Immuni, analizzare i dati per guardare oltre i dati dell’app

Esattamente tre settimane fa il Commissario straordinario (per l’attuazione e il coordinamento delle misure occorrenti per il contenimento e contrasto dell’emergenza epidemiologica COVID-19) ha finalmente reso disponibili alcuni dati relativi all’utilizzo della app di contact tracing italiana, la nota **immuni**.

Le 2 domande

Possiamo quindi chiederci: cosa possiamo capire a partire da questi dati, che comprendono il numero di volte che l'app è stata scaricata, e il numero di notifiche inviate e utenti positivi ogni giorno dallo scorso Giugno?
Possiamo capire se la app è efficace e quindi è una misura di contenimento della pandemia che mostra elementi di costo-efficacia? (cioè vale il denaro pubblico che vi si spende, in manutenzione correttiva, gestione server, operatività del call center e varie attività promozionali).

Anticipiamo subito una risposta per certi versi deludente: no, non possiamo capire granché. Infatti, molti ricercatori (tra cui chi scrive da circa 5 mesi), avevamo segnalato la necessità di ottenere quei dati insieme *ad altri* dati, tra cui, lo ricordiamo: il numero di utenti attivi, cioè di persone che usano effettivamente la app. Questo numero è ovviamente una porzione di quanti la scaricano, togliendo quanti l'abbiano provata per un po' e poi l'abbiano disattivata o disinstallata dal proprio cellulare. Il Governo Italiano non conosce questo dato, sebbene si ritenga fondamentale che una larga porzione di cittadini Italiani sia utente attivo di immuni affinché questa sia di qualche utilità*. Sulla base dei dati resi disponibili da altri gestori di app simili (che provano almeno a stimare questo numero), ad esempio swisscovid in Svizzera, è probabile che il numero degli utenti attivi sia compreso tra due terzi e tre quarti delle persone che l'hanno scaricata: noi prendiamo il 70% come stima conservativa; il numero di ore (giorni?) che passano da quando l'app riceve informazioni sui contatti stretti di una persona positiva a quando essa notifica queste persone (un bug scoperto qualche settimana fa, e poi apparentemente risolto, di cui ha parlato anche un servizio di Paternesi andato in onda su **Report**, riguardava proprio questa funzione); il numero di giorni che passano da quando la persona riceve la notifica e quindi può attivarsi per

determinare la sua positività sulla base di un test medico, a quando la stessa riceve l'esito del test: questo dato permetterebbe di stimare l'impatto sociale dei falsi positivi (che qualcuno, sempre in Svizzera, cerca di quantificare e riconoscere), in quanto si suppone che nel frattempo la persona notificata si sottragga al tessuto sociale e produttivo del paese scegliendo responsabilmente di mettersi in isolamento domiciliare fiduciario. E, infine, il numero più importante per determinare l'efficacia di immuni**: la proporzione di persone davvero positive (identificate come tali da un test molecolare o parimenti affidabile) tra quelle che hanno ricevuto una notifica di "contatto a rischio" (cioè contatto stretto di persona positiva). Ebbene, nessuno di questi dati è ancora disponibile, anche se sarebbe molto utile cercare di ottenere almeno una stima delle grandezze in gioco, anche su base semplicemente campionaria.

Guardare i dati, capire la realtà

Ad ogni modo, cosa possiamo quindi dire a partire dai pochi dati che sono condivisi quotidianamente dai gestori di immuni? Facciamo qualche considerazione di carattere qualitativo con la visualizzazione approntata da **Enrico Conte**, uno studente del corso in Data Science dell'Università di **Milano-Bicocca** in cui insegno. Ci basiamo su un semplice grafico della serie temporali dei dati medi a 7 giorni, così da "pulire il segnale" delle fluttuazioni che siamo abituati a vedere nel numero di positivi, che inevitabilmente riflette tanto l'andamento della pandemia quanto il diverso numero di test effettuati nei giorni festivi e prefestivi***.

1) Possiamo iniziare con una ovvietà: e cioè che il numero di download di immuni è cresciuto costantemente da quando è stata resa disponibile agli inizi dello scorso giugno (linea rosa nel grafico), anche se questo non significa necessariamente che sia cresciuto *monotonamente* (cioè senza flessioni) anche il numero di utenti attivi (linea rossa nel grafico). Questa crescita ha avuto tre momenti di crescita più significativi di altri: la prima, ovviamente, quando l'app è stata resa disponibile al pubblico Italiano, lo scorso Giugno; la seconda, in concomitanza dell'inizio delle scuole (ad esempio, il 14 settembre in Lombardia); e poi una terza, ai primi di Oottobre, quando la comunicazione da parte degli organi istituzionali e alcuni mass media è divenuta più insistente e, soprattutto, il numero di positivi è tornato purtroppo a crescere in modo esponenziale. I cittadini Italiani hanno

quindi riposto una qualche fiducia che questo strumento potesse essere utile per contenere l'aumento dei contagi; però da qualche giorno, anche in virtù di qualche scoraggiante notizia, questa fiducia sembra essere come sospesa, e infatti il numero di utenti attivi non aumenta praticamente più da qualche giorno, assestandosi a meno di 8 milioni.

2) Possiamo dire che il dpcm del 18 ottobre, quello che, tra le altre cose, ha fatto obbligo agli operatori sanitari che riscontrassero (o a cui venisse comunicata) la positività di una persona che dichiara di usare immuni di attivare la procedura di sblocco e notifica dei suoi contatti stretti****, *non ha ancora sortito l'effetto desiderato*: cioè di rendere utile immuni, permettendole di fare il suo lavoro (bene o male), cioè notificare le persone (considerate) a rischio. Guardando i dati (linea viola tratteggiata in Figura), possiamo vedere che la forbice tra la proporzione di utenti positivi sulla popolazione di potenziali utenti inizia a crescere solo ai primi di Ottobre (quindi anche prima del 18 Ottobre ma terribilmente in ritardo rispetto all'inizio del programma di tracciamento digitale), e in maniera molto meno marcata di quanto cresca il numero di positivi nella fascia della popolazione tra i 15 e i 75 anni (linea tratteggiata blu). Al momento, questo rapporto è ancora circa dieci volte inferiore***** a quello che dovrebbe essere se il meccanismo di notifica funzionasse come previsto.
3) Da ultimo, possiamo dire che le misure di confinamento (lockdown) a livello regionale hanno funzionato, in qualche misura: infatti si può vedere che dai primi di Novembre (il dpcm è del 3 Novembre) si assiste ad una chiara flessione nel numero di contatti stretti che hanno ricevuto una notifica per ogni singolo utente positivo (linea dorata in Figura): scendono da un massimo di 26 persone per utente positivo, a meno di 20; segno questo che le reti sociali degli utenti di immuni, o meglio la rete dei loro contatti stretti, si riducono. Queste reti sociali cambiano anche in virtù del numero di persone che usano immuni ovviamente (più utenti la

usano, maggiore è la proporzione dei propri contatti che può essere notificata), ma da qualche settimana, come possiamo vedere dal grafico, il numero di utenti di questa app è praticamente costante (circa 7 milioni, se vale la approssimazione di cui sopra), quindi possiamo dire che il calo è principalmente dovuto al fatto che gli utenti di immuni stanno frequentando (strettamente) meno persone di 3-4 settimane fa.

Una morale per una storia ancora da scrivere

Non possiamo dire molto altro e probabilmente nell'ammetterlo siamo delusi tanto quanto il lettore che ci ha seguiti fin

qui. Ma il punto è proprio questo: per mesi la comunità dei ricercatori o semplicemente dei cittadini desiderosi di capirci qualcosa non ha avuto alcun dato per capire come (e se) stesse funzionando l'app immuni. Da 3 settimane sono condivisi alcuni dati e ci si potrebbe anche accontentare ("meglio di niente") se in realtà questi non fossero utili a comprendere la cosa più importante: e cioè se i soldi pubblici spesi per questa iniziativa sono ben spesi oppure no; e se in futuro (prossimo o remoto) sia opportuno investire risorse e aspettative su questo genere di strumenti (e non altri). Non si consideri la questione frivola: abbiamo capito (forse), e a nostre spese, che investire nel digitale, nella ricerca, nell'assistenza sanitaria e nella prevenzione (campagne vaccinali, scorte di presidi, pianificazione strategica) è un investimento che può evitare di dover far fronte a spese praticamente insostenibili per qualsiasi economia nazionale (solo a titolo di esempio, ogni malato COVID ricoverato in un ospedale costa alla collettività dai 9 ai 22 mila euro a seconda della intensità di cura e dall'inizio della pandemia l'Italia ha registrato circa due milioni e mezzo di giornate di degenza in regime ordinario e 260.000 giornate in terapia intensiva per la COVID-19). Questi investimenti richiedono strategie e scelte politiche che non possono non basarsi su evidenze che emergano dall'analisi incrociata di dati provenienti da innumerevoli fonti e raccolti con diversi mezzi. *Il dato non è un lusso*, nelle nostre società globalizzate e iperconnesse, né una risorsa da centellinare a pochi privilegiati: è una necessità che permette di prendere decisioni corrette e tempestive e di adottare consapevolmente comportamenti efficaci. Eppure il dato (cioè definirlo, raccoglierlo, analizzarlo) *costa*, e richiede tanto competenze specialistiche (adeguatamente formate nelle nostre università e centri di ricerca) quanto una *numeracy* ampia e diffusa nella cittadinanza (inquieta che non esista ancora una parola Italiana per questo concetto).

È quindi opportuno che una nuova e più matura cultura del

dato si diffonda, tanto tra i decisori politici, quanto nella più vasta platea di chi li sceglie e aderisce più o meno convintamente alle loro politiche, affinché si faccia un uso sempre più consapevole e responsabile degli strumenti che la nostra società può mettere in campo per fronteggiare le crisi che si affacciano al nostro futuro (tra cui il pericolo di nuove epidemie, il cambiamento climatico e le migrazioni che questo comporta). Anche questo può suggerire il semplice guardare ai pochi dati che abbiamo sulla app immuni.

Note

* per fortuna non ci si appassiona più a stimare la quota minima necessaria perché la app sia utile; questa informazione è stata usata probabilmente più per stimolare l'adozione della app ("dobbiamo raggiungere il 60%"); poi come alibi per il suo insuccesso ("solo il 6% degli Italiani l'ha scaricata"); e ora, appurato che non si riesce a superare il tetto del 20%, si tende a fare di necessità virtù ("comunque, più utenti la usano e meglio è").

** lo ricordiamo solo in nota per non annoiare il lettore esperto, e ce ne sono molti oramai: l'efficacia di una qualsiasi app di contact tracing riguarda la sua capacità di identificare (e avvertire) un positivo prima che questi possa infettare altre persone e, quindi, possa suggerirgli di isolarsi, spezzando così la catena di contagio. Ovviamente una app non può sapere se un suo utente è positivo al coronavirus, ma può segnalargli questo rischio se la stima della probabilità che lo sia supera una certa soglia, determinata sulla base di un modello di rischio che considera per quanto tempo è stato vicino (cioè a meno di 2 metri) ad una persona che qualche giorno dopo è risultata positiva. L'errore di un'app del genere può essere di due tipi: falso negativo, quando l'app non notifica e non suggerisce di isolarsi ad una persona che è stata effettivamente contagiata da un altro positivo (ad esempio perché le ha stretto la mano per pochi secondi), lasciandola libera di contagiare altre persone. E il

falso positivo, quando l’app rileva un rischio elevato per una persona che non è stata contagiata: ad esempio perché le due persone, pur stando vicini per più di 15 minuti, indossavano adeguatamente la mascherina, oppure perché non si sono parlati, o si davano perfino le spalle, come su un mezzo pubblico di trasporto.

*** perché in Italia non si vedono più grafici di questo tipo?
**** *“è fatto obbligo all’operatore sanitario del Dipartimento di prevenzione della azienda sanitaria locale, accedendo al sistema centrale di Immuni, di caricare il codice chiave in presenza di un caso di positività”*
*****c’è qualche semplificazione in questa stima. Per il possibile, abbiamo cercato di considerare sia quanti italiani abbiano tra i 15 e i 75 anni (circa 46 milioni), sia quanti positivi alla COVID-19abbiano una età compresa tra 15 e 75 anni.

Immuni, analizzare i dati per guardare oltre i dati dell’app