore dello *Psychometrics Center* dell'università di Cambridge. Un'immagine (la trovate anche [a questa pagina](#)) dà un'idea

di ciò che succede.

Trait Prediction Engine



“MEGLIO DI TUA MADRE”. A Pisa, Popov ha affermato che, se ha sufficienti dati, il computer può conoscerti *meglio di tua madre*. Così, sono andata a vedere se era vero. E mi sono tenuta le schermate, che ora recupero per raccontarvele.

Dunque. Nel momento in cui faccio il test (novembre 2017) il sistema mi dice che su Facebook ho meno di 150 like (peraltro piuttosto datati).

Trascura i like nazionali (per esempio il like al Salone del Libro) e trascura tutti i testi, che ovviamente sono in italiano. La cosa da una parte mi fa sentire poco rilevante ma dall'altra, lo ammetto, mi conforta. Il sistema considera solo 6 like a pagine in inglese il cui significato, con ogni evidenza, è stato censito e catalogato. Non si tratta certo di pagine di nicchia: ci sono, per esempio, il Moma e la Harvard Business Review.

COMBINARE I DATI. Ciascun singolo like è irrilevante come un singolo punto nello spazio. Ma a essere potente è la triangolazione combinatoria, ben roduta sui 6 milioni di profili di utenti forniti spontaneamente, che hanno costituito il *database* di partenza. Ed è potente perfino se i punti da unire sono pochissimi: il minimo sarebbe 10.

Triangolando 6 like, il sistema riesce comunque a calcolare che sono

(probabilmente), una donna. Con un certo temperamento artistico. Che ha un atteggiamento contemplativo, una mentalità analitica ed è di orientamento *liberal*. E così via, azzeccandoci più che decentemente. E ancora: che non appartengo a un credo religioso.

NON CI CASCO! Aggiunge che la mia formazione coinvolge il giornalismo, la psicologia, il business... il sistema fa inoltre plausibili ipotesi sul mio grado di soddisfazione nella vita e sulla mia situazione sentimentale. Mi chiede anche se voglio confermare il mio genere o se voglio fare un ulteriore test di personalità, ma ovviamente non ci casco. Il sistema dichiara anche la probabilità che le sue ipotesi corrispondano alla realtà. Ovviamente le percentuali sono bassine, anche se sfiorano il 70% per quanto riguarda l'orientamento politico. Certo, se i like fossero 300 invece che sei, la probabilità che il sistema azzechi tutto, ma proprio tutto, diventerebbe altissima e sfiorerebbe il cento per cento.

INGANNARE IL SISTEMA? E certo, potrei ingannare il sistema mettendo *mi piace* a pagine di candidati conservatori americani, di musica techno, di motociclette o di cartoni animati o di astrofisica, ma l'obiettivo non è questo. Al sistema, di me come persona, o dei miei dati presi singolarmente e del senso che hanno per me, non importa un fico secco. Io sono solo un produttore (di dati da accorpare) e un prodotto, in quanto individuo accorpabile a un target omogeneo, l'accesso al quale può essere comprato e venduto.

IL PRODOTTO SEI TU. *Se qualcosa è gratis, vuol dire che il prodotto sei tu.* Lo scriveva Time, a proposito di Facebook, [già nel 2010](#), ma forse il peso di questa affermazione comincia a essere evidente solo adesso. I social media possono vendere un sacco di pubblicità mirata, quindi efficace, quindi desiderabile sul mercato, a prezzi stracciati e tali da mandare in crisi il complesso dei media classici, proprio perché ormai riescono a conoscere gli utenti "meglio delle loro madri".

DUE MILIARDI. Lo [scrive Huffington Post](#): Facebook possiede i dati di due miliardi di persone, [foto e messaggi privati compresi](#), e solo per questo motivo vale in Borsa 500 miliardi di dollari. E nel mondo ci sono decine di

migliaia di società che fanno lo stesso mestiere: estrarre, o vendere e comprare, ed elaborare dati degli utenti, utili a veicolare offerte mirate di qualsiasi cosa.

È la [personalizzazione di massa](#), bellezza. Quella che, proprio fondandosi sull'omologazione dei comportamenti, seduce con l'illusione dell'unicità. Un bel paradosso.

SOTTRARRE DATI. E ci sono molte altre entità che fanno una cosa più semplice ancora: sottrarre dati. Del resto, se i dati sono preziosi, e se vengono accumulati, è ovvio che qualcuno proverà prima o poi [a rubarli](#). Per dire: a fine 2017 Yahoo, il secondo servizio di posta elettronica al mondo, ha comunicato che [tutti i suoi 3 miliardi](#) di account-utente sono stati violati.

REGOLAMENTI. La buona notizia è che tra poco [verrà varato in Europa](#) un più severo regolamento per la protezione dei dati. Ma la questione, in realtà, va oltre la privacy, e anche oltre le complesse schede di consenso che appariranno in rete, e che quasi tutti compileranno, acconsentendo senza nemmeno leggerle.

PERMANENTI E COMMERCIALIZZABILI. La questione riguarda la nostra capacità di discernere tra le offerte, commerciali e non, che la rete ci imbandisce dopo averci profilati perbene. Riguarda la nostra capacità di cercare informazioni, in rete e altrove, che non siano necessariamente confezionate per noi.

Riguarda il fatto che tutto ciò che cerchiamo, guardiamo, compriamo, scriviamo e pubblichiamo in rete si trasforma in dati permanenti e commerciabili. E riguarda il lavoro che compiamo producendo dati per società che nemmeno pagano le tasse.

CATALOGATI DA UNA MACCHINA. La questione riguarda i nuovi sistemi di domotica, capaci di estrarre dati anche dai nostri gesti quotidiani, e riguarda i ben noti sistemi di geolocalizzazione, che registrano non solo i nostri percorsi, ma anche [i locali e i negozi che visitiamo](#) e le foto che scattiamo in quei luoghi. Riguarda la sensazione di poter essere analizzati, riconosciuti e catalogati da una macchina.

PROFILAZIONE DISCRIMINATORIA. La questione riguarda la possibilità che una profilazione sempre più accurata venga in futuro usata in maniera discriminatoria, nei confronti di chi cerca un lavoro, chiede un prestito, vuole affittare una casa o stipulare un'assicurazione.

Riguarda perfino il fatto che chi viene individuato come consumatore abbiente possa vedersi offrire prodotti più costosi, o che gli vengano presentati prezzi più alti per gli stessi prodotti.

LA MOLE DEI DATI. La questione riguarda la mole dei dati in sé: che cosa faremo con i [163 zettabyte](#) di dati che si accumuleranno nel 2025, compreso [il primo video di gatti](#) caricato su Youtube, i quasi 74 milioni (ad oggi) di video di gatti che l'hanno seguito e tutti gli auguri di buon Natale che la sottoscritta ha ricevuto nel 2009

LEGGI ANTITRUST. Ma, in primo luogo, la questione riguarda il fatto che, come scrive il sociologo William Davies su Internazionale (1.4.18), le leggi per la protezione della privacy e dei dati non basteranno: *distuggere la privacy in modi sempre più avventurosi è il lavoro di Facebook. Proprio come gli ambientalisti chiedono all'industria dei combustibili fossili di "lasciarli nel sottosuolo", quello che dovremmo fare è chiedere alla Silicon Valley di "lasciare le informazioni nelle nostre teste"... Il modo migliore per fermare tutto questo è quello che la Silicon Valley teme più di ogni altra cosa: le leggi antitrust.* La proposta di Davies è più che sensata. Ma non sempre le cose più sensate sono anche le più facili da fare.

Questo articolo esce anche su [internazionale.it](#)