

Chi ha paura dei big data?

di Riccardo Puglisi

L'utilizzo pervasivo dei dati sul comportamento online inizia con una scelta dell'utente: la rinuncia alla privacy per ottenere un servizio personalizzato. E forse non dobbiamo preoccuparci tanto della persuasione politica, quanto di quella commerciale.

Lo scandalo Cambridge Analytica

Forse per la prima volta, i big data relativi al nostro comportamento su internet e social network appaiono più come una minaccia al funzionamento dei meccanismi sociali e politici che come un modo intelligente di usare la tecnologia per fornirci contenuti personalizzati, dopo che abbiamo accettato “con un click” di ridurre la nostra privacy a favore del nostro social network o motore di ricerca preferito.

Lo scandalo del Datagate, come ben raccontato da Fabio Chiusi su [Valigia Blu](#), coinvolge il social network Facebook, la società Cambridge Analytica finanziata da Robert Mercer, miliardario fondatore dell'*hedge fund* Renaissance e di chiaro orientamento politico conservatore, alcuni ricercatori come Alexandr Kogan – diviso tra l'Università di Cambridge e l'Università statale di San Pietroburgo – e Christopher Wylie, che ha lavorato sia per l'(ex) ideologo di Donald Trump, Steve Bannon, che per Cambridge Analytica, e che poi è diventato la gola profonda per gli articoli del *Guardian* sul tema.

Lo scandalo sta nel fatto che dati raccolti da application terze a partire dai comportamenti individuali su Facebook non sono stati utilizzati soltanto a fini accademici, ma sono stati poi sfruttati per rendere più precisi ed efficaci i messaggi propagandistici utilizzati dalla campagna a favore di Brexit in Gran Bretagna e per la campagna presidenziale di Trump negli Usa. L'idea sottostante è quella di prevedere le caratteristiche psicologiche di ogni individuo sulla base di uno schema standard – il cosiddetto modello Ocean – che incorpora le cinque dimensioni di “Openness” (apertura a nuove esperienze), “Conscientiousness” (consapevolezza), “Extroversion” (estroversione), “Agreeableness” (socievolezza) e “Neuroticism” (Nevrosi) – collegando questi tratti della personalità a comportamenti online: si va da quelli più palesi come le pagine di Facebook su cui si mette il “like” ad altri meno visibili ai propri “amici digitali” – e dunque meno soggetti ad autocensure – come la frequenza e la distribuzione temporale dell'atto stesso di mettere like e di caricare foto su Facebook.

Se un ricercatore è in grado di prevedere i tratti della personalità di un individuo sulla base del suo comportamento online, e se è possibile costruire messaggi propagandistici più adatti a persone che hanno un certo profilo caratteriale, così da incrementare la propensione di costoro a diffondere i messaggi stessi o da modificarne il comportamento “esterno” (nelle scelte economiche o politico-elettorali), allora la bellezza e l'utilità di contenuti personalizzati su internet si trasformano in una trappola pericolosa, perché aumenta di molto la possibilità che qualcuno influenzi in maniera subdola e “inconscia” tali comportamenti, arrivando perfino a manipolare l'esito di elezioni cruciali come le presidenziali Usa o il referendum su Brexit.

Il meccanismo di predizione può funzionare soltanto se siamo in grado di misurare con un test standard i cinque tratti di personalità di individui di cui conosciamo anche il comportamento online. Ebbene, qui il contributo più rivoluzionario è forse quello di Michal Kosinski e coautori nell'articolo pubblicato su [Machine Learning Journal](#) nel 2013. Kosinski ha creato questo “ponte” convincendo centinaia di migliaia di persone a effettuare il test di personalità online in cambio dell'accesso alle informazioni relative al loro comportamento su Facebook. Tecniche statistiche

adatte ai big data permettono poi di trovare i comportamenti osservabili su Facebook che sono meglio capaci di prevedere i tratti di personalità, esattamente basandosi sugli individui di cui conosciamo entrambe le “sfere”. A questo punto si possono estrapolare i risultati del modello statistico e collegare direttamente i comportamenti osservabili sui social network a un messaggio oggetto di “microtargeting”, cioè adatto a massimizzare la probabilità di condivisione e persuasione per quel sottogruppo dotato di quelle caratteristiche.

Va chiarito però che per l’assistenza alla propaganda elettorale di Brexit e Trump, Cambridge Analytica non ha utilizzato i dati originali raccolti da Kosinski, quanto qualcosa che assomiglia a una loro replica, grazie all’opera di Alexandr Kogan, il quale ha assoldato un numero importante di utenti online tramite Mechanical Turk (un’applicazione di Amazon che paga qualche dollaro a coloro che effettuano compiti ripetitivi e non appaltabili a una macchina) e li ha sottoposti al test di personalità chiedendo di condividere i dati sul loro account Facebook.

Il pericolo della persuasione commerciale

Vale qui la pena formulare alcune riflessioni generali.

Come rilevato da Chiusi nell’articolo citato sopra, si nota una singolare diversità di approccio da parte dei media intorno all’utilizzo dei big data a fini elettorali: ottima cosa quando in ballo c’era Barack Obama, pericolosa manipolazione nel caso di Trump e Brexit. Ma è singolare come nessuno – o quasi – abbia riflettuto sul fatto che la valutazione cambia di molto se nell’utilizzo dei big data a fini elettorali c’è concorrenza o se, invece, c’è un monopolio da parte di un politico “pioniere” che per l’appunto “straccia la concorrenza” con la nuova tecnologia.

Si tornerebbe dunque ai buoni vecchi concetti economici di concorrenza e monopolio, tenendo però presente che stiamo probabilmente parlando di un gioco “a somma zero”, in quanto gli apparati tecnologico/propagandistici dei due team elettorali lottano per gli stessi voti: non c’è più l’orwelliano pericolo di una manipolazione inconscia dei risultati elettorali ma il più becero – e meno preoccupante – rischio di uno spreco di risorse. Ma ciò non vale per ogni campagna elettorale? E ancora: siamo davvero sicuri dell’efficacia persuasiva di queste tecniche propagandistiche? La ricerca accademica (ad esempio questo tweet di Brendan Nyhan, sempre citato da Chiusi) è molto più scettica sugli effetti di persuasione elettorale rispetto alla questione a monte della predicibilità dei tratti di personalità a partire dal comportamento online. Se lo scetticismo è fondato, il rischio che corriamo è quello – davvero ironico – di guardare all’inetto topolino della persuasione elettorale invece che alla poderosa ed efficace montagna della persuasione commerciale.