

IL CASO

L'influenza non si cura con Google

EUGENIA TOGNOTTI

E pensare che era la star del mondo Big Data, il Google Flu Trend (Gft), trionfalmente presentato nel 2008 come un prezioso strumento di salute pubblica. E lo sarebbe stato davvero se avesse mantenuto la promessa di monitorare in tempo reale i casi d'influenza sulla base dei termini di ricerca associati con quella malattia su Google: brividi, debolezza, febbre, mal di testa, tosse, mal di gola. Invece è stato un flop. Ha sovrastimato la prevalenza dell'influenza nella stagione 2012-2013 di oltre il cinquanta per cento.

CONTINUA A PAGINA 27

L'INFLUENZA NON SI CURA CON GOOGLE

EUGENIA TOGNOTTI
 SEGUE DALLA PRIMA PAGINA

E' stato impreciso sul picco della stagione influenzale. Ancora. Da agosto 2011 a settembre 2013 ha sbagliato le previsioni relative a 100 settimane su 108. Per non parlare dell'influenza pandemica da virus H1N1, comparsa sulla scena a Pasqua, quando il Gft aveva già fatto le sue previsioni, sottovalutando la dimensione del focolaio fuori stagione. Insomma, il tentativo di inseguire i focolai di influenza in base ai termini di ricerca ha fallito l'obiettivo che s'era proposto.

Dando per certo che la prima reazione della gente colpita da influenza o sindrome influenzale è quella di cercare informazioni sul web, Gft puntava a prevedere le epidemie, contando sul fatto che da Big Data arrivasse una valanga di informazioni tali da aiutarci a fare cose che sarebbero impossibili con un volume inferiore di dati, raccolti in modo tradizionale. Come quelli forniti dall'organismo di controllo della sanità pubblica degli Stati Uniti, il Cdc (Atlanta) sulla base di vi-

site mediche e di dati clinici e virologici.

Dopo aver raccolto più di cinquanta milioni di potenziali termini di ricerca - ogni genere di frase collegata a «influenza» - e confrontato la frequenza con cui la gente aveva cercato quei termini con il numero dei casi di influenza nel triennio 2003-2006 - il team di Google aveva individuato alcune decine di frasi adatte (45). La verifica effettuata durante l'epidemia influenzale del 2007 aveva confermato che, in effetti, le previsioni si avvicinavano ai livelli di malattia nel mondo reale. Cosa che giustificava l'entusiasmo, un po' esagerato, in verità, con cui il progetto Gft era stato presentato su Nature, che esaltava la possibilità di riuscire a prevedere un'epidemia d'influenza, battendo sul tempo il Cdc. L'era dei Big Data sembrava già qui. Promettendo di aprire la strada alla comprensione della diffusione delle malattie, monitorata con gli algoritmi. Il fatto è che i numeri restano numeri al di fuori di un contesto. Il Gft può «misurare» ciò che la gente cerca, ma non può cogliere il perché la gente va alla ricerca di certi termini abbinati a regolari epidemie stagionali del 2003-2008. Termini come «tosse» o «febbre», correlati con l'influenza, potrebbero essere in realtà sintomi di altre malattie stagionali. E che parole ha cercato la gente, impressionata dall'allarme per la minaccia pandemica del 2009?

In conclusione, la lettura della realtà basata sull'analisi dei Big Data è disseminata di trappole e lo spiega bene uno studio comparso sull'ultimo numero di Science intitolato «The Parable of Google Flu: Traps in Big Data Analysis» che esamina nel dettaglio le problematiche legate all'uso di grandi dati da aggregatori come Google. Gft - ha detto Ryan Kennedy, uno degli autori, professore di Scienze Politiche all'Università di Houston - è uno straordinario pezzo di ingegneria e uno strumento molto utile. Ma occorrerà apportare dei correttivi. «La nostra analisi dimostra che i risultati migliori si ottengono abbinando informazioni e tecniche, i big data con le metodologie più tradizionali». Nessun dubbio che i Big Data avranno alcune applicazioni importanti nei prossimi anni anche nel campo della Medicina. Ci sono già alcuni esempi di successo. Ma occorrerà aspettare la prossima stagione influenzale per vedere se Gft potrà entrare a farne parte.

