Privacy e scatole nere

Comunicare i processi opachi prodotti dalla tecnologia digitale

Laureando
Michele Invernizzi
./834165

Relatore
Michele Mauri

Correlatore
Antonio Lioy

—
Politecnico di Milano
Scuola del Design
LM — Design della Comunicazione
A.A. 2016/2017

INTRO

0 — **Premessa** **.9**

TEORIA E CONTESTO

1 — **Information privacy** **.15**
1.1 Una questione terminologica .15
1.2 Definizione dell'intorno .31
1.3 Caratteristiche .58

2 — **Opacità multi-livello** **.63**
2.1 Dati personali: problemi legati all'accesso .64
2.2 Dati personali: problemi legati all'utilizzo .77

3 — **Asimmetrie di potere** **.89**
3.1 Rispondere all'asimmetria: due approcci .95

4 — **Il ruolo del designer e della visualizzazione** **.121**

SPERIMENTAZIONI PROGETTUALI

5 — **Metodo e approccio sperimentale** **.129**

6 — **Esperimento nº 1: ReCon** **.133**
6.1 Il team di progetto .134
6.2 Applicazioni mobile e il controllo delle autorizzazioni .135
6.3 Idea e processo .140
6.4 ReCon 2.0 .145
6.5 Limiti, opportunità, riflessioni .158

7 — **Esperimento nº 2: Facebook.tracking.exposed** **.163**
7.1 Il team di progetto .165
7.2 Facebook e il suo algoritmo .166
7.3 Idea e processo .169
7.4 MetaTimeline: a timeline of timelines .178
7.5 Limiti, opportunità, riflessioni .186

CONCLUSIONI

8 — **Riflessioni finali** **.193**
8.1 Sullo sperimentare .195
8.2 Sul metodo .196
8.3 Prospettive future .198

INTRO

We feel free because we lack the language to articulate our unfreedom.

Slavoj Žižek

0 — Premessa

Viviamo in una "società dell'informazione", dove i dati personali sono diventati una preziosa risorsa economica. Ogni giorno utilizziamo servizi che ci chiedono semplicemente di accettare un'informativa sulla privacy, col tacito accordo di poter poi in cambio raccogliere informazioni su di noi. Informazioni che vanno dal numero di click fatti su un sito alla precisa posizione geografica durante tutto l'arco della giornata. Oltre ad aver migliorato numerosissimi aspetti della nostra vita, questo ha anche fatto sì che istituzioni pubbliche e aziende private abbiano ora la possibilità di ammassare sempre maggiori quantità di dati e, col continuo progresso tecnologico, di analizzarli sempre più velocemente ed efficacemente.

In una società dove condividere informazioni personali per accedere ad ogni servizio diventa la norma, e dove algoritmi proprietari prendono decisioni per gli utenti in base ai loro dati, si viene a creare una pericolosa asimmetria di potere tra il detentore dei dati e l'individuo. Se da una parte questo ha preoccupanti ripercussioni sul potere che gli Stati acquisiscono sui loro cittadini – e.g. le rivelazioni

di Edward Snowden sulla sorveglianza di massa praticata dall'NSA[1], che non saranno però oggetto di questa tesi – anche nel settore privato la condivisione di dati personali ha delle conseguenze dirette sulle persone: social network, motori di ricerca, acquisti online, etc. portano tutti alla profilazione degli utenti a scopi commerciali.

Ci sono diversi attori che tentano di contrastare questa asimmetria. Associazioni, enti, aziende e singoli attivisti si impegnano quotidianamente per alimentare un dibattito sulla privacy che è fondamentale, sia per lo sviluppo di pratiche aziendali che tengano conto dei diritti dei consumatori che per la tutela dei diritti dei cittadini in una società moderna e perennemente connessa.

Dal punto di vista del designer che deve collaborare in progetti sulla privacy, la questione diventa metodologica. Come si possono mappare processi inizialmente invisibili, coperti da un velo di complessità tecnologica e, a volte, di barriere legali? Come queste caratteristiche influenzano il processo progettuale? Gli strumenti e metodi di cui il designer dispone sono ancora adeguati o vanno riadattati? Ne servono di nuovi? Per rispondere a queste domande verrà utilizzato un approccio sperimentale, che parte dall'osservazione della realtà, o di ciò che di essa si può vedere, permettendo allo stesso tempo di generare riflessioni sull'evoluzione di cultura e pratica del design.

[1] Nel giugno 2013 l'ex consulente della National Security Agency ha reso pubblici una serie di documenti che rivelavano i dettagli di diversi programmi di sorveglianza di massa del governo statunitense e britannico, fino ad allora segreti.
https://it.wikipedia.org/wiki/Edward_Snowden

Questa tesi riporta in particolare due esperimenti: la riprogettazione di un'interfaccia per un'applicazione mobile che monitori le informazioni mandate a terze parti dal nostro smartphone; la creazione di uno strumento digitale per portare alla luce la presenza di un algoritmo che influenza la nostra percezione della realtà su Facebook. In questo modo verranno esplorati due aspetti fondamentali: le opacità relative all'accesso dei dati – chi li detiene, quando e quanto spesso vengono utilizzati – e le opacità relative al loro utilizzo – come i dati vengono elaborati e le conseguenze per l'utente.

TEORIA E CONTESTO

Allora si aprirono gli occhi di tutti e due
e si accorsero di essere nudi; intrecciarono
foglie di fico e se ne fecero cinture.

Genesi 3, 7

O credenti, evitate di far troppe illazioni,
ché una parte dell'illazione è peccato. Non vi
spiate e non sparlate gli uni degli altri. [...]

Corano 49:12

1 — Information privacy

Per creare una solida base su cui sviluppare la struttura di questa tesi, è fondamentale analizzare lo stato dell'arte in cui si trova il dibattito intorno alla privacy ad oggi. Per fare questo sono state percorse due strade parallele: da una parte la ricerca bibliografica ha permesso di costruire il vocabolario necessario ad affrontare la tematica, oltre che a fornire i punti di vista e gli avanzamenti di varie discipline sul tema; dall'altra un'analisi del contesto al di fuori del mondo accademico poteva far emergere punti di discussione importanti per l'opinione pubblica ma non affrontati – oppure già risolti – in accademia.

1.1 Una questione terminologica

Privacy è certamente una parola di difficile definizione (Solove, 2016). Un po' come libertà o design, il termine ha al suo interno moltissime sfumature che lo rendono un "termine ombrello", frutto della sua progressiva evoluzione nel tempo e all'acquisizione di nuove accezioni. Oltre alle tracce presenti nei testi sacri di diverse religioni, già Aristotele

nella sua Politica introduceva dei concetti simili riferendosi alla distinzione tra una sfera pubblica, relativa alla vita politica – *polis* – ed una sfera invece privata, della famiglia – *oikos* – (DeCew, 2015). Diversi studi antropologici hanno inoltre dimostrato che varie sfumature del termine esistono in moltissime culture in tutto il mondo (Altman, 1977), a riprova del fatto che esso non sia solo frutto della società moderna né tanto meno un prodotto del mondo occidentale. Certamente però molto del significato odierno della parola ha le sue radici nello sviluppo di una società sempre più individualista nel XIX secolo e particolarmente nel mondo anglosassone, dove la crescente classe borghese arricchita dall'industrializzazione reclamava maggiori tutele riguardo la gestione dei propri affari privati (Iaselli & Gorla, 2015).

Al giorno d'oggi si relaziona alla sfera della privacy tutta una serie di situazioni che variano su uno spettro estremamente vasto, ad esempio: la copertura delle parti intime del proprio corpo; l'uso di tende alle finestre di casa; la pubblicazione delle intercettazioni di un personaggio politico; l'ingiunzione a consegnare i libri contabili di un'azienda durante un processo; il diritto di segretezza del voto; la raccolta indiscriminata di informazioni personali da parte dei governi; la costruzione di pubblicità ad hoc in base ai dati di navigazione online. È chiaro quindi come possano sorgere problemi a livello legislativo e giuridico quando si cerca di segnare i confini che delineino il diritto alla privacy e le sua violazione (Solove, 2006), oppure a livello economico, quando si tenta di stabilire dei modelli per valutarne il valore (Acquisti, 2016). Oltre a diritto ed economia, ogni disciplina che si è occupata di una qualche sfaccettatura di privacy ha dovuto costruire una propria definizione. Anche ai fini di questa tesi è necessario quindi stabilire cosa si intende parlando di privacy, gettando così le basi per un linguaggio comune e sottolineando allo stesso tempo le limitazioni che questa definizione crea.

Nella letteratura una delle definizioni più citate è quella data da Altman nel 1977 che, nonostante sia stata scritta in un periodo molto diverso da quello presente, racchiude l'attributo comune a tutte le sfumature della privacy: per lo psicologo americano essa è "il controllo selettivo dell'accesso all'individuo" (Altman, 1977)[2]. Privacy non vuol dire cioè necessariamente il nascondere o celare noi stessi e tutto ciò che ci riguarda ma è piuttosto la facoltà di poter stabilire quale sia il confine di ciò che riteniamo sensibile rispetto a ciò che consideriamo condivisibile con il resto della società (e tutti i discorsi derivati su quando sia giusto avere suddetta facoltà). Privacy non è l'opposto di condivisione, quanto piuttosto il controllo sulla condivisione.

Ora, la qualità di questa definizione è che accomuna

[2] *[...]privacy as the selective control of access to the self[...]*, traduzione mia.

tutte le accezioni di privacy evidenziandone l'essenza, ma risulta poi troppo generica e vaga per una riflessione più specifica all'ambito di questa tesi. Bisogna perciò stabilire un perimetro preciso entro al quale applicare questa definizione. Kang (Kang, 1998) raggruppa le varie accezioni del termine in tre gruppi principali. Il primo, che viene definito *space privacy*, si riferisce alla parte più materiale e tangibile della privacy, ovvero il confine pubblico-privato in senso di territorio e spazio personale (p. e.: l'atto di chiudersi nella propria camera ma anche il baccano proveniente da una festa al piano di sopra o le chiamate indesiderate di marketing durante i pasti); il secondo, o *decisional privacy*, riguarda la possibilità di auto-definirsi come individuo o fare scelte significative senza pressioni dall'esterno, come ad esempio la libertà di usare contraccettivi o di abortire, oppure di poter votare liberamente; infine l'*information privacy*, ovvero il controllo su come le informazioni personali – qualsiasi informazione che può unicamente identificare un individuo – vengono ottenute, processate, divulgate ed utilizzate.

Le rivelazioni di Edward Snowden[3] nel 2013 sui programmi di spionaggio di massa negli USA e in altre nazioni occidentali hanno riportato l'information privacy al centro di numerosi dibattiti e hanno evidenziato come scenari che poco tempo prima potevano sembrare usciti da un romanzo distopico, siano ormai la realtà dei fatti grazie ad una pericolosa combinazione di innovazioni tecnologiche e concessioni legislative. Se nel settore pubblico incombe la minaccia di un Grande Fratello, anche in quello privato sono state sollevate questioni simili: la diffusione di Internet e dei social network, la crescita di superpotenze economiche in campo tecnologico, i continui progressi nella capacità di calcolo dei processori e l'abbassamento dei costi di archiviazione hanno reso la raccolta, elaborazione e utilizzo di enormi quantità di dati un'operazione estremamente economica e vantaggiosa (Acquisti, 2016).

Usando le parole dello scrittore ed economista Nick Srnicek:

> [...]we can say that the digital economy refers to those businesses that increasingly rely upon information technology, data, and the internet for their business models. This is an area that cuts across traditional sectors – including manufacturing, services, transportation, mining, and telecommunications – and is in fact becoming essential to much of the economy today. [...] In the twenty-first century, on the basis of changes in digital technologies, data have become increasingly central to firms and their relations with workers, customers, and other capitalists. The platform has emerged as a new business model, capable of extracting and controlling immense amounts of data, and with this shift we have seen the rise of large monopolistic firms. (Srnicek, 2016)

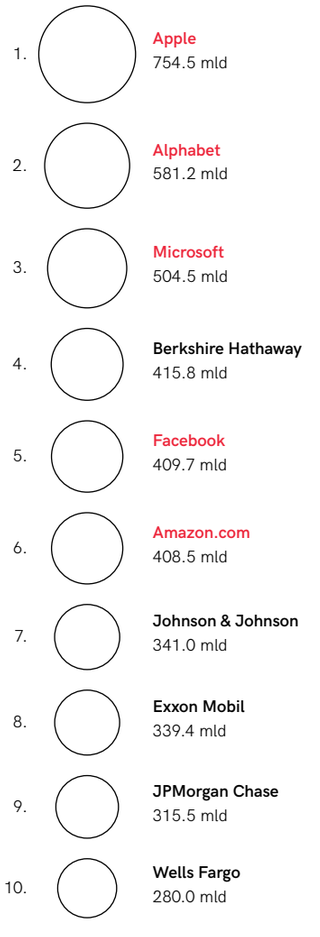


FIG. 01 Classifica delle aziende con la maggior capitalizzazione azionaria. La capitalizzazione è il valore delle azioni di una compagnia ed è usata per stimarne il valore totale. In rosso sono segnate le aziende che si occupano di IT e il valore è espresso in dollari ($).
Fonte: http://dogsofthedow.com
Data: 29 Marzo 2017

[3] I documenti consegnati a varie testate giornalistiche mondiali tra cui il *The Guardian*, il *Der Spiegel* e il *Washington Post*, provavano la stretta collaborazione tra diversi governi – anche quello italiano – ma anche con partner commerciali come Google, Facebook, Apple e Microsoft.
https://en.wikipedia.org/wiki/Global_surveillance_disclosures_(2013-present)

È interessante notare come le preoccupazioni e il dibattito intorno alla privacy, sia legata ai dati generati online che nel mondo reale, siano sempre in qualche modo legate alla tecnologia.

Uno dei primi esempi di cui si ha traccia in una pubblicazione scientifica è del 1890 (Kang, 1998). In un articolo della neofondata (1887) Harvard Law Review, Samuel D. Warren e Louis D. Brandeis descrivono il concetto di privacy come il "diritto di essere lasciato solo". I due avvocati si occupavano di responsabilità civile e le recenti diffusioni di fotografia e quotidiani avevano scatenato non poche preoccupazioni riguardo l'abilità dell'individuo di poter celare la propria vita privata da intrusioni e diffamazioni. Nuove ondate di preoccupazione sorsero con la diffusione dei telefoni e delle tecnologie per le intercettazioni obbligando numerosi governi a espandere le leggi che ne regolavano l'utilizzo. Successivamente negli anni '70 i computer e lo sviluppo delle telecomunicazioni resero evidente la facilità ed economicità della raccolta di dati personali (Kang, 1998; Holvast, 2008), fattori che si sono acuiti nel tempo con la diffusione di Internet negli anni '90 e nuovamente nell'ultimo decennio con il "Web 2.0" e i social network (Acquisti, 2016).

Rielaborando la definizione di Altman e circoscrivendola al solo ambito dell'information privacy (che d'ora in poi verrà semplicemente chiamata privacy per semplicità), ai fini di questa tesi ci si riferirà a privacy intendendo "il controllo selettivo dell'accesso, elaborazione e utilizzo alle informazioni personali dell'individuo". Nonostante le questioni legate alla sorveglianza di massa e alla violazione della privacy da parte dei governi sia di grande rilevanza, questa tesi si concentrerà sui problemi e sulle opportunità legate al settore privato e digitale, perché è qui che più il designer può dare il suo apporto più concreto. Gli smartphone e i computer con cui ci si collega a internet sono infatti i principali punti di accesso alle informazioni e alla cultura nella società digitale odierna e le aziende che posseggono o controllano questi punti di accesso hanno un potere enorme sui modi in cui le persone esperiscono il Web. Sono numerose le associazioni, gli enti e gli attivisti che si concentrano proprio sul contrastare un sistema che offre come unica possibilità quella di sacrificare parte delle proprie libertà personali per poter accedere alle potenzialità di un mondo tecnologico e collegato alla rete. Il designer può usare le sue abilità di traduttore e mediatore per connettere il contesto, gli approcci di discipline diverse e l'utente finale per allargare il dibattito sulla privacy digitale alla partecipazione dell'opinione pubblica.

Non verranno trattate perciò ad esempio le carte fedeltà dei supermercati o il tracciamento degli acquisti tramite

carta di credito nei negozi, quanto piuttosto le transazioni servizio-dati quando si legge il giornale su Internet, quando si fa una ricerca con Google, quando si acquista su Amazon, quando di usa Facebook e Instagram, quando si scarica un'applicazione sullo smartphone.

Attraverso la tesi verranno utilizzate le parole persone, utenti, consumatori e cittadini interscambiabilmente di proposito, per sottolineare che sono tutte sfaccettature importantissime della stessa medaglia: come utenti di Internet è facile dimenticarsi che ciò che facciamo online ha ripercussioni anche nella vita reale e che, quando la risorsa essenziale che regge un'economia sono i dati personali, vengono influenzati non sono solo i propri diritti di consumatori ma di cittadini ed esseri umani.

APPROFONDIMENTO

L'EVOLUZIONE DEGLI STUDI SULLA PRIVACY

Durante la creazione di un vocabolario comune di termini, utile a poter affrontare con maggior chiarezza il tema principale di questa tesi, è emerso come sono moltissime le discipline che nel corso dell'ultimo secolo hanno affrontato da diversi angoli il discorso privacy. Analizzare quanto differenti campi di ricerca hanno esplorato questa tematica nel tempo e quali sono gli argomenti principali connessi ad essa potrebbe aiutare a collocare questa tesi in un contesto di ricerca, posizionandosi vicino a campi che trattano temi simili.

(→METODO: Compendio metodologico)

Per iniziare è necessario recuperare i paper o articoli scientifici che trattano di privacy. Una via possibile è quella di utilizzare un database di riassunti e citazioni per articoli di pubblicazioni riguardanti la ricerca. Sicuramente non conterrà l'intero corpus di tutti gli articoli mai scritti sulla privacy, ma se il database in questione è una fonte riconosciuta ed autorevole in campo accademico, potrebbe essere un'approssimazione accettabile. Scopus, fondato nel 2004 dalla casa editrice Elsevier[4], contiene più di 57 milioni di documenti scientifici appartenenti a quasi tutti i giornali e le riviste più rispettati ed influenti nella ricerca. Esso restituisce, data una chiave di ricerca, tutti i documenti che la contengono come parola chiave, lasciando inoltre esportare un file di testo in

[4] Scopus è il database online di riassunti e citazioni per articoli scientifici della Elsevier, maggior editore mondiale in campo medico e scientifico. *https://www.scopus.com*

formato .csv[5] che riporta varie statistiche come il numero di documenti pubblicati divisi per anno, per area di ricerca, per fonte, per affiliazione del ricercatore principale, nonché una lista delle parole chiave che più spesso vengono associate al termine ricercato. Filtrando una per volta le varie aree di ricerca, è possibile fare emergere l'importanza di tematiche specifiche, oltre all'importanza di fare ricerca sulla privacy nei diversi campi.

Tenendo ben a mente che Scopus non contiene tutti i paper mai scritti, che le categorie che assegnano il campo di ricerca non sono mutualmente esclusive (p. e.: un articolo può essere contemporaneamente nella categoria *engineering* e *computer science*) e che le parole chiave non sono sempre la rappresentazione accurata dei temi specifici trattati, è stato ritenuto comunque un buon proxy[6] per avere un'idea abbastanza rappresentativa, seppur distorta, dell'evoluzione di un discorso relativo alla privacy nell'ambiente scientifico.

Immesso il termine "*privacy*" nella barra di ricerca, Scopus ha trovato circa 83.000 documenti, dal 1984 ad oggi. Ci sono articoli, pubblicazioni per conferenze, lettere, capitoli di libri e varie altre tipologie di documento. È stato ritenuto più corretto mantenerle tutte per via delle diverse modalità con cui operano diversi campi accademici. Definite le aree che contavano almeno 2.000 pubblicazioni e aggregate nella categoria *altro* le rimanenti, è stata una per volta filtrata la ricerca ed esportato il file csv corrispondente in modo da poter comparare le categorie tra di loro.

→ FIG. 02

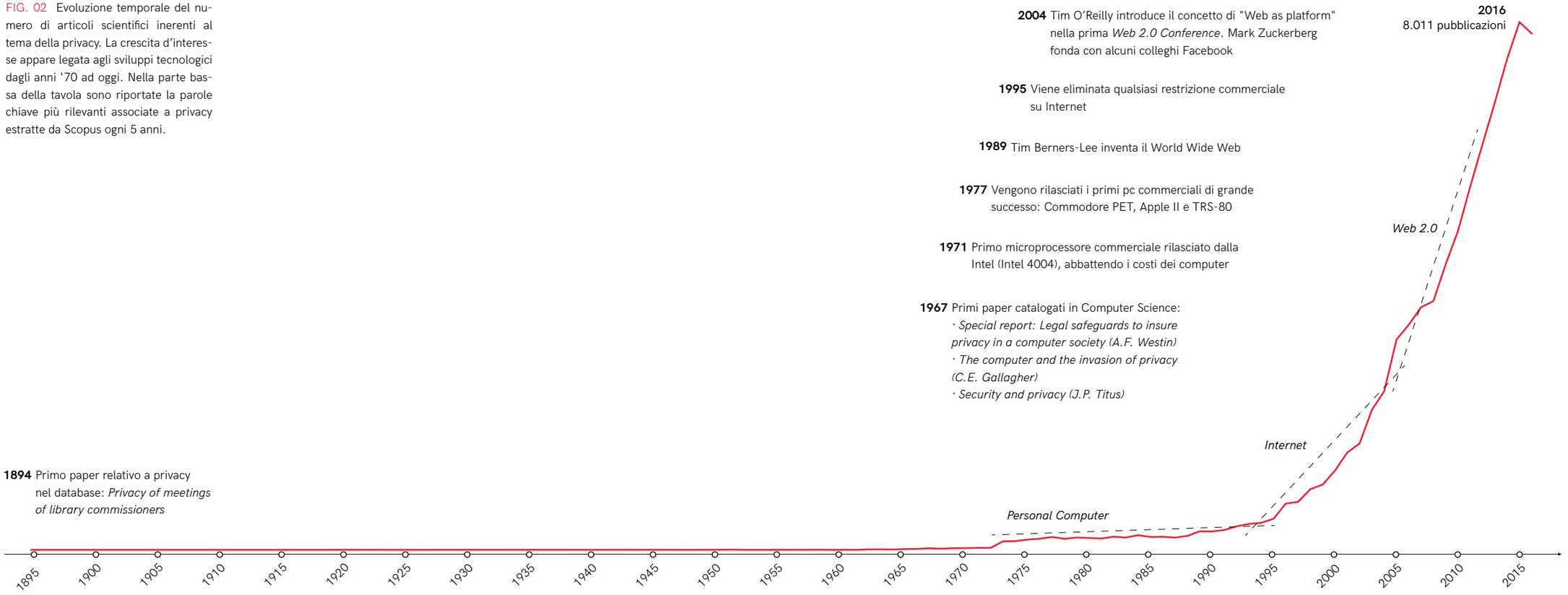
Possiamo fare subito alcune osservazioni. Seguendo l'evoluzione del numero di documenti totale, prima degli anni '70 la privacy era quasi assente nella ricerca, mentre da quel momento in poi segue una crescita esponenziale fino ad arrivare ai circa 8.000 documenti all'anno nel 2015-2016. Guardando le singole discipline, si nota come sia l'Informatica il maggior contribuente a questo incremento, arrivando nell'ultimo decennio a ricoprire più della metà delle pubblicazioni a discapito della Medicina che per 20 anni (1970-1990) era stata la principale fonte di ricerca. Anche Ingegneria e Matematica hanno aumentato l'interesse per la privacy nel tempo, anche se probabilmente molto spesso legate all'Informatica in campi come la crittografia e gli studi sugli algoritmi. Tutti gli altri campi sembrano invece abbastanza costanti, in termini di percentuale sul totale, nel tempo.

La dicotomia tra i due periodi temporali pre e post-computer è visibile anche guardando le parole chiave assegnate agli articoli. Filtrando infatti i risultati a intervalli di cinque anni emerge come, mentre fino a metà degli anni '90 molte delle tematiche sono in relazione ad aspetti legali, entrando nel nuovo millennio si inizia a parlare di *data privacy*, di si-

[5] Acronimo di *comma-separated values*, è uno dei formati di testo più comunemente usati per manipolare dati tabulari.

[6] Un proxy è un server che funge da intermediario tra il proprio computer e il resto dell'Internet. *È come mandare qualcuno ad un'asta: sei un famoso miliardario che non vuole farsi vedere di persona, ma quel vaso Ming ti piace proprio tanto. Quindi mandi una persona fidata per scommettere in tua vece. Tu mantieni l'anonimato ma ottieni ciò che volevi (fonte: Sidway Dictionary).*

FIG. 02 Evoluzione temporale del numero di articoli scientifici inerenti al tema della privacy. La crescita d'interesse appare legata agli sviluppi tecnologici dagli anni '70 ad oggi. Nella parte bassa della tavola sono riportate la parole chiave più rilevanti associate a privacy estratte da Scopus ogni 5 anni.



1970	1975	1980	1985	1990	1995	2000	2005	2010	2015
Human	Privacy	Privacy	United States	Privacy	Human	Privacy	Data Privacy	Privacy	Data Privacy
Adult	Jurisprudence	Human Rights	Privacy	United States	Privacy	Humans	Security Of Data	Data Privacy	Privacy
Female	Human Rights	United States	Human Rights	Human Rights	United States	Human	Privacy	Human	Cryptography
Male	Civil Rights	Jurisprudence	Human	Legal Approach	Confidentiality	United States	Human	Humans	Human
Age	Informed Consent	Civil Rights	Civil Rights	Civil Rights	Humans	Data Privacy	Humans	Network Security	Network Security
Age Factors	Legal Rights	Confidentiality	Legal Approach	Human	Legal Aspect	Security Of Data	Cryptography	Cryptography	Security Of Data
Fitness	Legislation	Human	Confidentiality	Jurisprudence	Legal Approach	Confidentiality	Internet	Privacy Preserving	Security
Legislation	Government	Legal Aspect	Jurisprudence	Legal Rights	Female	Internet	Network Protocols	Authentication	Authentication
Middle Age	State Interest	Legal Rights	Government	Legal Aspect	Genetics	Legal Aspect	Confidentiality	Internet	Mobile Security
Physical Fitness	Abortion, Induced	Informed Consent	Legal Aspect	Genetics	Civil Rights	Medical Record	United States	Security	Privacy Preserving

1.
Computer Science
Data Privacy
Privacy
Cryptography
Security Of Data
Network Security
Privacy Preserving
Security
Authentication
Internet
Algorithms

2.
Medicine
Human
Privacy
Article
Humans
Confidentiality
United States
Female
Priority Journal
Male
Informed Consent

3.
Engineering
Data Privacy
Privacy
Cryptography
Security Of Data
Security
Authentication
Algorithms
Network Security
Internet
Privacy Preserving

4.
Social Sciences
Privacy
Human
Article
Humans
Data Privacy
United States
Confidentiality
Ethics
Female
Internet

5.
Arts & Humanities
Privacy
Human
Humans
Article
Ethics
Confidentiality
Data Privacy
Female
United States
Informed Consent

6.
Mathematics
Data Privacy
Cryptography
Security Of Data
Privacy
Network Security
Privacy Preserving
Authentication
Algorithms
Artificial Intelligence
Security

7.
Business
Privacy
Data Privacy
Electronic Commerce
Security Of Data
Internet
Laws And Legislation
Data Protection
Security
Trust
Information Management

8.
Biochemestry
Human
Privacy
Humans
Article
Priority Journal
Data Privacy
Confidentiality
Genetic Privacy
Review
Female

9.
Decision Sciences
Data Privacy
Privacy
Security Of Data
Cryptography
Information Systems
Security
Privacy Preserving
Internet
Information Management
Electronic Commerce

10.
Nursery
Human
Humans
Article
Privacy
Confidentiality
United States
Female
Ethics
Male
Legal Aspect

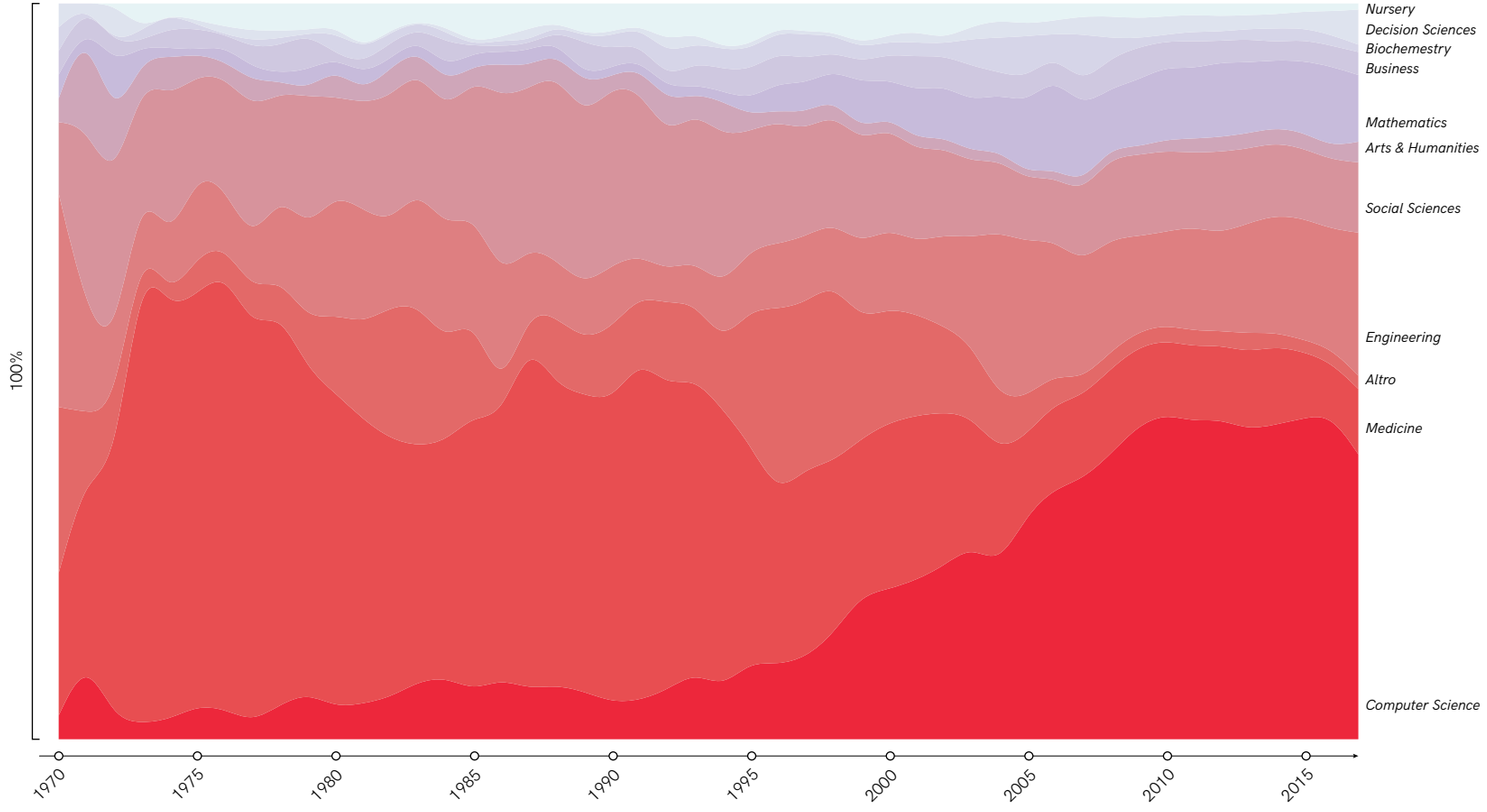


FIG. 03 Evoluzione temporale della ripartizione per disciplina degli articoli scientifici inerenti al tema privacy (in percentuale). Nella parte bassa della tavola sono riportate le parole chiave più rilevanti per ogni area.

curezza e di crittografia. Questo dipende ovviamente dall'esplosione di paper scientifici in Informatica, perciò per vedere le differenze tematiche dobbiamo filtrare le parole chiave per disciplina.

Informatica, Ingegneria e Matematica condividono molti temi, soprattutto sull'information privacy, sulla sicurezza informatica e sugli algoritmi. Parallelamente Medicina, Biochimica e le discipline umanistiche hanno in comune termini più relativi alla sfera umana, tra cui l'etica e la riservatezza. È interessante notare come la parola *"etica"* non sia presente tra le tematiche più trattate dalle cosiddette STEM[7].

Anche se il Design come disciplina è ancora praticamente assente nel panorama di ricerca che si occupa di studiare la privacy nelle sue molteplici sfaccettature, potrebbe diventare un ponte utile ad accorciare il divario tra diversi ambiti grazie alla sua propensione a riuscire a connettere aree tematiche, contesti di utilizzo e utenti tra loro (Baule & Caratti, 2016). Ad esempio progettando soluzioni basate sulle conoscenze tecniche dell'Informatica, fondamentali per il contesto di azione odierno, ma tenendo in considerazioni tutte le implicazioni etiche analizzate dalle scienze sociali.

7 ▲ Questa sigla - Science, Technology, Engineering and Mathematics - indica i corsi di studio volte a incrementare la competitività in campo scientifico e tecnologico.

1.2 Definizione dell'intorno

Parallelamente alla ricerca bibliografica, fondamentale per creare una solida base di conoscenza e per delineare lo stato attuale della ricerca accademica riguardante il tema, è utile anche analizzare cosa viene detto al di fuori di essa. Fare ciò potrebbe rivelare nuove prospettive o svelare delle dicotomie tra il dibattito accademico e il discorso pubblico.

Già diversi autori hanno dimostrato come il Web sia uno specchio utile a studiare la realtà. È uno specchio parziale e incompleto, che può mostrare solo parte della realtà, ma che allo stesso tempo se usato ricordando i suoi limiti permette di fare affermazioni fondate sul mondo reale (Mauri, 2015).

Una volta intrapresa la strada dello studio del web, ci si scontra subito con una nuova serie di problemi: dov'è il luogo (virtuale) migliore per iniziare la nostra ricerca? Quale sarà la dimensione temporale da scegliere? Come si può raccogliere dati archiviati e formattati diversamente in varie parti del web, su un medium così instabile e mutevole? Queste e molte altre domande metodologiche e tecniche impongono di dover progettare dei metodi che possano adattarsi alla "irregolarità" del web. Richard Rogers, precursore di questa tipologia di analisi sociologica, li ha definiti *Digital Methods* (metodi digitali): invece di adattare

il web ai metodi di ricerca tradizionali, essi sfruttano la volatilità del medium a vantaggio dell'analisi (Mauri, 2015).

Si può pensare al web come composto da una serie di *device*, o dispositivi. Google, Facebook, Reddit, Youtube, Instagram, etc. sono dispositivi con le loro peculiari caratteristiche e funzioni, ed ognuno di essi fornisce una serie di "oggetti digitali" che possono tornare utili per studiarli (Rogers, 2013). Link, post, tag, profili, sono tutti esempi di oggetti digitali. Ogni Digital Method serve a riadattare gli oggetti digitali perché siano utilizzabili per la ricerca.

Per studiare le tematiche che ruotano intorno alla privacy è necessario quindi definire quali device possano aiutare a fare affermazioni lecite sul mondo reale e di conseguenza quali metodologie utilizzare, riadattare o creare per esplorarli. Nello specifico, analizzare device che trattano di privacy a diversi livelli di specificità potrebbe rivelare delle modifiche nelle tematiche trattate man mano che ci si muove da un ambiente più generale ad uno più "addetto ai lavori". Si è perciò deciso di analizzare Wikipedia, Medium e ArsTechnica. Wikipedia rappresenta la conoscenza collettiva, l'enciclopedia generale dove è facile approcciarsi ad un argomento sconosciuto; Medium è una piattaforma di blogger in cui persone più o meno legate al tema della privacy decidono esprimere la loro opinione, offrendo così la possibilità di studiare gli attori in qualche modo attivi nel discorso; infine ArsTechnica è un magazine online che si occupa di tecnologia e del suo legame con la società, ospitando quindi giornalisti specializzati sul discorso dell'information privacy. Ovviamente non bisogna dimenticare che ogni piattaforma ha i suoi limiti e le sue faziosità, che verranno man mano sottolineate durante l'analisi.

WIKIPEDIA

(→METODO: Compendio metodologico)

Wikipedia, essendo un'enciclopedia online, può essere utilizzata per studiare come diverse tematiche, ad un primo livello generale, sono connesse tra loro. Proprio come una normale enciclopedia, l'obiettivo è quello di introdurre in modo il più completo ma breve possibile lo scibile, perciò dovrebbe connettere le tematiche chiave intorno al termine cercato per restituire una spolverata generale di esso. A differenza di una enciclopedia cartacea però, i temi sono connessi tra loro da link, oggetti digitali su cui possiamo fare leva per studiare le relazioni tra argomenti. Quasi ogni pagina possiede infatti una sezione chiamata "Voci correlate", in inglese *See Also* (il numero di pagine che non contiene questa sezione è minimo e quindi trascurabile come errore), dove gli autori di una data pagina posizionano tutte le tematiche che sono abbastanza vicine a quella pagina ma che meritano una spiegazione separata. Ciò permette di creare una sorta di mappatura tematica di ciò che è ritenu-

to importante dall'opinione pubblica connesso alla privacy.

Con un'ulteriore analisi, si può anche vedere come questa mappatura cambi in diverse aree linguistiche, studiando le voci correlate in ogni versione della pagina ma in diverse lingue. Questo non sarà esattamente sovrapponibile ad un'area geografica precisa, però darà comunque un'approssimazione accettabile di cosa interessa a diverse aree del globo. L'approssimazione sarà più vicina in lingue più localizzate come l'italiano o il tedesco e un po' meno in inglese che è più diffuso, anche se è possibile ritenerlo comunque accettabile perché nazioni come Stati Uniti, Regno Unito e Australia sono molto affini, soprattutto sulle politiche sulla privacy.

Il Digital Method scelto è *Seealsology*[8], sviluppato da DensityDesign[9] in collaborazione con médialab SciencePo[10], creato appositamente per generare una rete di tematiche a partire da una o più pagine di Wikipedia. Per quanto riguarda le lingue, sono state scelte inglese, italiano, francese e tedesco perché nel mondo occidentale rappresentano le nazioni che esercito la maggior influenza a livello politico e tutte erano analizzabili con Seealsology (per ragioni tecniche oltre che pratiche, Seealsology non supporta tutte le lingue esistenti su Wikipedia).

Analizzate le voci correlate di ogni lingua, l'output sono delle reti dove ogni nodo è una pagina Wikipedia ed ogni connessione è una voce correlata che lega insieme due pagine. La grandezza del nodo indica il numero di altre pagine che citano quel nodo, quindi più è grande più pagine lo contengono nella sezione "Voci correlate".

→ FIG. 04

Come ci si poteva aspettare l'inglese è la lingua dove il discorso è più esteso: certamente in seguito alle rivelazioni del 2013 di Edward Snowden i due temi "privacy" e "sorveglianza" si sono strettamente legati, tanto che anzi quest'ultimo sembra quasi prevalere con un insieme di nazioni con programmi di sorveglianza di massa, nomi di *whistleblowers*[11] e di programmi di spionaggio. Vi sono poi altri cluster rilevanti, anche se secondari rispetto al discorso sorveglianza, che descrivono dei temi legislativo-politici ("*personality rights*", "*privacy law*" e nomi di associazioni per i diritti digitali come l'EFF e Privacy International), tecnologici ("*privacy software*", "*authentication*", "*I2P*") e di lotta alla criminalità ("*identity theft*", "*phishing*").

Un nodo molto interessante è "*Web literacy*", probabilmente indice di un discorso abbastanza maturo anche nell'opinione pubblica riguardo all'importanza di approcciare criticamente il Web oltre a conoscerne le caratteristiche. Questa pagina non esiste in nessuna altra lingua.

8 ■ *http://tools.medialab.sciences-po.fr/seealsology*

9 ▲ Laboratorio di ricerca del Dipartimento di Design del Politecnico di Milano.
http://www.densitydesign.org

10 ▲ Laboratorio di ricerca dell'università parigina Sciences Po.
http://www.medialab.sciences-po.fr

11 ● Individui che denuncino pubblicamente o riferiscano alle autorità attività illecite o fraudolente all'interno del governo, di un'organizzazione pubblica o privata o di un'azienda. In italiano significa letteralmente "soffiatore di fischietto".

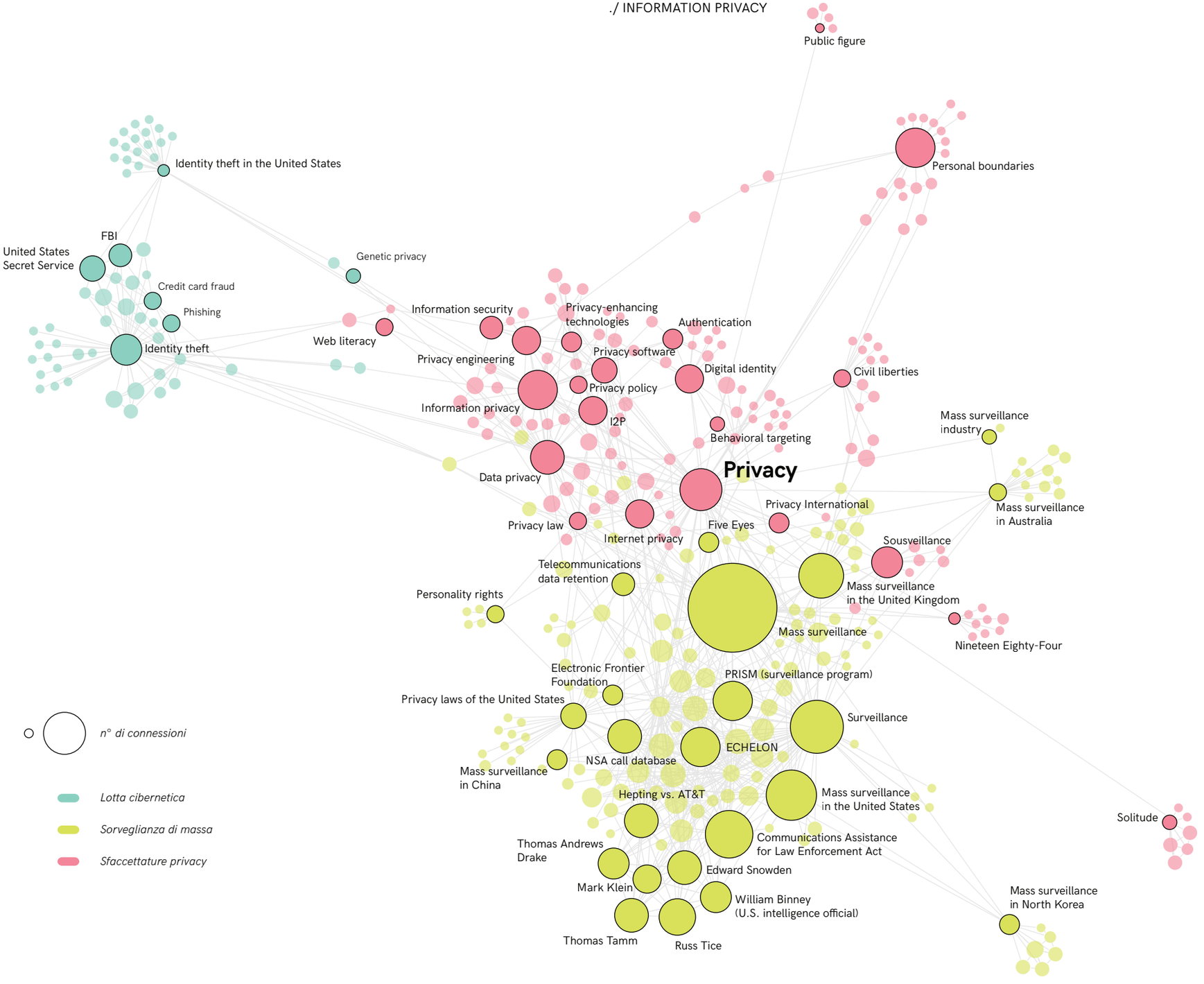


FIG. 04 Mappa di correlazione tra pagine di Wikipedia, versione inglese. Ogni nodo è un articolo di Wikipedia e due pagine sono collegate tra loro se una cita l'altra nella sezione "Voci correlate".

La rete francofona è molto interessante: anch'essa connette i due temi "*vie privée*" e "*surveillance globale*" ma con meno forza rispetto alla rete anglofona. Anche se è possibile che l'influenza della tematica della sorveglianza di massa sia dovuta ad autori canadesi per via della vicinanza con gli Stati Uniti, certamente molto del discorso ruota intorno alla Francia ed alla situazione europea. Tra i nodi si può infatti trovare varie commissioni, regolamenti e leggi francesi o europee volte a difendere i diritti della privacy dei cittadini. Inizia a delinearsi una dicotomia che sarà ritrovata anche più avanti tra la diversa attitudine dei governi appartenenti all'Unione Europea e quello statunitense nel regolamentare la privacy. Leggi generali da una parte e auto-regolamentazione dell'industria dall'altra.

Ci sono inoltre due tematiche che troviamo solo qui. Il primo è la "fuga di informazioni" ("*fuite d'information*"), nodo che fa da ponte tra la sorveglianza di massa e la sicurezza informatica ("*sécurité de l'information*"), tema che nella lingua anglofona non sembrava assolutamente rilevante ma qui è molto influente. Il secondo è l'etica ("*Éthique*"), una prospettiva che dovrà necessariamente diventare centrale in futuro se si vogliono trovare seriamente delle soluzioni che tengano conto delle singole persone piuttosto che dei meri vantaggi economici.

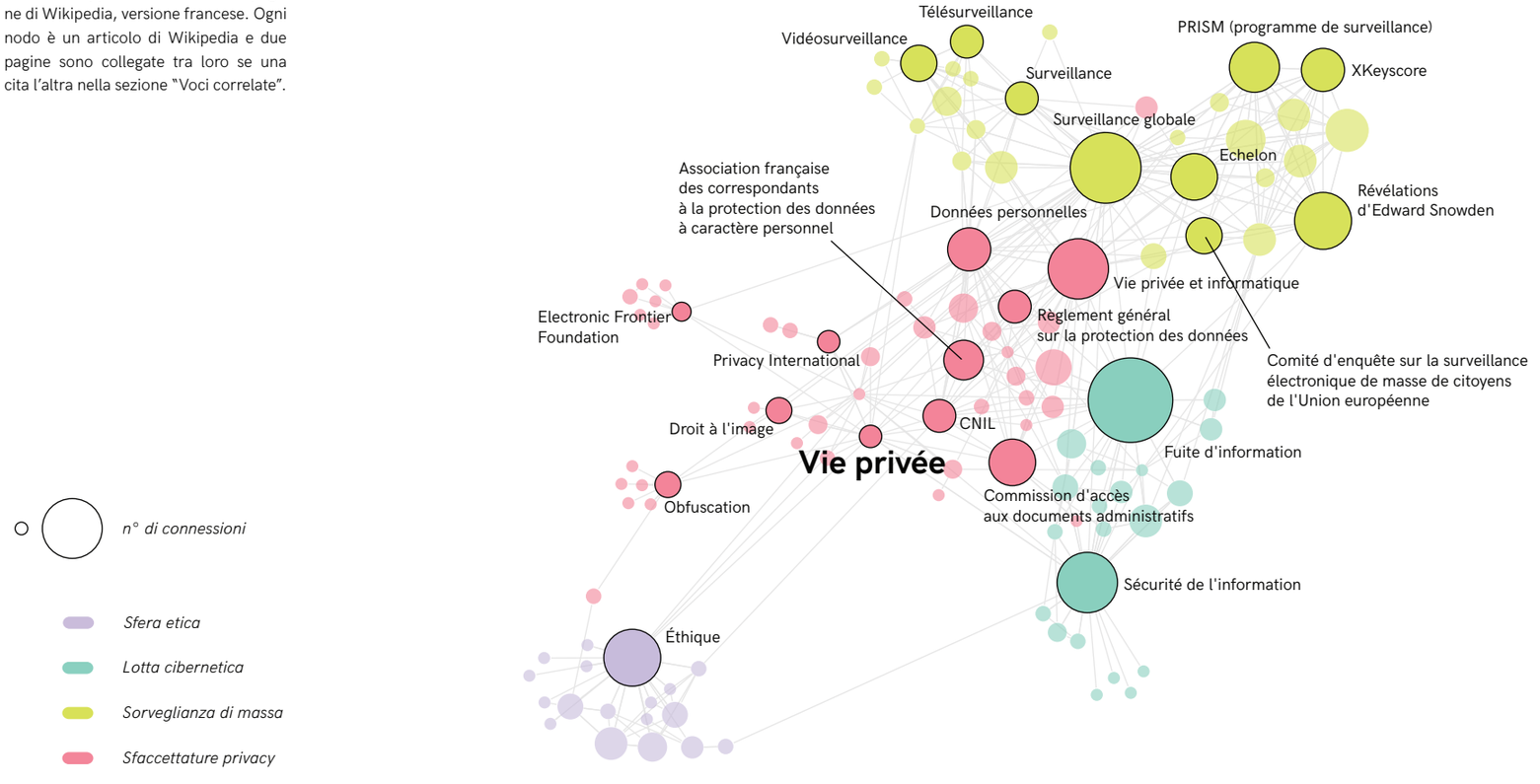


FIG. 05 Mappa di correlazione tra pagine di Wikipedia, versione francese. Ogni nodo è un articolo di Wikipedia e due pagine sono collegate tra loro se una cita l'altra nella sezione "Voci correlate".

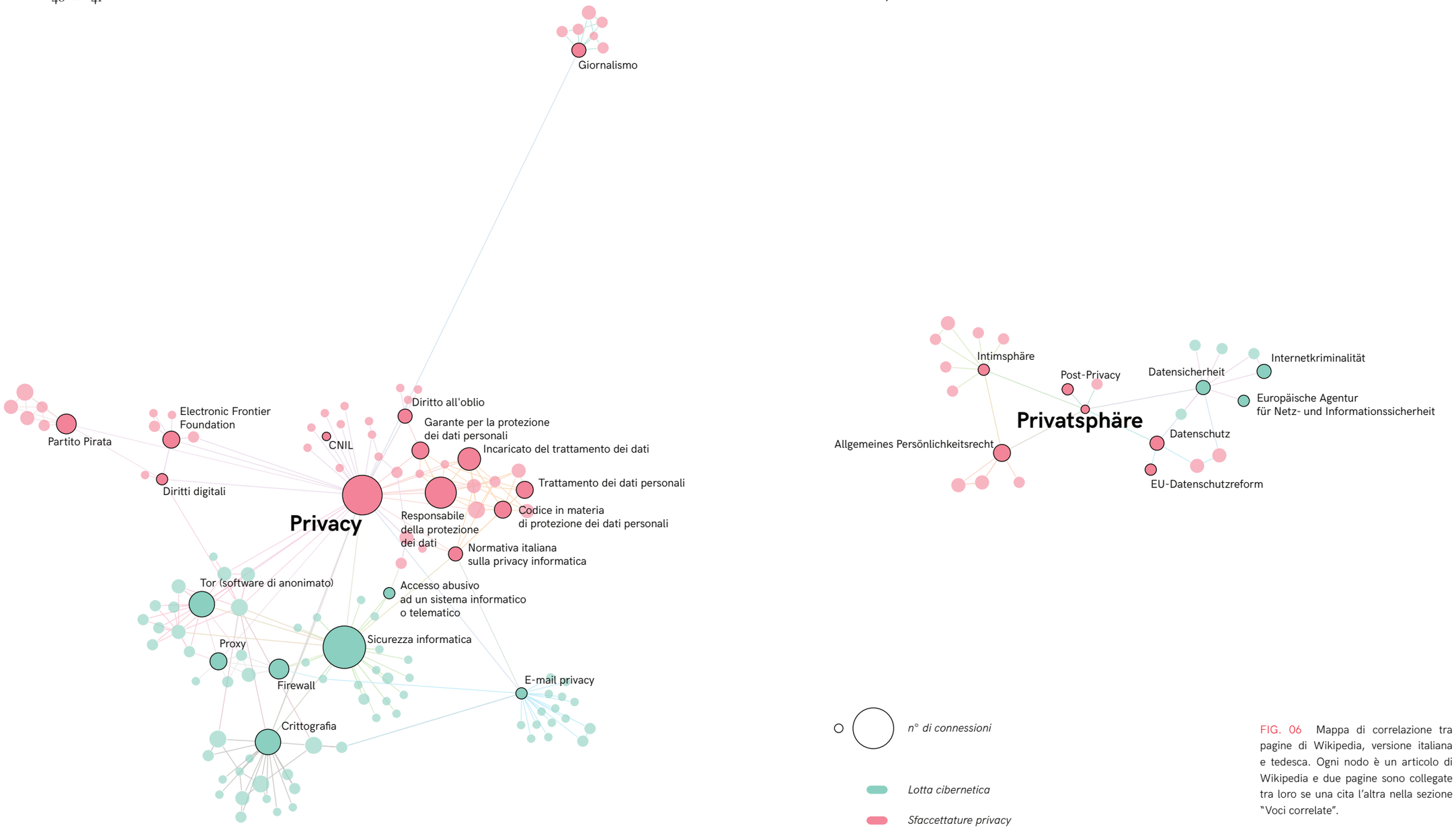


FIG. 06 Mappa di correlazione tra pagine di Wikipedia, versione italiana e tedesca. Ogni nodo è un articolo di Wikipedia e due pagine sono collegate tra loro se una cita l'altra nella sezione "Voci correlate".

Le altre due reti, quella italiana e tedesca, sono assai più povere e le tematiche connesse meno varie. Manca completamente il discorso sulla sorveglianza di massa e degli sforzi per la trasparenza di persone come Snowden, o perlomeno non sono in alcun modo legati alla privacy (le pagine sono infatti presenti in entrambe le lingue, ma sembrano non avere alcuna connessione). È come se in queste sfere linguistiche il dibattito fosse ancora in uno stadio embrionale: vengono citate le leggi europee e il tema della sicurezza informatica ma il discorso italiano sembra fissato solo su questioni tecnologiche ("*Tor*", "*proxy*", "*firewall*", "*crittografia*") mentre quello tedesco sul concetto di identità e sfera personale ("*intimsphäre*", "*allgemeines persönlichkeitsrecht*").

È quindi possibile concludere che c'è un discorso abbastanza sviluppato intorno alla privacy e che, come ci si poteva aspettare, esso è sia legato al rapporto tra tecnologia e legislazione, sia influenzato dagli avvenimenti d'attualità, in particolare sulla sorveglianza di massa. Questi avvenimenti però sembrano avere effetto solo localmente.

L'opportunità, vista con l'occhio del designer della comunicazione, potrebbe essere quella di lavorare insieme alle associazioni di diritti digitali sulla consapevolezza dell'opinione pubblica, al fine di far maturare un dibattito più informato e costruttivo.

MEDIUM

Medium è una piattaforma di pubblicazione online nata nel 2012 dove opinionisti di ogni campo e professione scrivono degli argomenti che stanno loro più a cuore. Su Medium chiunque può scrivere gratuitamente ma la piattaforma ha accresciuto molto la sua visibilità negli anni tanto che molto spesso a scrivere si possono trovare rinomati professionisti di molti settori tra cui design, tecnologia, economia e politica. Insieme a Twitter è forse il device più consono alla ricerca di opinioni su un tema (almeno in termini di rilevanza) ma a differenza di quest'ultimo gli utenti non sono limitati a stare in 140 caratteri e in più Medium è proprio specializzato nell'essere una sorta di archivio delle opinioni. L'ipotesi che muove l'analisi è che qui ci sia ancora l'eco di tutti i temi emersi nello studio di Wikipedia, ma che se ne aggiungano altri più specifici e che magari l'opinione pubblica non abbia ancora colto.

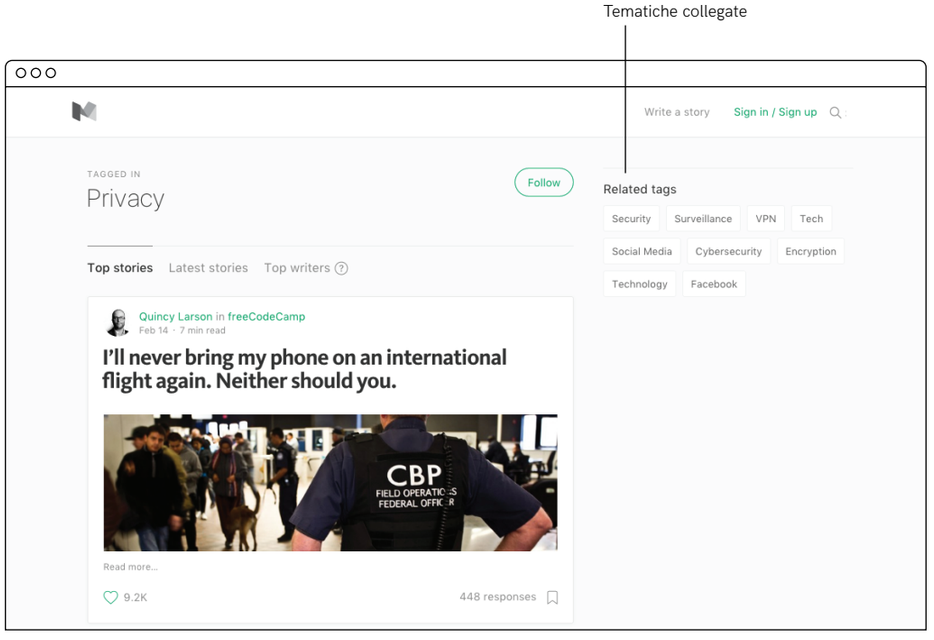
(→METODO: Compendio metodologico)

Subito sorge un ostacolo: mancano strumenti adatti ad analizzare la piattaforma e metodi digitali specifici (oppure sono estremamente difficili da trovare). Il primo passo è stato perciò di prendere in prestito metodi digitali simili e riadattarli per sfruttare le specificità di Medium. Ogni articolo viene taggato dall'autore con delle parole chiave che riassumono il senso del pezzo; queste tag[12] vengono messe in una sezione apposta in fondo all'articolo e tutti le

[12]• Termine associato ad un'informazione, che ne facilita la classificazione. *Sono come le etichette messe sugli scatoloni nei traslochi, che permettono di capire dove sono andati a finire i calzini invernali.*

mettono perché sono anche il sistema con cui l'algoritmo di Medium categorizza il corpus di articoli e gestisce il motore di ricerca interno. Inoltre se si cerca una parola chiave nella barra di ricerca della *Home page*, Medium restituirà gli articoli più rilevanti legati a quella determinata parola chiave (con un algoritmo abbastanza oscuro, probabilmente in base a fattori come il numero di commenti, di raccomandazioni, se chi scrive è affiliato ad una rivista/organizzazione importante, etc.) suggerendo anche una serie di tag correlati. La scelta di questi è anch'essa oscura come anche il numero di articoli restituiti, non essendo sicuramente la totalità di essi. Nonostante i suddetti limiti, si può presumere che le parole chiave siano quelle che compaiono di più insieme alla query cercata e che gli articoli siano filtrati per rilevanza e siano un buon filtro per analizzare le tematiche collegate alla privacy.

È stato creato quindi un tool su misura che estraesse, data una chiave di partenza e una profondità di ricerca, tutti i termini che creano il contorno di quella chiave, fino alla profondità definita. In questo caso, le parole correlate a "*privacy*", fino al secondo livello di profondità.



IMG. 01 Schermata di Medium che mostra la posizione delle tematiche correlate.

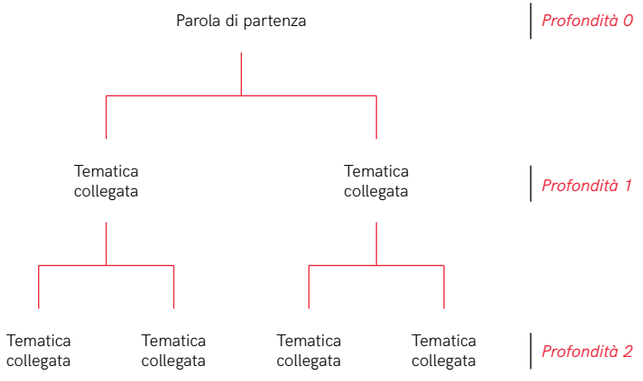


FIG. 07 Schema che illustra il principio delle tematiche correlate e della profondità di ricerca.

I primi risultati non sono esattamente quelli sperati: le parole associate a privacy sono molto generiche e non aiutano molto ad inquadrare il dibattito. Si nota una forte predominanza di un discorso tecnologico in cui risaltano, oltre ovviamente a "*Tech*" e "*Technology*", colossi digitali come Apple e i vari social media oppure parole legate alla sicurezza informatica. Non c'è però quasi nessun altro tema specifico. Una possibile spiegazione risiede nel luogo in cui sono state cercate le parole chiave. Prese infatti dalla pagina generale, è probabile che mostrino solo le parole che compaiono nella maggioranza degli articoli, finendo così per riportare solo i tag più generici. Per far emergere le tematiche specifiche degli articoli è stato necessario cambiare approccio.

Creando un nuovo tool, sono state estratte le tag scritte alla fine di ogni articolo, prendendo tutti i risultati caricati dalla pagina di Medium cercando la parola "*privacy*". Con le tag estratte possiamo analizzare tutte le parole chiave associate a privacy negli articoli più rilevanti di Medium (e quindi quelli che un qualsiasi lettore andrà a leggere aprendo la piattaforma) ma anche con che frequenza due tag, e quindi due tematiche, appaiono assieme quando sono legate all'argomento privacy.

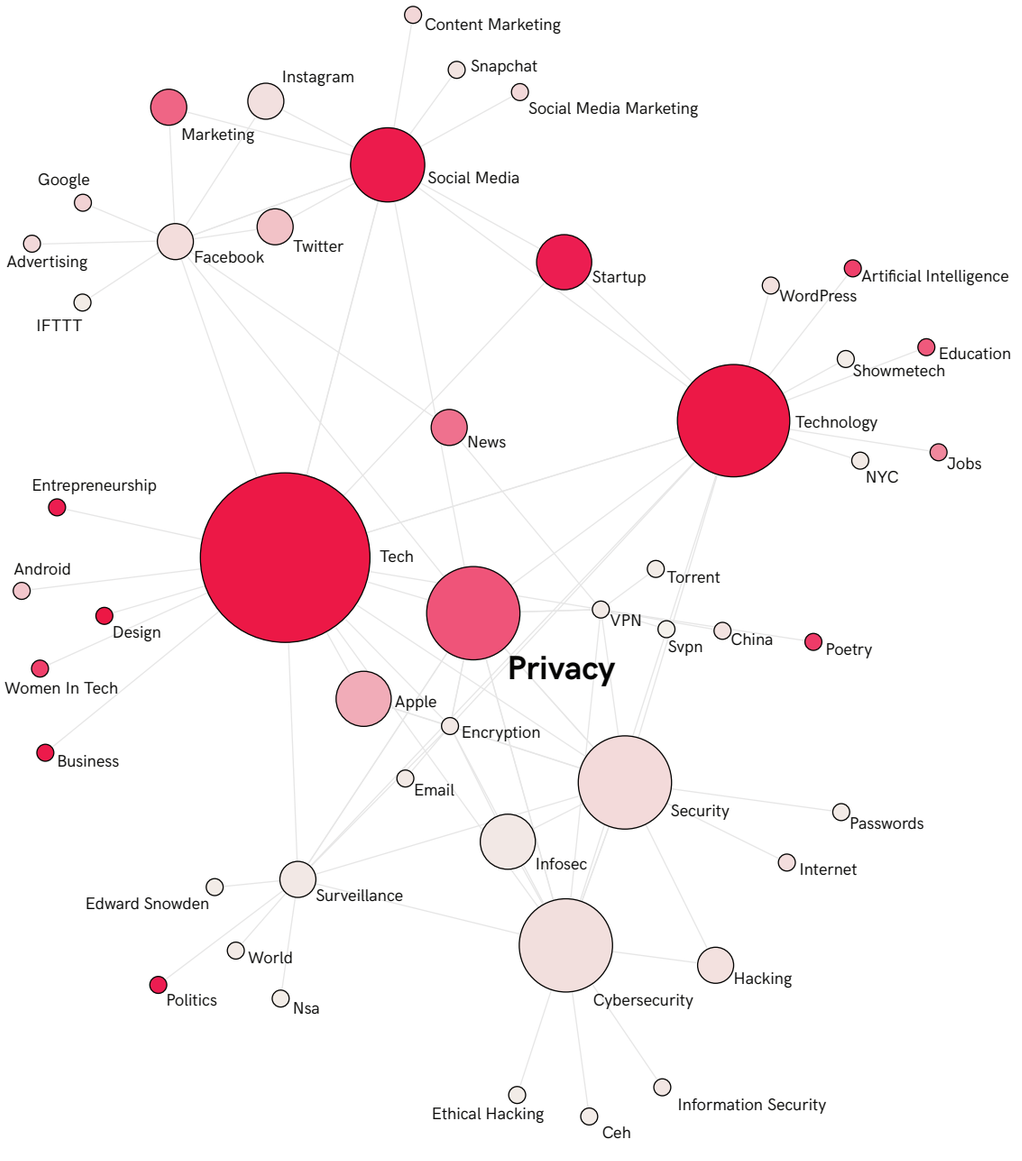


FIG. 08 Rete che mostra le parole chiave associate a privacy su Medium, prendendole dalla sezione "related tags". L'intensità di colore rappresenta la rilevanza del tema sulla piattaforma.

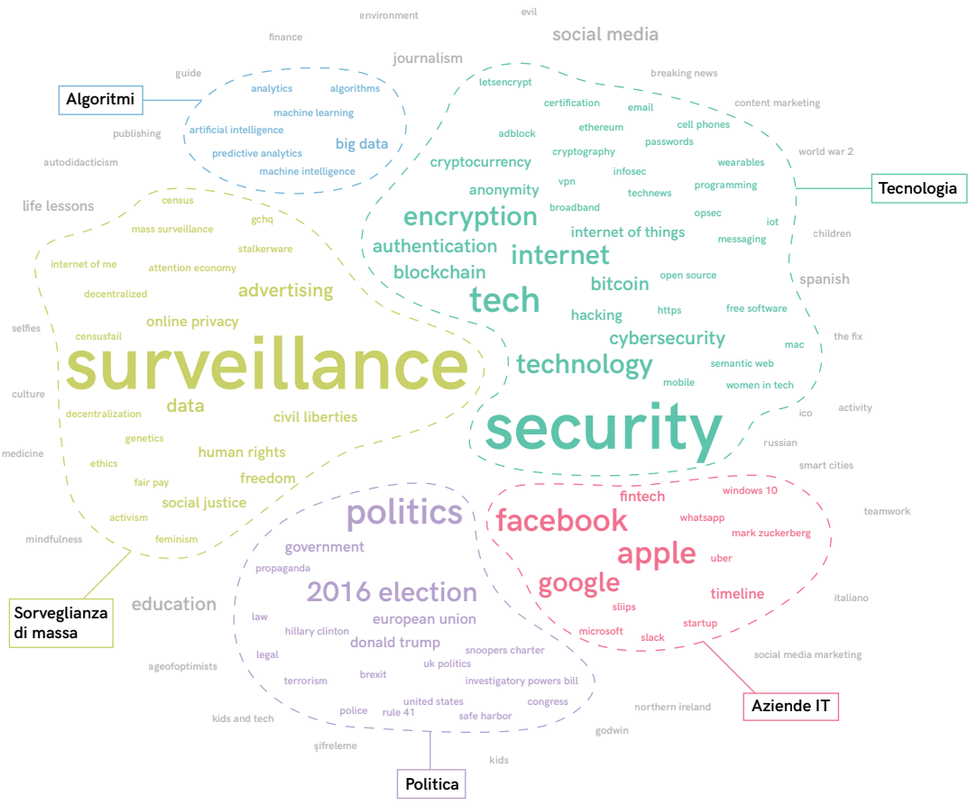


FIG. 09 Nuvola di parole chiave associate ai primi 100 articoli legati al tema privacy su Medium. La grandezza del carattere rappresenta la frequenza di apparizione negli articoli e il colore raggruppa le parole per categoria.

Visualizzando tutte le parole chiave trovate negli articoli, escono questa volta dei gruppi di parole interessanti. Il cluster sulla sorveglianza di massa è ancora il più importante ma ora più della metà delle tag è in qualche modo legato ad un discorso tecnologico, in varie sfumature come sicurezza informatica, tool per proteggere la propria privacy o nomi di grandi aziende digitali. Un altro gruppo che ha guadagnato influenza rispetto all'analisi di Wikipedia è quello politico: risalendo agli articoli con quelle parole chiave, si notano le paure collegate ad un aumento dei poteri dati ai servizi segreti con l'elezione di Donald Trump e la disputa legale tra il governo americano e Apple, il che spiegherebbe anche la dimensione della tag legata all'azienda di Cupertino.

Appare un cluster di parole che pur essendo piccolo mostra una sfaccettatura del discorso che non era ancora affiorata. "*Algorithms*", "*machine learning*", "*artificial intelligence*", etc. indicano un legame tra privacy e il crescente uso di algoritmi nei servizi che utilizziamo quotidianamente, una relazione che verrà indagata in maggior dettaglio più avanti.

Il secondo passo è visualizzare quanto spesso due o più termini compaiono assieme, con la supposizione che maggiore sia la frequenza, maggiore sia la vicinanza nel dibattito tra due tematiche. Essendo Medium una piattaforma basata sulle opinioni, ci si aspetta che il legame tra tag segua molto i fatti di attualità rispetto a Wikipedia, dando più rilevanza a ciò che viene discusso recentemente. Nella rete ogni nodo rappresenta una parola chiave, con grandezza differente in base a quante volte compare nei vari articoli. Ogni nodo è collegato ad altri ogni volta che questi compaiono assieme in un articolo, con lo spessore della connessione a indicare la frequenza con cui questo avviene.

→ FIG. 11

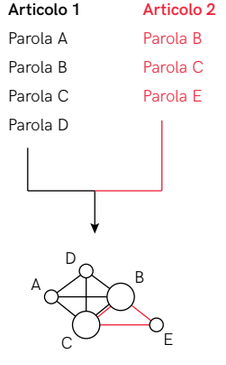


FIG. 10 Mappare visualmente la co-occorrenza tra parole chiave su diversi articoli attraverso una rete.

Seguendo l'ossatura della rete si nota, come previsto, un riflesso dei maggiori avvenimenti che hanno in qualche modo acceso il dibattito intorno alla privacy: la diatriba legale tra il governo statunitense e Apple riguardo ai fatti di San Bernardino oppure i rischi legati alla creazione di una "bolla di notizie" su Facebook e la sua influenza sulle elezioni americane. In particolare proprio quest'ultimi hanno stimolato molto la riflessione sul ruolo degli algoritmi nel quotidiano (oltre ovviamente a tutto ciò che riguarda le fake news, che non hanno però legami con la privacy), anche se nella rete non è presente un legame diretto con le tag "*machine learning*" o "*algorithms*", le quali appaiono separate dal resto del corpus.

L'analisi su Medium ha sottolineato lo stretto rapporto tra privacy e tecnologia ed ha fatto emergere, anche se flebilmente, la tematica dell'intelligenza artificiale che apparentemente può sembrare slegata da un dibattito incentrato sui dati personali ma che al contrario è fondamentale in quanto la funzione di questi algoritmi è proprio quella di elaborare enormi quantità di informazioni sulle persone e prendere delle decisioni partendo proprio da esse.

Medium si è rilevato un device utile ad analizzare la forma del dibattito attuale intorno ad una tematica, tenendo però conto dei suoi limiti di rappresentatività. Essendo

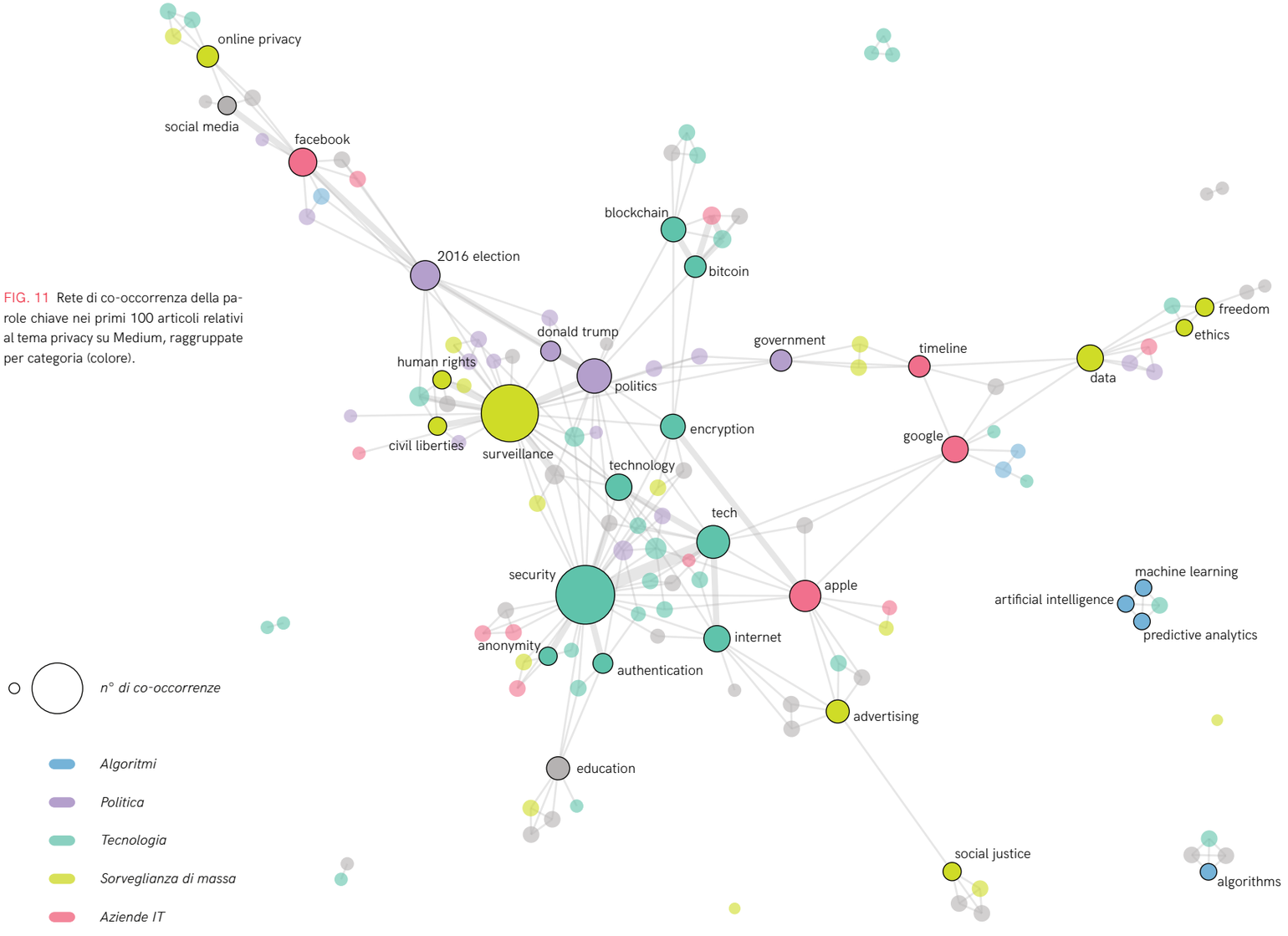


FIG. 11 Rete di co-occorrenza della parole chiave nei primi 100 articoli relativi al tema privacy su Medium, raggruppate per categoria (colore).

molti degli scrittori ritenuti da Medium "rilevanti" abitanti degli Stati Uniti, il punto di vista è concentrato particolarmente in quella direzione. Nel caso della privacy, dove il discorso è certamente più maturo che in altri Paesi come si è potuto constatare anche nell'analisi di Wikipedia, è un limite accettabile.

Nella creazione degli strumenti per studiarlo è affiorato inoltre un loro grande ostacolo, dovuto alla volatilità specifica di Medium: essendo una piattaforma in continua evoluzione, è stato necessario riscrivere più volte il codice con le logiche di estrazione delle informazioni perché la struttura del sito si modificava leggermente nel tempo. Ai fini della riproducibilità è perciò essenziale che il tool sia tenuto sempre aggiornato (in alternativa un programmatore più esperto potrebbe cambiare le logiche per rendere l'estrazione più stabile). L'importanza di "seguire il medium" (Rogers, 2013) viene ribadita qui più che mai.

ARS TECHNICA

ArsTechnica è un sito d'informazione tecnologica attivo dal 1998, che tratta tutto ciò che è legato al mondo di hardware e software ma in particolare ha una sezione specificatamente dedicata alle *technology policy*, quindi al legame tra regolamentazione e tecnologia. Visto che il discorso sulla privacy è strettamente legato ai due temi, un'analisi

(→METODO: Compendio metodologico)

di una piattaforma specializzata di questo genere poteva aggiungere nuove prospettive di ricerca.

Il primo approccio è stato quello di estrarre i primi 100 articoli per rilevanza presenti nella sezione "policy" di ArsTechnica con l'ipotesi che anche senza cercare direttamente per pezzi relativi al tema privacy, esso sarebbe emerso autonomamente per la sua importanza. Ancora una volta, come per Medium, si può supporre che le parole chiave associate ad ogni articolo possano essere un'approssimazione sufficiente delle tematiche trattate al suo interno. Questa volta però esse non sono riportate da nessuna parte, se non in alcune variabili nascoste nel codice che però non vengono mai visualizzate. In particolare sono presenti due tipi di parole chiave: un gruppo inserito a fini commerciali, cioè per assegnare ogni articolo ad una categoria per le agenzie pubblicitarie; un gruppo deciso dall'autore del pezzo, con parole non standardizzate, forse create per una veloce categorizzazione utile all'editore. Mentre le prime sono molto generiche e vaghe, le parole chiave inserite dall'autore danno una sintesi efficace del contenuto dell'articolo, nonostante siano senza uno standard definito.

Dopo avere creato uno strumento per estrarre le tag dagli articoli, sono state rappresentate in una rete, sempre per far emergere le tematiche che ricorrono insieme. Si può vedere subito che la parola "*privacy*" non è così centrale come

```
var ars = {…,
    "AD":{"kw":["section_tech-policy","discipline","broadband-business-3","internet-regulation","policy","culture"], …},
    …}

var digitalData = {…,
    "keywords":{"display":"advertisers|ajit-pai|isps|online-privacy|type:report"},
    …}
```

FIG. 12 Le variabili che nascondono le parole chiave in un qualsiasi articolo su ArsTechnica. Per l'analisi è stata usata solo la seconda, contenente i termini inseriti dall'autore.

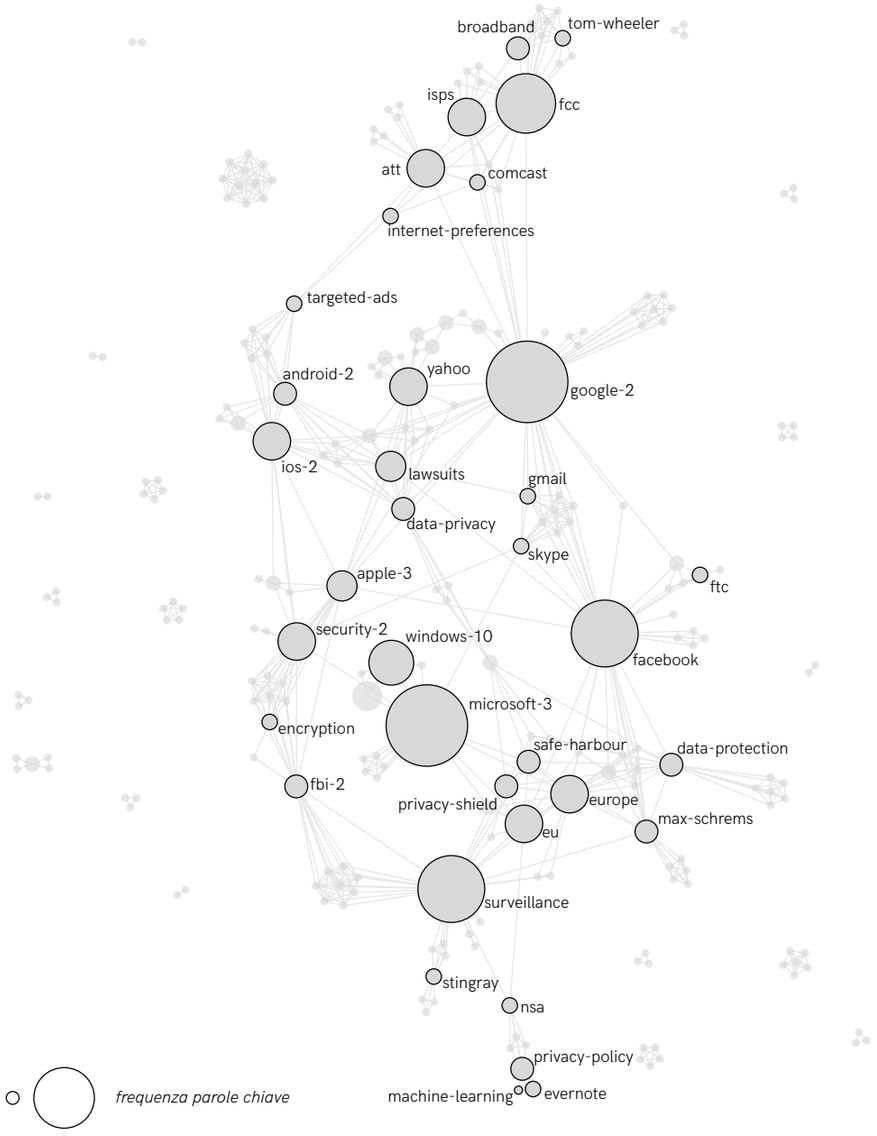
FIG. 13 Rete di parole chiave estratte dai primi 100 articoli della sezione policy di ArsTechnica.

si era ipotizzato e presenta anche una visione abbastanza monotematica legata alla sorveglianza di massa e alla violazione di privacy da parte di hacker.

Un secondo tentativo è stato quindi portato avanti estraendo lo stesso tipo di tag ma questa volta dai primi 100 articoli più rilevanti cercando esplicitamente "*privacy*" nella barra di ricerca.

→ FIG. 14

Anche questa volta però il risultato aggiunge ben poco a ciò che già si conosceva. La maggior parte della rete è occupata dai nomi delle grande aziende della Silicon Valley, citate per avvenimenti singoli più che per tematiche definite. Parte del problema è da ricercarsi nell'uso ufficioso dei tag da parte dei giornalisti di ArsTechnica: non vincolati da una convenzione progettata dall'alto, le parole chiave



riflettono i temi specifici di ogni articolo e perciò i filoni principali rimangono nascosti. Seppur non suggerisca nuove strade da percorrere, l'analisi di ArsTechnica ribadisce comunque legami e tematiche uscite da Wikipedia e Medium ed è perciò una conferma della strada percorsa fino a qui.

FIG. 14 Rete di parole chiave estratte dai primi 100 articoli riguardanti il tema privacy su ArsTechnica.

Lo studio di Wikipedia, Medium e ArsTechnica è stato di grande aiuto per avere una prima, veloce infarinatura delle diverse tematiche che si diramano a partire dalla parola privacy. L'analisi delle tre piattaforme ha confermato il focus sull'information privacy scelto nel capitolo 1.1, mostrando come il dibattito sulla privacy sia strettamente legato alla tecnologia ma in particolare a Internet e alla propria vita digitale. È stato anche possibile identificare alcuni degli attori principali che partecipano al dibattito attraverso le connessioni che formavano nei vari *device*: enti governativi e personaggi politici; associazioni per i diritti civili e agenzie indipendenti dallo Stato; i colossi della Silicon Valley.

1.3 Caratteristiche

Prima di andare a parlare delle problematiche che possono sorgere nella vita di tutti i giorni legate alla privacy, è importante cercare di capire meglio alcune delle sue proprietà. Certamente parte della difficoltà di darle una definizione chiara e univoca è dovuta agli attributi che la caratterizzano.

Prima di tutto la privacy è un processo che, anche se è presente universalmente in tutto il mondo, varia poi da cultura a cultura (Altman, 1977). Non solo, nella stessa cultura può cambiare da persona a persona e per lo stesso individuo essa può variare a seconda del contesto e nel tempo (Altman, 1977; Acquisti, 2016). Così se in generale un persona è restia all'idea di comunicare la sua posizione geografica durante tutto l'arco della giornata ad un estraneo, potrebbe senza problemi essere felice di farlo quando in cambio può usare quando gli serve un servizio di navigazione (la maggior parte delle applicazioni per smartphone di navigazione non smettono infatti di tracciare la posizione una volta che vengono chiuse). Tutto ciò potrebbe cambiare nel momento che questa stessa persona abbia un hobby che vuole tenere segreto.

Un altro aspetto importante è che non è detto che più informazioni vengano mantenute private, più si è soddisfatti del proprio livello di privacy. Ognuno ha un livello desiderato del proprio confine pubblico-privato: scostarsi troppo da questo livello, sia condividendo troppo (*crowding*) che nascondendo troppo (*isolation*), ha ripercussioni negative per la persona (Altman, 1977). Nei momenti di tristezza ad esempio si cerca istintivamente il conforto nelle parole dei propri cari in quanto la condivisione aiuta la guarigione psicologica, oppure a volte si desidera rivelare alcuni segreti per creare un legame intimo con un'altra persona (Acquisti, 2016).

Si possono osservare alcune proprietà della privacy anche quando essa diventa un bene di scambio. Nel momento in cui ad esempio si fornisce nome, cognome, anno di nascita, preferenze musicali, lista di amici, etc. a un social network come Facebook o Twitter, si sta implicitamente scambiando i propri dati in cambio di un servizio, in questo caso la possibilità di essere sempre facilmente in contatto con i propri cari. È così che i dati personali, e quindi la propria privacy, acquistano allo stesso tempo sia un valore come bene finale – l'importanza che viene data alle informazioni sensibili – sia uno come bene intermedio – il valore che gli si dà per accettare la transazione del servizio (ibidem). Soprattutto il secondo viene spesso tralasciato o dimenticato perché accade in maniera quasi implicita, tantoché quasi tutte le aziende che scambiano il loro servizio per le informazioni dei propri utenti, lo pubblicizzano

come gratuito nonostante una transazione economica effettivamente ci sia.

Inoltre questo scambio unisce tratti di natura diversa: tangibili, nel beneficio immediato che si trae dal servizio o nella consapevolezza di aver ceduto parte delle informazioni personali; intangibili, nel momento che queste informazioni verranno utilizzate o rivendute ad altre aziende in futuro oppure che esse potrebbero essere usate impropriamente senza il proprio consenso; incommensurabili, ovvero le opportunità e i rischi di vivere in una società dove si può conoscere tutto di una data persona (ibidem), in quanto la privacy non è solamente un diritto individuale ma è parte costitutiva di ogni società e ne contribuisce al benessere generale (Solove, 2006).

Tutte queste caratteristiche, unite alla sfuggevolezza della definizione di privacy, creano una serie di situazioni quando si interagisce con servizi online o sugli smartphone che per gli utenti sono poco chiare, opache. Esse portano ad un fenomeno noto come “privacy paradox” (Acquisti, 2016; Norberg, 2007), per cui anche se gli utenti si dicono interessati a salvaguardare la propria privacy, i loro comportamenti online non sembrano rispecchiare queste preoccupazioni.

[Blackboxing is] the way scientific and technical work is made invisible by its own success. When a machine runs efficiently, when a matter of fact is settled, one need focus only on its inputs and outputs and not on its internal complexity. Thus, paradoxically, the more science and technology succeed, the more opaque and obscure they become.

Bruno Latour

2 — Opacità multilivello

Come introdotto nel capitolo precedente, l'utilizzo di servizi online, di applicazioni mobile e la consultazione di siti web non è quasi mai realmente gratuita: avviene sempre una transizione economica di un servizio in cambio di informazioni personali che vengono poi rivendute a terze parti o rielaborate internamente. Il ciclo di vita dei dati inizia con la collezione, seguita dall'elaborazione ed infine si conclude con la disseminazione che può essere verso terze parti o di ritorno all'utente sotto forma di personalizzazione del servizio.

Solove individua in questi passaggi numerosi rischi di "attività dannose" che possono essere messe in atto da governi, aziende e persone con fini malevoli per ledere la privacy dell'individuo (Solove, 2006). Egli distingue nella sua tassonomia quattro gruppi. L'accesso alle informazioni (*information collection*) può essere fatta tramite sorveglianza intrusiva o facendo pressioni per far rivelare informazioni; l'elaborazione (*information processing*) può generare usi impropri, discriminazione o aggregazione eccessiva di informazioni riguardo la stessa persona; la disseminazione

(*information dissemination*) diventa pericolosa quando i dati vengono usate a danno dell'individuo per diffamare, ricattare ed estorcere, oppure quando viene rotto il patto di confidenzialità rivelandoli a terze parti. Solove individua infine un ultimo gruppo legato all'invasione nella vita privata (*invasions*) che comprende le intrusioni e le manipolazioni indirette che portano l'individuo a dover modificare le proprie abitudini.

Tutte e quattro le categorie dipendono in qualche modo dalle qualità etiche e morali di aziende e governi ma, mentre disseminazione e invasione sono legate quasi esclusivamente alla volontà e onestà dei soggetti che hanno ottenuto i dati, accesso ed elaborazione sono contrastabili anche dal basso, attraverso soluzioni che portino trasparenza laddove il processo risulta opaco.

2.1 Dati personali: problemi legati all'accesso

La principale fonte di opacità che si genera nel momento in cui si decide di utilizzare un servizio che richiede i propri dati per funzionare viene chiamata "posizione d'informazione asimmetrica" (Acquisti, 2016): una delle due parti ha un potere su un pezzo d'informazione che l'altra parte non ha. Quando ad esempio si installa un'applicazione per smartphone che aiuta a monitorare la nostra attività giornaliera (Fitbit, Google Fit, Runtastic, etc.), viene dato il permesso alle aziende proprietarie dell'app di raccogliere tutta una serie di dati personali e di usarli o cederli a terze parti senza poterne più venire informati o avere un minimo controllo su di essi, in conformità con le *privacy policies*[1] accettate installando l'applicazione. Queste privacy policies si rivelano quasi sempre essere documenti inaccessibili e complessi (Cate, 2010), accettate senza prestarci troppe attenzioni piuttosto che lette attentamente. Allo stesso modo, quando si vuole ordinare la cena comodamente da casa utilizzando servizi online come Foodora, Deliveroo o Uber Eats, viene chiesto in cambio di creare un profilo su queste piattaforme. Il profilo permetterà di ricevere offerte su misura e personalizzazioni, ma allo stesso tempo obbligherà ad acconsentire ad alcune condizioni – spesso sotto forma di caselline con la spunta – che daranno la libertà a queste aziende di monitorare che punti dello schermo vengono fissati, quanto in basso si naviga nella pagina dei ristoranti, oppure di vendere le proprie abitudini a compagnie pubblicitarie.

È importante sottolineare che anche prima di aver installato l'app o usato la piattaforma online c'era un'asimmetria, però essa era a favore dell'utente in quanto era l'unico in possesso delle informazioni. Nella transazione questa asimmetria si è così invertita, lasciando la persona ignara

[1] • Documento informativo che descrive il trattamento (accumulazione, elaborazione e trasmissione) dei dati personali dell'utente. *Sono come un contratto di affitto in cui si paga un affitto (i dati) per accedere "gratutitamente" alla casa, ma il contratto può cambiare da un momento all'altro senza preavviso*.

di quanto e come i suoi dati verranno utilizzati, nonché quali conseguenze questo avrà per lei (Acquisti, 2016).

A nascondere la creazione di questa posizione svantaggiosa concorrono diversi fattori: alcuni sono progettati più o meno malevolmente dalle aziende per tutelarsi legalmente nelle loro attività, anche sfruttando le caratteristiche della privacy introdotte nel capitolo precedente, altri sono intrinsechi nella natura stessa della transazione dato-servizio (che è comunque in qualche modo riconducibile ad un'azione ben progettata in quanto parte di un sistema economico).

Innanzitutto lo scambio è asincrono e sproporzionato tra le parti. Usando un qualsiasi servizio che vive dei dati raccolti dai suoi utenti, il beneficio legato ad esso è immediato mentre l'utilizzo delle informazioni personali avverrà in un futuro non ben definito e senza nessuna notifica. In questo modo, mentre per l'azienda il trasferimento di dati è la parte essenziale della transazione, per l'utente la sottigliezza e l'intangibilità con cui questa avviene ne fa percepire meno l'onerosità o perlomeno è molto più difficile darle il giusto peso o addirittura riconoscerla come tale. Tutta l'attenzione è rivolta alla praticità del servizio che si riceve (ibidem). Per fare un esempio concreto, ogni volta che si cercano notizie su un motore di ricerca come Google o Yahoo!, si sta implicitamente accettando che vengano raccolti dati su cosa si cerca, come vengono definite le chiavi di ricerca, quali link vengono cliccati, ma anche l'IP[2] del computer, la posizione geografica, il browser usato per navigare. Mentre l'appagamento dovuto all'aver trovato ciò che si cercava – o lo sconforto per la situazione opposta – è immediato e concreto, il costo di aver rivelato parte di sé stessi o delle proprie abitudini passerà molto probabilmente inosservato, frutto di uno scambio invisibile e facilmente dimenticato.

Un'altra fonte di opacità sono le privacy policy. Esse sono documenti legali che descrivono alcune o tutte le modalità con cui una parte raccoglie, usa e condivide i dati raccolti sui suoi clienti. Come accennato prima, anche se queste informative dovrebbero essere presenti proprio per eliminare questa asimmetria e tutelare la privacy dei consumatori, la loro crescente lunghezza, complessità e mutevolezza le rende molto spesso degli ostacoli alla trasparenza (Cranor, 2012). Non solo, le privacy policy di un'azienda non descrivono ovviamente in alcun modo cosa faranno "aziende terze" (*third-parties*) quando riceveranno i dati dei clienti della prima, aumentando notevolmente la difficoltà per essi di prendere una decisione consapevole. Le informative sulla privacy sono interamente auto-regolate dalle industrie, nel senso che non esistono leggi precise (né in USA né all'interno dell'UE) che ne delineino le caratteristiche ma solo un

[2] • Identificativo numerico assegnato ad ogni dispositivo che si connette a Internet. *È come il numero di telefono. Il sistema telefonico funziona perché ogni numero è differente. I numeri contengono anche informazioni geografiche generali, come la nazione e la regione. (fonte: Sideways Dictionary)*

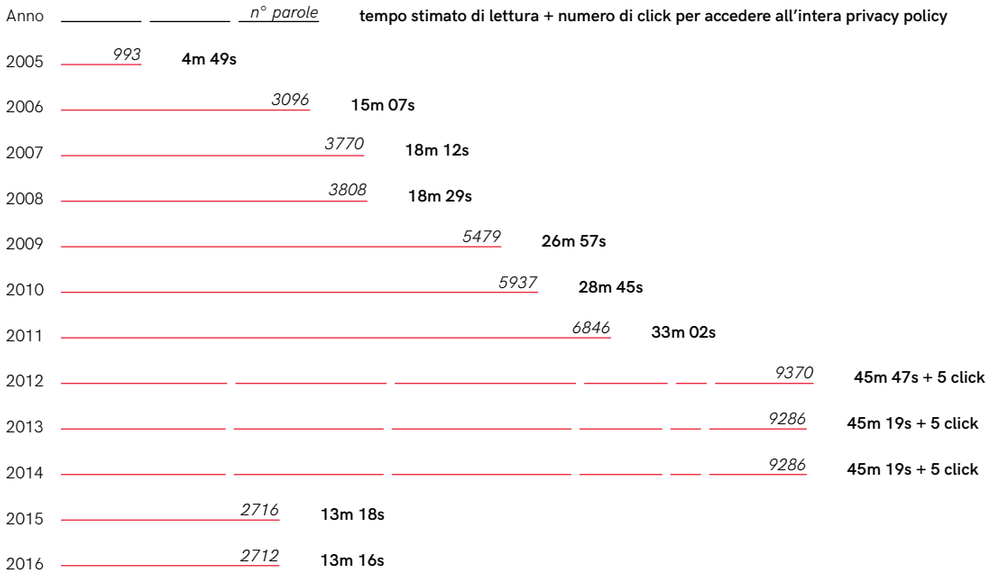


FIG. 15 Lunghezza della privacy policy di Facebook dalla sua fondazione ad oggi. La riduzione degli ultimi anni è sicuramente un passo positivo, ma la difficoltà nella terminologia usata rimane comunque un grande ostacolo.
Fonti: https://web.archive.org/web/ /http://www.facebook.com/policy.php; http://niram.org/read*

insieme di linee guida suggerite dall'OCSE[3]. È quindi chiaro come le aziende non abbiano nessun interesse nell'essere più chiare o specifiche, avendo molto più flessibilità per poter operare a loro vantaggio.

Esistono invece delle differenze a livello legislativo riguardo all'ultimo punto che contribuisce a portare opacità nell'accesso ai dati: le opzioni di default e le modalità con cui esse vengono presentate all'utente quando acconsente ad utilizzare un servizio online. Mentre in UE sono state definite norme precise su quali tipi di informazioni necessitano esplicitamente il consenso dell'utente (sistema chiamato *opt-in*), negli Stati Uniti questi principi sono di nuovo lasciati all'auto-regolazione, creando un sistema di "avviso e consenso"[4] che si basa sulla volontà dell'utente di voler espressamente essere escluso dalla raccolta dati (sistema *opt-out*)(Acquisti, 2016; Cate, 2010; Cranor, 2012). In entrambi i casi è però facile confondere o depistare l'utente attraverso dei semplici accorgimenti di linguaggio o presentazione della scelta. Diversi studi hanno infatti dimostrato come la scelta di una costruzione positiva o negativa della frase ("Accosento a..." oppure "Non acconsento a...") influiscano in maniera non indifferente sul numero di persone che spuntano la casella per il trattamento dei dati (Johnson et al., 2002), oppure viceversa se queste caselle sono già spuntate come opzione di default. Queste scelte sono appositamente progettate per non proporre mai entrambe le alternative né per sottolineare l'avvenimento di un compromesso tra la scelta che si sta facendo e le sue conseguenze (ibidem).

Tutti gli accorgimenti progettuali citati sembrano in apparenza essere stati creati per lasciare il controllo della propria privacy nelle mani del consumatore/utente ma si rivelano un'enorme fonte di opacità per la maggior parte di essi non possedendo le adeguate conoscenze tecniche o la necessaria consapevolezza per poter efficacemente prendere decisioni per proteggere i propri dati personali (Acquisti, 2016). Subito dopo quella spunta selezionata o quel bottone cliccato, la traccia della transazione scompare alla vista, invisibile ma perpetua. Tutto ciò che appare è l'utilità e la comodità della nuova applicazione o piattaforma.

[3] Organizzazione per la cooperazione e lo sviluppo economico.
https://www.oecd.org

[4] Notice and choice, traduzione mia. (Acquisti, 2016; Cate, 2010; Cranor, 2012)

APPROFONDIMENTO

GOOGLE LOCATION HISTORY

Parlando di privacy Apple, Twitter, Facebook, Google e molti altri colossi digitali lavorano incessantemente per migliorare la trasparenza verso i propri utenti: la loro posizione dominante dipende molto dalla fiducia che le persone ripongono riguardo al loro operato ed un passo falso potrebbe danneggiarne gravemente l'immagine e il fatturato. Google in particolare spiega abbastanza approfonditamente il cosa, come e perché della sua raccolta dati[5], rilasciando ogni anno un rapporto sulla trasparenza[6] (ad esempio riportando ogni richiesta di dati da parte dei governi e la percentuale di richieste accettate e negate) e permettendo all'utente di poter scaricare in qualsiasi momento i dati archiviati su di lui[7]. Come breve esperimento durante il periodo di analisi, si è perciò provato a scaricare l'archivio delle posizioni geografiche registrate da Google. L'azienda ha già un'interfaccia progettata benissimo[8] che mostra ai propri utenti tutti i luoghi dove essi sono stati, permettendo di risalire ai propri movimenti nel passato. Quello che non mostra è quanto spesso vengono richieste le coordinate geografiche. L'obiettivo era quindi quello di testare se questa possibilità ridesse ad un utente qualsiasi il controllo sulla condivisione delle proprie informazioni, rendendo effettivamente più trasparente il processo di accesso alle informazioni dell'azienda californiana.

(→METODO: Compendio metodologico)

Tramite la funzione *Takeout*[7] è stata quindi scaricata la cronologia delle posizioni geografiche archiviate da Google, il quale fornisce i dati in formato JSON[9] o KML[10]. Già appare una prima barriera alla fruizione: un utente poco esperto avrebbe una discreta difficoltà nel capire come aprire file in questi formati, rendendo praticamente impossibile la comprensione di ciò che essi contengono. Il file scaricato – in formato JSON per poter poi essere usato e visualizzato più facilmente – contiene i dati geografici dal 18 Settembre 2015 al 5 Febbraio 2017 ed è talmente pesante (circa 79 Mb, tantissimo per un file di testo) che ben pochi editor sono in grado di aprirlo. Si è deciso di limitarlo però al solo periodo di tesi, quindi dal 19 Luglio 2016 al 5 Febbraio 2017, perché prima di questo periodo erano stati disattivati per lungo tempo i servizi di geolocalizzazione e perciò non era possibile avere un dataset continuo di tutto il periodo. Inoltre questo ha permesso di avere anche un file di dimensioni più gestibili (comunque enorme, 54 Mb).

Una volta aperto, il JSON non presenta alcun commento o spiegazione per poter essere compreso a fondo: sicuramente ogni istante di tempo – espresso in Epoch time[11], illeggibile per una persona umana – riporta un momento in cui è stata richiesta la posizione da Google, con riportate le coordinate geografiche e la precisione della misurazione, però il significa-

5 ▪ *https://privacy.google.com*

6 ▪ *https://www.google.com/transparencyreport*

7 ▪ *https://takeout.google.com/settings/takeout*

8 ▪ *https://www.google.com/maps/timeline?pb*

9 • Javascript Object Notation. È un formato per lo scambio di dati, molto facile da leggere per l'uomo e molto facile da analizzare per la macchina. Esempio:

```
var posizione = {
        "nome" : "Milano",
        "latitudine" : 45.4773,
        "longitudine" : 9.1815,
        "tipologia" : "città"
    };
```

10 • Keyhole Markup Language. È un linguaggio per gestire dati geospaziali inizialmente introdotto da Google. Esempio:

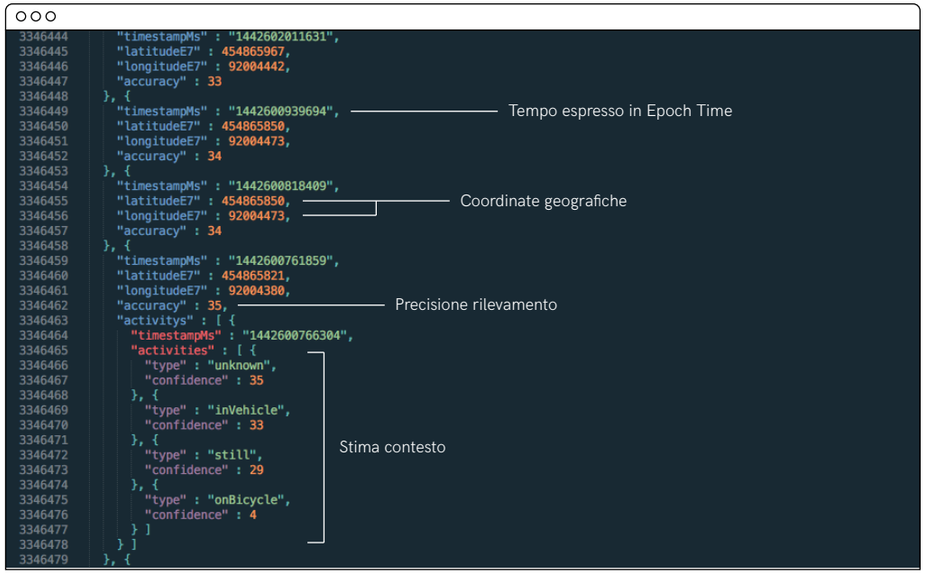
```
<kml>
<Placemark>
  <description>Milano</description>
  <name>Milano</name>
  <Point>
    <coord>45.4773,9.1815,0</coord>
  </Point>
</Placemark>
</kml>
```

11 • Tempo in secondi dalla mezzanotte del 1° Gennaio 1970 (chiamata *epoca*). Il 27 Aprile 2017 alle 09:00 saranno passati 1.493.276.400 secondi dall'*epoca*.

to preciso della voce "activitys" non è chiarissimo, anche se si può supporre che siano i momenti in cui Google rileva che c'è un cambio di coordinate geografiche e perciò tenta di stabilire se il soggetto è fermo, in macchina, in bici, etc.

Siccome il dataset è troppo grande per poter essere analizzato manualmente, è stato creato uno script che estrapolasse il numero di rilievi al giorno e li raggruppasse poi per ore. Così sarebbe stato poi possibile visualizzare il numero di rilievi nel tempo ed avere un'immagine chiara di quante volte Google monitorasse la posizione geografica.

IMG. 02 Struttura del JSON relativo alla *Location History*. Il testo in bianco sono un'interpretazione delle voci, alcune delle quali alquanto ambigue.



Il primo passo è stato quello di mappare il numero di accessi alla localizzazione richiesti ogni giorno, visualizzandoli attraverso una mappa di calore che riportasse in orizzontale il giorno e in verticale il mese. L'ipotesi era quella di poter notare qualche schema ricorrente, magari dovuto a spostamenti molto grandi o attività elevate. Apparentemente però, paragonando il risultato con l'interfaccia fornita da Google[12], non si nota nessuna ricorrenza: il numero di accessi richiesti per giorno varia da 190 a 4.300 – cioè circa 180 volte all'ora – ma non sembra avere correlazioni con distanza percorsa o movimento frequente. L'unica cosa che si può vedere è che i giorni col più basso numero di posizioni rilevate corrispondono con viaggi all'estero, dove il segnale prendeva poco o affatto.

Guardando il dataset da un'altra prospettiva, mostrando cioè la variazione oraria di un dato giorno attraverso una linea temporale e sovrapponendo la linea di ogni giorno rilevato, si è quindi cercato di vedere se uscisse più chiaramente un pattern di qualche genere. È interessante vedere come emergano due tipi di giornata, una in cui l'ubicazione è rilevata da Google dalle 10 alle 20 circa volte l'ora – quindi ogni 3-6 minuti – e un'altra dove viene rilevata intorno alle 180 volte l'ora, ovvero in media circa ogni 20 secondi. La divisione è maggiormente visibile durante la notte, anche se non c'è alcuna corrispondenza con le serate passate fuori. Ciò può essere visto con maggior chiarezza se isoliamo alcune giornate specifiche.

12 *https://www.google.com/maps/timeline?pb*

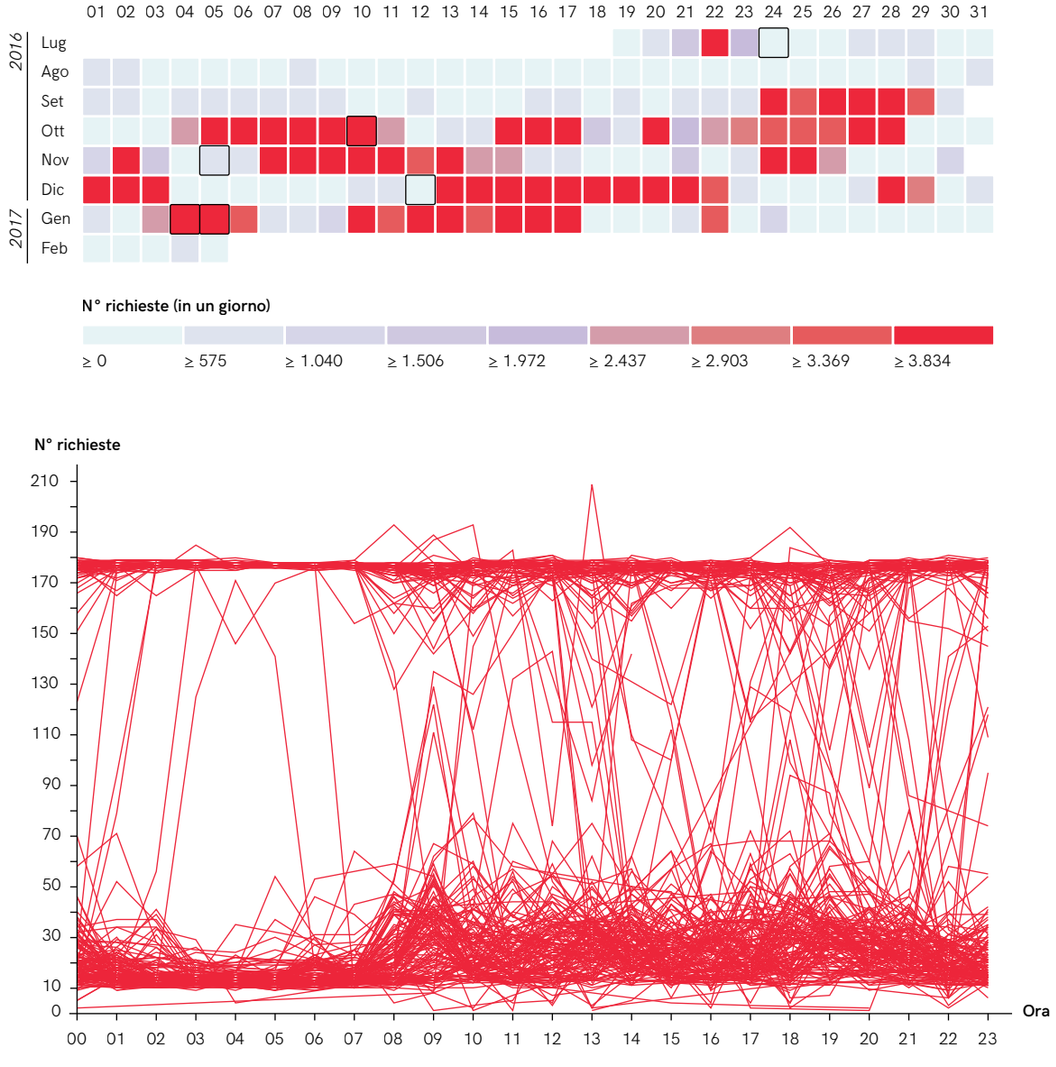
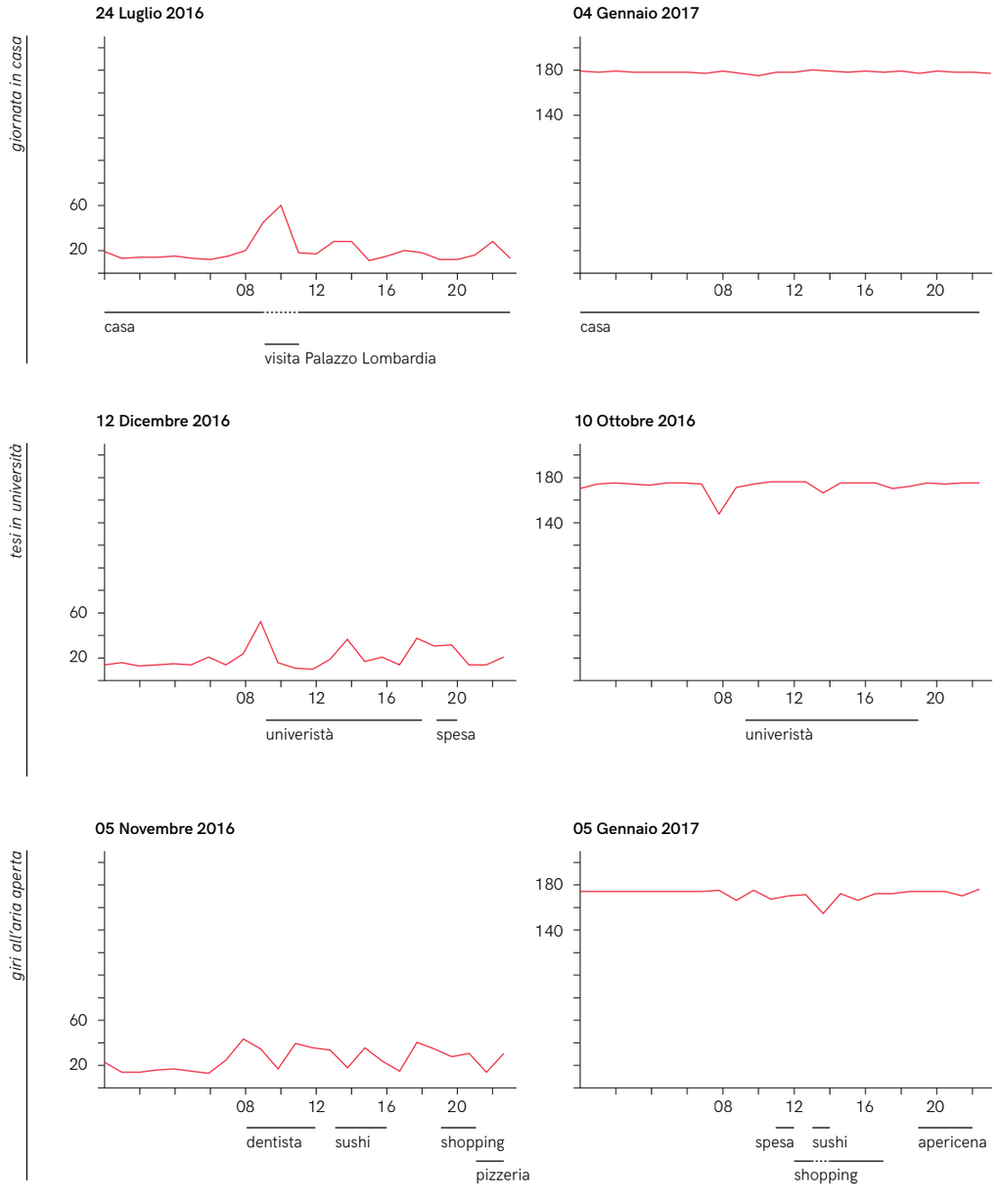


FIG. 16 Rappresentazione da diverse prospettive del numero di accessi alla geolocalizzazione da parte di Google, dal 19 Luglio 2016 al 05 Febbraio 2017.



Avere una misura della quantità di rilevazioni che prese al proprio smartphone potrebbe provocare un certo senso di spiazzamento. Pensandoci è abbastanza ovvio che per fornire un servizio così preciso e puntuale esso debba rilevare le coordinate geografiche molto frequentemente, però poterlo vedere lo rende improvvisamente più reale. Nonostante i passi compiuti da Google per rendere i suoi processi più trasparenti e ridare parte del controllo in mano all'utente siano sicuramente positivi, molto deve essere ancora fatto perché ci sia una vera trasparenza sulla raccolta dati. Questa si realizzerebbe veramente solo quando ogni dato e metadato raccolto sull'utente gli siano restituiti integralmente, in modo facilmente fruibile, così che anch'esso possa realmente giovare delle informazioni che rilascia. In particolare l'archivio restituito grazie alla funzione *Takeout* è trasparente solo in apparenza, in quanto senza le necessarie conoscenze tecniche esso risulta inutilizzabile.

2.2 dati personali: problemi legati all'utilizzo

Una volta raccolti, i dati vengono immagazzinati e rielaborati sia per rivenderli a terze parti che per restituirli agli utenti come parte integrante del servizio, ad esempio il suggerimento di prodotti affini su Amazon o di film simili su Netflix. La gestione di questi dati è lasciata in mano a degli algoritmi che permettono così di automatizzare il processo e di poter servire autonomamente e rapidamente milioni di persone. Con la progressiva digitalizzazione della vita e il conseguente aumento esponenziale di dati raccolti, sia in termini di quantità che di dimensioni, il ruolo di questi algoritmi è diventato sempre più centrale in moltissimi ambiti della quotidianità: essi decidono cosa verrà restituito quando si fa una ricerca su Google, quali amici o notizie vedere su Facebook, chi verrà assunto al lavoro e in quale scuola superiore andrà il proprio figlio (Tufekci, 2015). Ma cos'è esattamente un algoritmo?

Al livello più semplice possibile un algoritmo non è nient'altro che una serie di operazioni, definite da un numero finito di regole, che conduce al risultato in un numero finito di passaggi. Concettualmente anche una ricetta di cucina è un algoritmo: unendo gli ingredienti con l'ordine scritto ed usando gli strumenti, le temperature e i tempi giusti, un cuoco abile trasformerà sempre un algoritmo per la crostata di mele in una deliziosa crostata di mele. Ov-

viamente gli algoritmi che trattano i dati sono definiti da linee di codice, ovvero una serie di regole definite da un programmatore usando un linguaggio che il computer può comprendere.

Per semplicità essi possono essere divisi in due macrogruppi (Diakopoulos, 2014). Da una parte ci sono gli algoritmi che possono essere chiamati statici o espliciti, dove il risultato che si vuole ottenere è già conosciuto e quindi vengono descritte una ad una tutte le regole per andare al passaggio successivo – ad esempio il riordino di un elenco in ordine alfabetico o il calcolo della temperatura media mensile a Milano negli ultimi 100 anni. Questi algoritmi hanno sempre, anche nei casi più complessi, elementi ben definiti in entrata (*input*) e restituiscono sempre un risultato (*output*) definitivo e univoco, "corretto" (Tufekci, 2015). Il codice con cui sono scritti è trasparente, nel senso che con le conoscenze tecniche adeguate è facilmente revisionabile e riproducibile.

Ci sono poi algoritmi dinamici, che servono per trattare quantità enormi di dati multi-dimensionali ed eterogenei. Le regole definite dal programmatore in questo caso non sono i singoli passaggi da seguire ma piuttosto delle regole che siano dinamiche e flessibili rispetto ai dati (Diakopoulos, 2014). È questo il caso degli algoritmi di *machine-learning* (anche detti di "apprendimento automatico"), che cercano di prendere decisioni in base a pattern imparati analizzando i dati. Sono proprio una combinazione di algoritmi di questo tipo che fanno ormai sempre più parte della quotidianità: grazie ad essi infatti è possibile dare un senso ad una mole di informazioni che risulterebbe incontrollabile per una gestione solo umana, perché definire una logica precisa per prendere decisioni su questi dati sarebbe totalmente inefficiente (Burrel, 2016). Essi migliorano l'esperienza quotidiana enormemente ma contengono nella loro natura numerosi livelli di opacità che è necessario conoscere e tenere presente, soprattutto per il potere e la rilevanza che acquisiscono di giorno in giorno sulle persone.

La descrizione dettagliata di questo tipo di algoritmi esula dagli obiettivi di questa tesi, ma alcuni principi generali verranno presentati in maniera semplificata per poter rendere più chiara la spiegazione riguardo all'opacità. Nonostante le numerose varianti in cui possiamo classificare i diversi approcci di machine-learning, tutte seguono il medesimo principio: l'algoritmo riceve un set di dati da cui cerca di estrarre dei pattern generali su cui poi baserà le decisioni prese su nuovi informazioni. Per fare un esempio più concreto, si supponga di voler creare un algoritmo che riconosca il tipo di albero presente in una foto ogni volta che gliene si presenta una. All'algoritmo verranno sottoposte migliaia di foto di alberi etichettate col nome dell'albero

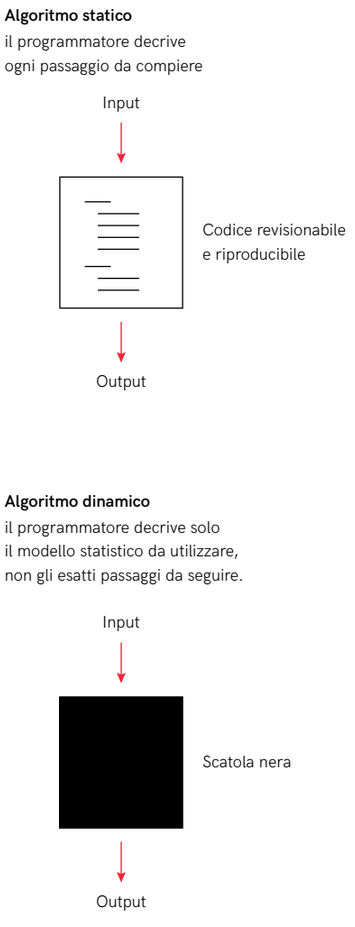
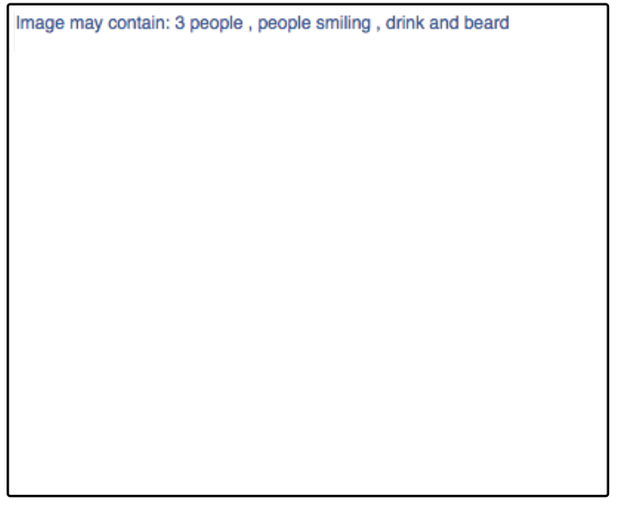


FIG. 17 Schema semplificato della differenza tra algoritmi statici e dinamici.

corretto (questo viene chiamato “addestramento supervisionato”). Esso definirà internamente una logica che assegna ad ogni tratto dell’albero e ad ogni pixel dei punteggi che peseranno più o meno nella scelta finale per decidere se l’albero in questione sia un abete o una palma. Addestrato l’algoritmo, gli si potranno sottoporre nuove immagini di albero mai viste e lui userà e riaddestrerà continuamente la sua logica interna per dire il nome dell’albero.



IMG. 03 Il testo assegnato ad ogni immagine caricata su Facebook viene creato automaticamente proprio usando un algoritmo simile a quello descritto.

Qui già affiora un primo e importante livello di opacità: la complessità non risiede né nel dataset, per quanto vasto possa essere, né nel codice scritto dai programmatori, comprensibile ad un occhio esperto, quanto piuttosto nella “scatola nera” che l’algoritmo crea automaticamente mentre apprende (Burrel, 2016). Il termine scatola nera vuole proprio sottolineare questa incapacità di poter vedere con chiarezza cosa accade al suo interno, tanto che neanche gli stessi programmatori possono dire con esattezza come un algoritmo di machine-learning arrivi ad un determinato output (Gillespie, 2014). Ma se neanche i creatori sono pienamente consapevoli della scelta dell’algoritmo, di chi è la responsabilità quando esso sbaglia? Certo ha poca importanza se una mail viene catalogata erroneamente come spam anche se veniva da una persona affidabile, ma se la posta in gioco fosse l’assunzione di un candidato rispetto ad un altro o addirittura la sentenza da infliggere ad un accusato, le conseguenze di una simile decisione sarebbero molto gravi sul piano etico e meriterebbero un discorso a parte che però esula da questa tesi.

Per poter comprendere meglio le altre fonti di opacità è utile, citando Diakopoulos, separare le tipologie di decisioni che l’algoritmo può prendere in categorie, perché ognuna di esse nasconde delle problematiche sia dovute alla complessità dell’algoritmo che al rischio di intenti malevoli dei

creatori (Diakopoulos, 2014). Secondo l'output finale può avvenire per assegnazione di priorità, classificazione, associazione o filtraggio. Raramente queste categorie esistono da sole ma vengono concatenate per prendere decisioni più complesse o che necessitano di una trasformazione dell'informazione.

ASSEGNAZIONE DI PRIORITÀ

Assegnazione di priorità significa creare una classifica dando risalto ad alcuni elementi a spese di altri. Per fare questo vengono forniti all'algoritmo dei criteri generali su cui esso si basa per creare delle metriche e definire la classifica. La città di New York ad esempio, avendo un personale ridotto di ispettori di sicurezza, usa questo genere di algoritmo per stilare l'ordine in cui verranno ispezionati gli edifici.

La scelta dei criteri iniziali influisce molto sul risultato e visto che spesso questi non sono resi pubblici, è difficile per un osservatore esterno comprendere i pesi assegnati a tutti i fattori che hanno contribuito alla graduatoria. Su questioni più delicate si possono infatti notare i rischi: alcune zone degli Stati Uniti adottano già la "sorveglianza predittiva" (*predicting policing*) dove le forze di polizia vengono distribuite per la città secondo pattern riconosciuti da un algoritmo che analizza numerosi fattori tra cui lo storico dei crimini commessi e la distribuzione geografica delle diverse comunità. Essendo però gli algoritmi coperti da copyright ed il loro funzionamento secretato, è impossibile dire con certezza tutti i fattori presi in analisi e quanto ognuno di essi influisca nella decisione finale. Inoltre l'addestramento dell'algoritmo è basato su dati storici e raccolti da persone, e quindi non divulgabili per questioni di privacy, creando un conflitto tra la necessità di rendere gli algoritmi trasparenti ed al tempo stesso tutelare le persone coinvolte. L'utilizzo di dati personali inoltre porta il rischio di rafforzare stereotipi e pregiudizi umani.

CLASSIFICAZIONE

In questo caso l'algoritmo assegna ad ogni elemento una categoria basandosi sui dati con cui è stato addestrato. Appartiene a questa categoria il riconoscimento facciale nelle foto su Facebook o l'algoritmo di Amazon che decide in quali regioni sarà disponibile il servizio Amazon Prime Now. Un recente articolo su Bloomberg[13] ha però mostrato come in molti grossi centri urbani degli Stati Uniti le zone escluse dal servizio coincidessero con incredibile precisione ai quartieri a maggioranza afro-americana. Amazon ha subito chiarito che fattori demografici come la razza sono stati perentoriamente esclusi dall'algoritmo.

Sebbene Amazon non avesse alcuna intenzione malevo-

[13] *https://www.bloomberg.com/graphics/2016-amazon-same-day*

la, è possibile che l'algoritmo abbia ereditato tutti i pregiudizi impliciti del dataset su cui è stato addestrato, in quanto tiene probabilmente (sempre probabilmente perché ovvie motivazioni di competitività mantengono il codice protetto) conto di fattori come tasso di criminalità e disparità di reddito. Nei mesi seguenti alla pubblicazione dell'articolo Amazon ha poi raggiunto i quartieri esclusi a Boston, New York e Chicago.

ASSOCIAZIONE

Per associazione si intende l'individuazione di relazioni tra due o più entità. Nell'output finale assomiglia alla classificazione in quanto definisce in pratica due classi: la relazione c'è oppure no. Su Wikipedia "vivono" dei *bot* (anche loro sono algoritmi) che creano link tra pagine ogni volta che notano una relazione tra esse (Diakopoulos, 2014). Un altro esempio sono i suggerimenti a film simili su Netflix. Anche qui i rischi sono legati a faziosità dei creatori o pregiudizi involontari nel dataset. Nel 2007 se si fosse tentato di cercare su Google "*she invented*", l'algoritmo ci avrebbe gentilmente suggerito: "*did you mean '**he** invented'?*" (Gillespie, 2014). Anche in questo caso sicuramente non si tratta di una manipolazione sessista da parte degli ingegneri di Google ma molto semplicemente la maggior frequenza con cui la parola "lui" precede "inventò" sull'intero Web.

È interessante però come in qualche modo gli algoritmi facciano affiorare tutti i pregiudizi passati o presenti insiti nella cultura umana ed è importante tenerlo presente quando si lascia che essi facciano affermazioni su sulla società (ibidem).

FILTRAGGIO

Quest'ultima tipologia include o esclude gli elementi in base a criteri definiti. È molto spesso lo step finale delle tipologie precedenti e le complementa. Utilizzando applicazioni come Flipboard, le notizie che si leggono ogni giorno sono filtrate a partire da un set che è stato previamente categorizzato, associato alle proprie preferenze e riordinato per rilevanza. Il rischio che si presenta è che vengano presentate sempre e solo le cose che piacciono di più o che piacciono alla maggior parte delle persone, entrambi fattori che possono minare la capacità di formulare opinioni diverse. Già nel 2011 Eli Pariser aveva sottolineato questo problema chiamandolo "*filter bubble*" (Pariser, 2011) e nell'ultimo anno soprattutto con le elezioni americane il termine è stato molto discusso riferendosi all'algoritmo di Facebook. Come sempre questi algoritmi sono proprietari e oscuri, perciò anche per ricercatori e giornalisti è impossibile indagare sui meccanismi interni e ricostruirne le logiche.

Un ultimo esempio estremo di filtraggio è la censura: governi che tengono sotto controllo i social network ed il web in generale per sopprimere opinioni rivoluzionarie è una prospettiva alquanto oscura e non così remota. Un rapporto dell'organizzazione indipendente Freedom House[14] uscito nel Novembre 2016 analizza la libertà su Internet in 65 nazioni che contengono l'88% degli utenti aventi accesso al web[15]. In 38 di queste nazioni – in cui vivono il 60% degli utenti – delle persone sono state arrestate per via delle loro attività sui social media.

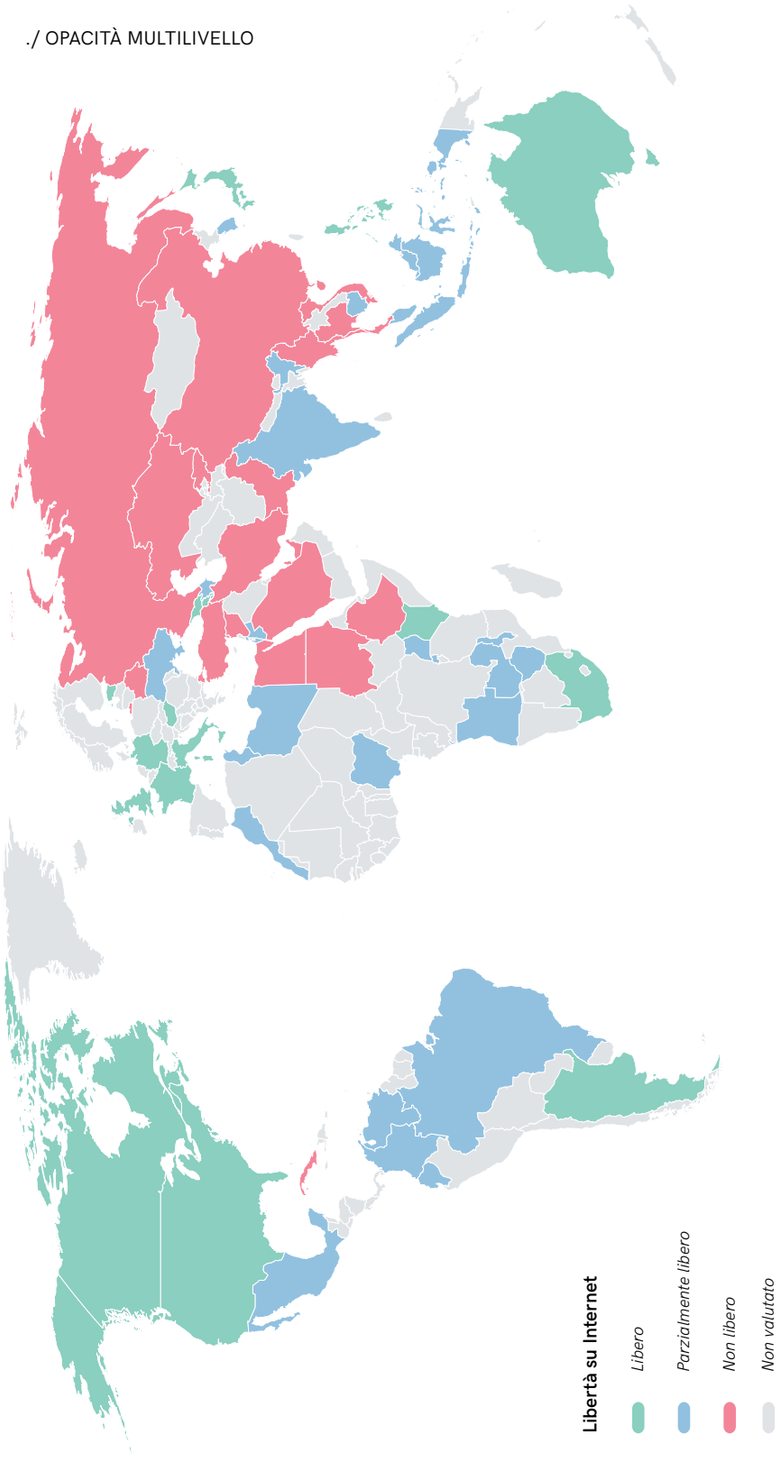
Citando nuovamente Diakopoulos, l'intento qui non è di demonizzare gli algoritmi, bensì di riconoscere che anch'essi operano con faziosità e pregiudizi esattamente come le persone (Diakopoulos, 2014). L'argomentazione che le scelte di un computer siano necessariamente imparziali ed eque non può essere accettata, in quanto il codice di un algoritmo è stato progettato da persone e i dati su cui si allena sono stati raccolti secondo criteri definiti da persone. La sua presunta oggettività è contaminata dalla soggettività dei suoi creatori.

Se non è facile, forse addirittura impossibile, aprire queste scatole nere per capirne meglio l'interno, si può almeno cercare di rendere visibile ed apparente la loro presenza.

14 ▲ Organizzazione di controllo indipendente per l'espansione di libertà e democrazia nel mondo. *https://freedomhouse.org*

15 ■ *https://freedomhouse.org/sites/default/files/FOTN_2016_BOOKLET_FINAL.pdf*

FIG. 18 Mappa che rappresenta la libertà su Internet nel mondo, secondo l'analisi svolta nel 2015-2016 da Freedom House.
Fonte nazioni vettoriali: Free Vector Maps.com



Transparency should be proportional to the power that one has. The more power one has, the greater the dangers generated by that power, and the more need for transparency.

Julian Assange

Dear subscriber, you have been registered as a participant in a mass disturbance.

SMS mandato dalle autorità ucraine a tutti i partecipanti alle proteste di Kiev del Febbraio 2014

3 — Asimmetrie di potere

È già stato visto come la raccolta di dati personali ribalti l'asimmetria di informazione, ed il detentore originale dell'informazione cede i vantaggi che aveva nell'esserne l'unico possessore perdendo il controllo su come essa verrà utilizzata. Possedere delle informazioni che nessun altro conosce è una forma di potere, ed ora l'individuo lo ha perso a favore dell'azienda, istituzione o governo che ne hanno guadagnato molto. Più informazioni possono essere raccolte su quell'individuo, più il divario tra poteri si allargherà. Un commerciante che conosce o può stimare, basandosi su comportamenti precedenti, il prezzo di prenotazione di un albergo che una data persona è disposta a spendere, può modificare a suo vantaggio le offerte (Acquisti, 2016); un governo totalitario può incarcerare un insegnante che sospetta possa ostacolare il regime, basandosi sulle informazioni che ha raccolto su di lui. Nonostante questo avvenisse già anche prima dell'avvento del digitale, la possibilità di poter raccogliere dati facilmente e continuamente riduce notevolmente gli ostacoli che potevano esserci in precedenza. Avendo le risorse per poter raccogliere ed elaborare

quantità sempre maggiori di dati su una grossa fetta della popolazione, su un'intera nazione o potenzialmente sul mondo intero, l'asimmetria di potere diviene pericolosissima. Sempre più associazioni per la difesa dei diritti digitali e attivisti richiedono e lavorano alla creazione di misure per ristabilire una sorta di equilibrio.

Il concetto non è relegato al tema della privacy: per ogni diritto riconosciuto ai cittadini di uno stato, servono misure per tutelarlo ed enti che se ne facciano carico. Per questo ci sono enti istituiti appositamente dalla legge per fare da garanti (Garante nazionale dei diritti delle persone detenute o private della libertà personale, Autorità per le garanzie nelle comunicazioni, etc. oppure in campo privato i sindacati). Anche per la privacy in varie nazioni sono state istituite delle autorità garanti – in Italia è chiamato "Garante per la protezione dei dati personali" – però la rapidità con cui il progresso tecnologico modifica il mercato e l'invisibilità con cui questo processo di innovazione avviene, rende gli strumenti delle autorità garanti poco efficaci (Acquisti, 2016; Cranor, 2012). Molto spesso vengono prese delle misure solo quando uno scandalo scuote l'opinione pubblica proprio perché è quasi impossibile poter monitorare ciò che accade dietro le quinte di servizi governativi e aziende private. Sarebbe chiaramente impossibile e controproducente richiedere trasparenza assoluta perché si andrebbe a minare da una parte la sicurezza nazionale e dall'altra l'innovazione e il beneficio economico apportato alla società, però i rischi di un'eccessiva permissività sono altrettanto gravi e non vanno perciò sottovalutati: sorveglianza di massa e ineguaglianza sociale.

Bruce Schneier, esperto di crittografia e personaggio di spicco sulle tematiche di privacy e sicurezza nonché membro del Berkman Klein Center[1] e dell'EFF[2], scrive[3]:

> Cameras make sense when trained on police, and in offices where lawmakers meet with lobbyists, and wherever government officials wield power over the people. Open-government laws, giving the public access to government records and meetings of governmental bodies, also make sense. These all foster liberty.
> Ubiquitous surveillance programs that affect everyone without probable cause or warrant, like the National Security Agency's warrantless eavesdropping programs or various proposals to monitor everything on the internet, foster control. And no one is safer in a political system of control.

Pensare che minacce del genere siano solo remote possibilità che accadono in lontani Paesi meno sviluppati è errato. Mettendo da parte le note rivelazioni di Edward Snowden sul programma di sorveglianza di massa praticato dall'NSA e fino al 2013 categoricamente negato dalle istituzioni americane, nell'ottobre 2016 l'*Investigatory Powers Tribunal*, l'unico organo giudizario con i poteri per investigare

[1] Centro di ricerca di Harvard (Boston) che si occupa dell'esplorazione e studio dell'Internet.
https://cyber.harvard.edu

[2] Electronic Frontier Foundation. Organizzazione no-profit che si occupa della difesa dei diritti del cittadino nel mondo digitale.
https://www.eff.org

[3] *https://www.schneier.com/blog/archives/2008/03/privacy_and_pow.html*

i servizi segreti britannici, ha denunciato la raccolta segreta e illegale da parte dei servizi segreti di un enorme volume di dati confidenziali e abitudini online di cittadini inglesi, avvenuta dal 1998 al 2015, ovvero dall'inizio del programma alla sua divulgazione pubblica, per ben 17 anni senza adeguate salvaguardie o supervisioni[4]. Ancora, nel novembre 2016, è stato rivelato che il corpo di polizia di Montreal, Canada, stava monitorando i movimenti e i messaggi del giornalista de *La Presse* Patrick Lagacé per scoprire l'identità delle sue fonti; un'ulteriore investigazione ha poi reso pubblico che in realtà le intercettazioni coinvolgevano non uno ma ben sei giornalisti[5].

Nel settore privato un sistema incontrollato può diventare terreno fertile per discriminazioni e diseguaglianze sociali ed economiche. Agli inizi di novembre 2016 Admiral, una delle più grandi compagnie assicurative del Regno Unito, aveva annunciato che avrebbe iniziato un programma speciale dove i costi delle assicurazioni venivano calcolati e decisi da un algoritmo basato sull'attività di Facebook dei richiedenti[6]. Già il giorno seguente l'annuncio, la compagnia ha però dovuto fare marcia indietro dopo che Facebook ha minacciato di chiudere i rapporti ricordando che le sue policy proibivano espressamente l'uso dei dati personali "per prendere decisioni di eleggibilità"[7]. Senza una risposta concreta per proteggere i consumatori, altre aziende in futuro percorreranno strade simili e per le persone rimarrà sempre meno possibilità di scelta, soprattutto quando lo svantaggio economico di non aderire a questo tipo di iniziative diventa così grande che consegnare i propri dati personali è l'unica alternativa.

Forse la difficoltà più grande sta nell'impercettibilità con cui le aziende online possono apportare cambiamenti ai loro servizi unita all'influenza che quest'ultimi possono avere nel quotidiano: il 25 marzo 2014 la PNAS[8], la rivista ufficiale americana dell'Accademia Nazionale delle Scienze, ha pubblicato un paper scritto da alcuni ricercatori affiliati a Facebook[9] e che ha fatto infuriare l'intera comunità accademica[10]. L'articolo riportava come fosse possibile influenzare lo stato d'animo di migliaia di persone semplicemente facendo in modo che l'algoritmo che controlla il *News Feed* del social network filtrasse solo post positivi o solo negativi. Le accuse principali della comunità accademica erano quelle di non aver rispettato alcun principio etico perché i partecipanti all'esperimento (689.003 persone) erano ignari di essere stati usati nell'esperimento e perché il processo di revisione del paper, nonostante fosse stato pubblicato su una rivista importante, avesse seguito un percorso molto particolare. Questo era il primo studio reso pubblico fatto da/con Facebook, ma nessuno può sapere quanti altri ne potrebbero essere stati fatti nell'inconsapevolezza generale.

4 ■ *https://www.theguardian.com/world/2016/oct/17/uk-security-agencies-unlawfully-collected-data-for-decade*

5 ■ *http://montrealgazette.com/news/local-news/police-surveillance-scandal-montreal-to-study-spying-issue-behind-closed-doors*

6 ■ *https://www.theguardian.com/technology/2016/nov/02/admiral-to-price-car-insurance-based-on-facebook-posts*

7 ■ *https://www.theguardian.com/money/2016/nov/02/facebook-admiral-car-insurance-privacy-data*

8 ▲ Proceedings of the National Academy of Science. È una delle riviste scientifiche multidisciplinari più rilevante al mondo.
http://www.pnas.org

9 ■ *http://www.pnas.org/content/111/24/8788.full*

10 ■ *https://www.theguardian.com/technology/2014/jun/30/facebook-emotion-study-breached-ethical-guidelines-researchers-say*

Aldilà dell'efficacia opinabile dei risultati mostrati nell'esperimento, esso dimostra la facilità con cui gli algoritmi che fanno parte della vita quotidiana non siano strumenti imparziali e oggettivi ma possano essere manipolati per modificare la percezione della realtà.

Per contrastare la crescente asimmetria che minaccia i diritti dei cittadini non possiamo affidarci unicamente alla protezione di misure legislative, per l'inadeguatezza della macchina burocratica a stare al passo col progresso tecnologico, né possiamo rimetterci alla sola auto-regolamentazione di un mercato che, anche se più flessibile ai cambiamenti, alla fine farà sempre gli interessi economici dei più forti. È perciò necessario un ponte tra questi due estremi e le iniziative dal basso, dai cittadini, lo sono: esse possono colmare il vuoto dove la legge non può arrivare con la flessibilità dell'innovazione.

A questo lavorano attivamente numerose organizzazioni e associazioni di attivisti per la difesa dei diritti digitali tra cui Privacy International[11], Electronic Frontier Foundation (EFF), Tactical Tech Collective[12], Freedom House e European Digital Rights (EDRi)[13].

[11] *https://privacyinternational.org*

[12] *https://tacticaltech.org*

[13] *https://edri.org*

3.1 Rispondere all'asimmetria: due approcci

Per farsi un'idea migliore su cosa sia già stato fatto nel passato attraverso iniziative dal basso, un valido punto di partenza è ancora una volta il Web. Usando il termine "dal basso" si vogliono intendere tutte quei progetti fatti da associazioni, gruppi di attivisti, redazioni giornalistiche, gruppi di ricerca, cittadini singoli, etc. che si occupano di tenere vivo il dibattito sulla privacy e che non sono legati a governi e aziende del settore (iniziative dall'alto).

Durante il periodo di ricerca della tesi, sono state raccolte e mappate diverse soluzioni anche molto diverse tra loro, che andavano da software a guide di buone pratiche ma anche font, opere d'arte e progetti di provocazione. Man mano che venivano catalogate, tutte queste soluzioni apparivano essere in qualche modo riconducibili a due sottogruppi, con due approcci completamente diversi: progetti per celarsi o nascondersi e progetti per contrastare l'asimmetria di potere attraverso la trasparenza e la consapevolezza dell'utente. Essendo stati validi spunti e ispirazioni per la parte progettuale, le prossime pagine riporteranno alcuni dei casi studio trovati, suddivisi secondo i due principali approcci. Un'ulteriore e più ampia mappatura dei progetti raccolti può essere poi trovata sull'apposita pagina web[14].

[14] *https://micheleinvernizzi.com/privasea*

NASCONDERSI

In questo gruppo sono stati raccolti progetti il cui obiettivo è quello di occultare il più possibile le proprie informazioni personali. Molti di questi strumenti sono stati creati per persone la cui sicurezza personale, o addirittura la vita, sono messe a repentaglio per via delle informazioni in loro possesso. Informatori, attivisti, perseguitati politici, utilizzano modalità per rimanere anonimi e preservare così la loro incolumità. Sempre più vengono utilizzati anche da altre persone per fuggire da ciò che essi sentono come una raccolta indiscriminata e ingiusta.

Molte di queste soluzioni propongono alternative open-source[15] guidate da un trattamento responsabile dei dati o dalla totale assenza di tracciamento. Questo vuol dire ad esempio che i dati raccolti sull'utente vengono anonimizzati, cancellati dopo brevi periodi, analizzati solo in aggregati e mai condivisi con terze parti, oppure addirittura che nessuno dato o metadato[16] associato ad un utente viene mai mantenuto sui server. L'esempio più famoso è forse Tor[17], acronimo di The Onion Router ("il router a cipolla"), un sistema di comunicazione per Internet che rende il tracciamento degli utenti molto più difficile grazie all'incapsulamento della comunicazione in diversi strati di crittografia – da cui nasce il riferimento alla cipolla.

Una variante di questo approccio è chiamato *data obfuscation*, ovvero la modifica e la deformazione dei dati personali in modo da distorcere l'immagine che gli altri si fanno di una persona. Questa operazione viene fatta anche platealmente, come atto politico, ad esempio riempiendo la propria cronologia di ricerca di migliaia di pagine di materiale contraddittorio e illegale[18], in una sorta di grido di protesta contro i cosiddetti data brokers, aziende il cui lavoro è aggregare tutti i dati rintracciabili delle persone per mettere ognuno in una categoria specifica e infine vendere le informazioni alle agenzie pubblicitarie – o a chiunque paghi.

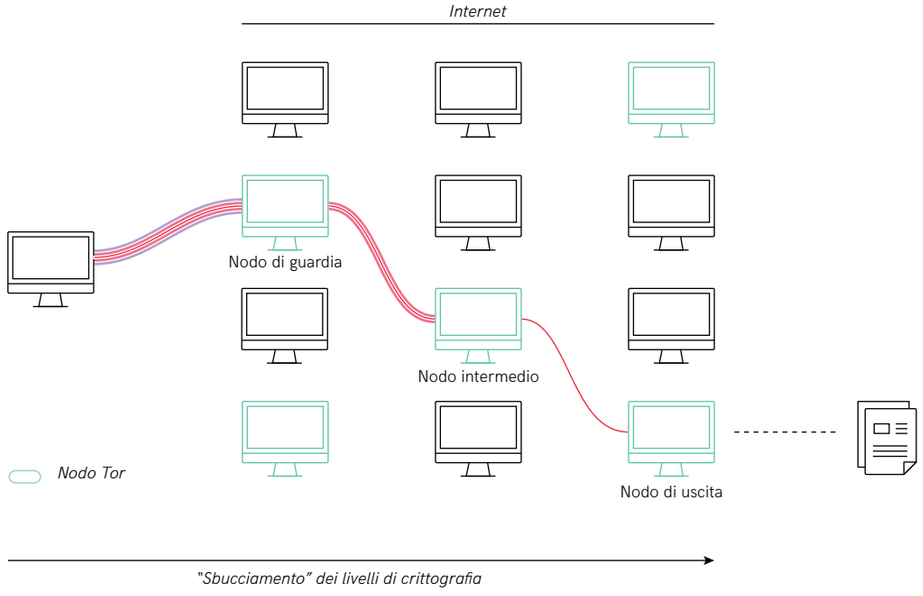


FIG. 19 Rappresentazione schematica della pratica dell'*onion routing* su cui si basa Tor. Ogni nodo ha contatti solo col nodo precedente e con quello successivo e non ha quindi modo di sapere l'intero percorso.
Ispirazione: https://www.bof.nl/onswerk/internetvrijheid-toolbox

15 ● Codice libero e aperto a modifiche ed estensioni. *È come fare una torta per un amico, e poi donargli la ricetta. La torta è il codice, solo più buone, e la condivisione della ricetta è la parte open source.*
(fonte: Sideways Dictionary)

16 ● Informazione che descrive un insieme di dati. *È come tutte le informazioni sulla busta di una lettera. Anche se non mostra il contenuto può dire molto – da dove viene, a chi è indirizzata, quanto pesa. Da lì, si può iniziare a inferire sul contenuto – se viene da una banca, probabilmente non è una lettera d'amore.*
(fonte: Sideways Dictionary)

17 ■ *https://www.torproject.org*

18 ▲ Un esempio è *Ruin my search history*.
http://ruinmysearchhistory.com

CASO STUDIO

Security in-a-box

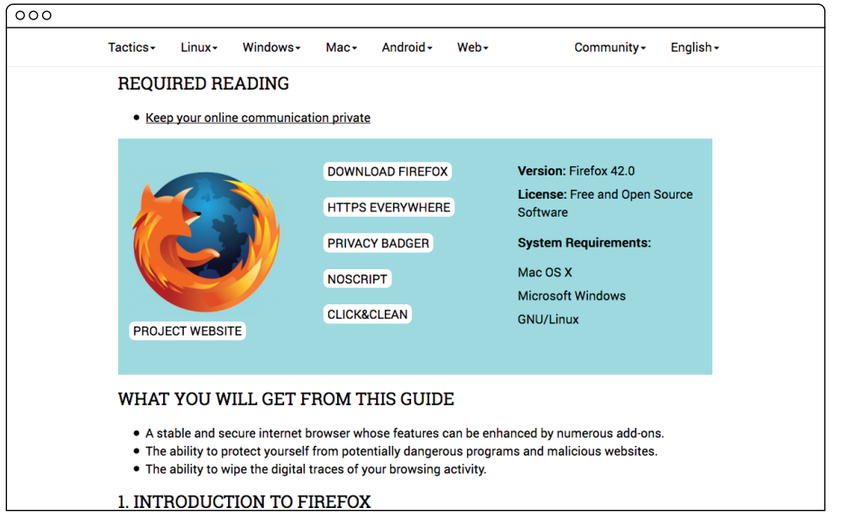
Progetto sviluppato da Front Line Defenders[19] e Tactical Technology Collective[20], Security in-a-box è una raccolta selezionata di strumenti e buone pratiche per rimanere anonimi e protetti mentre si naviga online. Guide e strumenti sono catalogati per sistema operativo (Linux, Windows, Mac OS, Android) in modo da coprire tutte le possibili piattaforme usate comunemente. L'obiettivo del progetto è quello di far conoscere ad un pubblico più ampio le tecnologie normalmente usate da giornalisti ed attivisti politici che necessitano di misure per salvaguardare la propria incolumità.

https://securityinabox.org/en

[19] *https://www.frontlinedefenders.org*

[20] *https://tacticaltech.org*

IMG. 04-05 Screenshot estratti dal sito di security in-a-box.

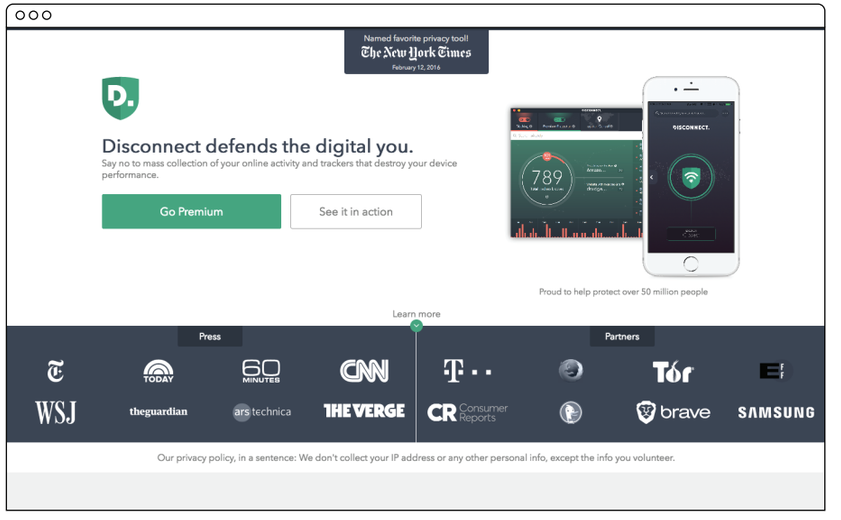


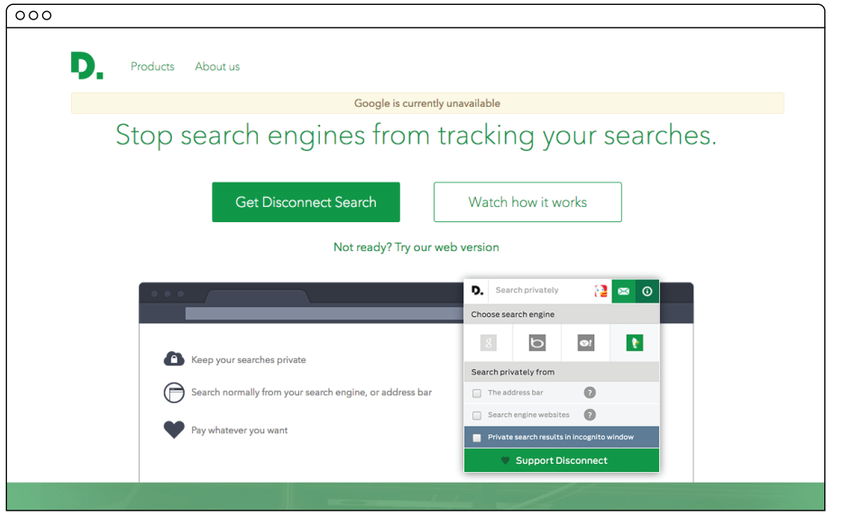
CASO STUDIO

Disconnect

Prodotto dall'omonima compagnia fondata nel 2011 specificatamente per produrre strumenti per proteggere la privacy degli utenti del web, Disconnect offre diverse funzionalità: a seconda del pacchetto scelto il tool va dall'essere un semplice filtro anti-pubblicità e motore di ricerca senza tracciamento (gratuitamente), fino ad offrire un servizio di VPN[21] dedicato e livelli di sicurezza aggiuntivi alla connessione Wi-fi (50 dollari all'anno).

https://disconnect.me





IMG. 06-07 Screenshot estratti dal sito di disconnect.me.

21 • Virtual Private Network. Si tratta di una rete privata e cifrata ma che usa un protocollo di trasmissione pubblico (come Internet). *È come essere insieme ad un altra persona, gli unici due a parlare un determinato linguaggio. Potreste urlarvi cose segrete da una parte all'altra di una stanza affollata e nessuno capirebbe ciò che vi state dicendo. (fonte: Sideways Dictionary)*

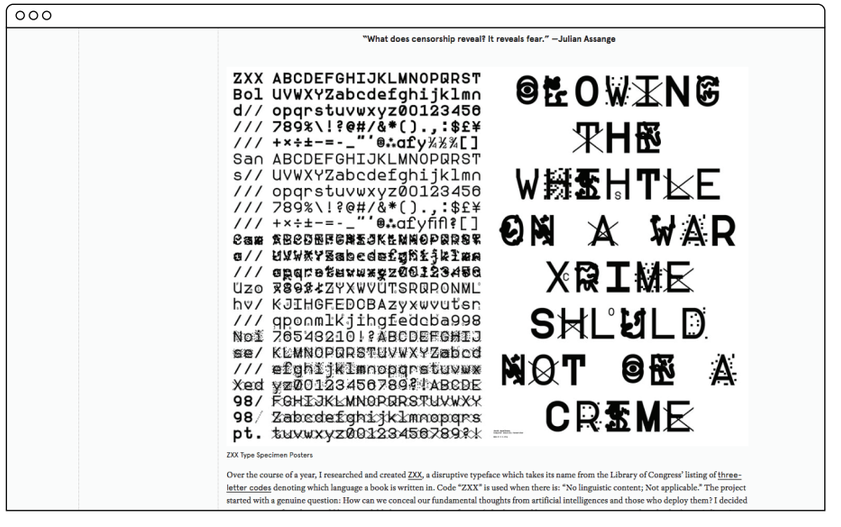
CASO STUDIO

ZXX

ZXX è un font sviluppato nel 2013 dal designer coreano Sang Mun[22]. Il font prende il nome dal codice con cui la Libreria del Congresso etichetta i libri senza contenuto linguistico ed ha la caratteristica di non poter essere riconosciuto da software di scansione del testo. È un interessantissimo caso di come il Design può contribuire al dibattito sulla privacy.

http://blogs.walkerart.org/design/2013/06/20/sang-mun-defiant-typeface-nsa-privacy

22 ■ *http://sang-mun.com*



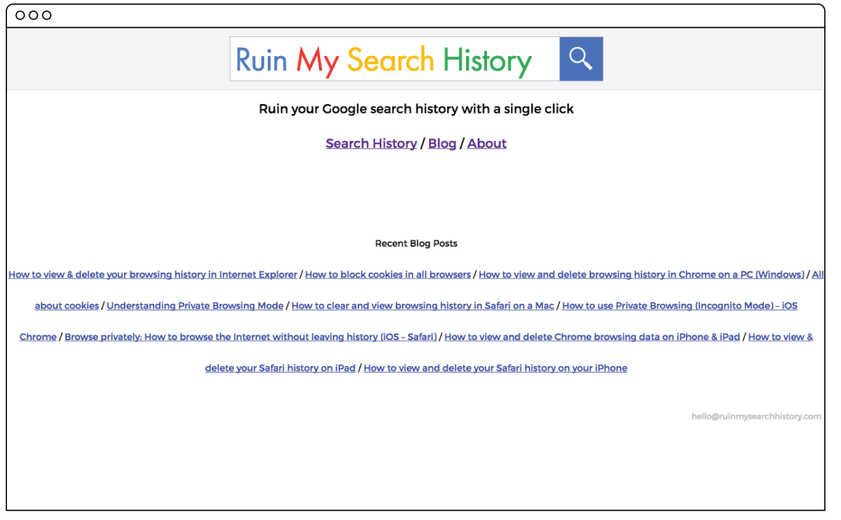
IMG. 08-09 Screenshot estratti dal blog della Walker Art Center relativi al progetto ZXX.

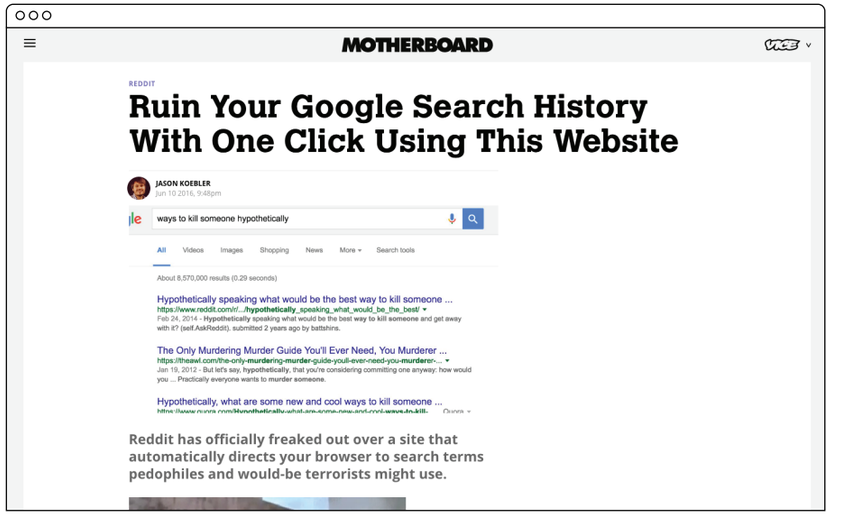
CASO STUDIO

Ruin My Search History

Esempio di *data obfuscation*, Ruin My Search History permette di inondare la propria cronologia di ricerca con migliaia di siti delle più svariate tipologie, molti dei quali anche parecchio compromettenti. L'idea è che un sovraccarico di informazioni chiaramente falsificate renda la profilazione da parte delle aziende inutilizzabile.

http://ruinmysearchhistory.com





IMG. 10-11 Screenshot estratti dal sito di Ruin My Search History e da un articolo su Motherboard relativo all'estensione per browser.

CASO STUDIO

AdNauseam

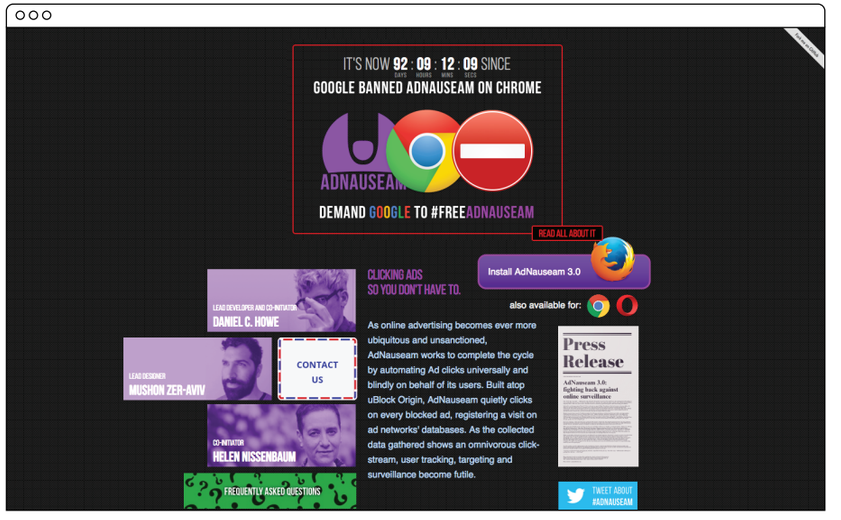
Altro caso di *data obfuscation* progettato come atto di protesta contro la profilazione online, AdNauseam è un'estensione per browser[23] che blocca tutte le pubblicità che verrebbero altrimenti mostrate mentre si naviga. A differenza di un normale ad-blocker[24] però, AdNauseam clicca ogni pubblicità che blocca, in modo che non sia possibile creare una profilazione reale dell'utente.

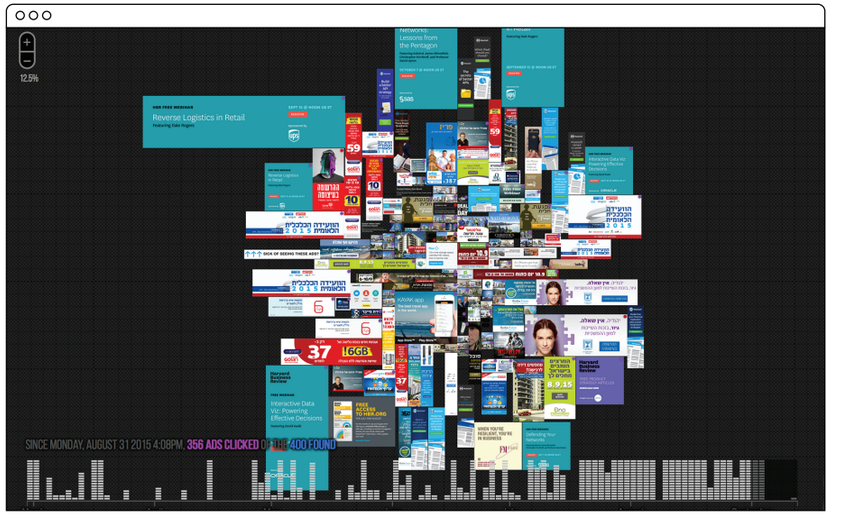
https://adnauseam.io

23 • È l'applicazione usata per navigare sul Web. *È come una televisione, prende il segnale e lo trasforma in immagine in modo che tu possa vederlo.* *(fonte: Sideways Dictionary)*

24 • Qualsiasi estensione usata per bloccare la pubblicità. *È come se la televisione fosse in grado di eliminare la pubblicità appena ne rileva una.*

IMG. 12-13 Screenshot estratti dal sito di AdNauseam.





PORTARE TRASPARENZA

Il secondo approccio raccoglie invece tutti quei progetti atti a educare l'opinione pubblica sul funzionamento del mondo della raccolta dati, portando alla luce dei meccanismi altrimenti nascosti e creando trasparenza in processi altrimenti opachi. L'idea è che una maggior consapevolezza nell'utilizzo dei servizi online e più in generale della tecnologia che ci circonda, possa permettere di instaurare un dibattito più produttivo per trovare soluzioni più efficaci ad un sistema economico che esclude il possessore dei dati da qualsiasi partecipazione (Acquisti, 2016).

I progetti raggruppati in questa categoria si prefiggono di creare quella conoscenza di base necessaria per comprendere un sistema complesso e ricco di tecnicismi, oltre a voler mostrare come i propri dati vengono presi ed elaborati, sottolineando come non sia un problema distante ma tocchi tutti in prima persona. A volte sono proposti come provocazioni, ad esempio sotto forma di giochi come in Snitch Hunt[25], dove l'utente diventa analista e deve scoprire più informazioni possibili su alcuni individui, scoprendo allo stesso tempo le reali potenzialità di possedere metadati[26] su un gruppo di persone. Altre volte utilizzano i dati di chi li prova per spiegare il funzionamento di tecnologie con cui si interagisce a volte inconsapevolmente. Ne è un esempio Predictive World[27], dove connettendosi con Facebook o fornendo alcuni dati personali – che verranno cancellati alla fine della sessione – viene mostrato all'utente cosa potrebbe generare un algoritmo predittivo in base alle informazioni personali analizzate, stime che vanno dall'aspettativa di vita, alla probabilità di essere un fumatore di marijuana o alla stabilità emotiva. Anche se l'algoritmo in questione è stato creato dal Centro Psicometrico dell'Università di Cambridge[28], dà una misura degli algoritmi che vengono usati nella vita reale per prendere decisioni sulle persone.

[25] *https://snitchhunt.org*

[26] • Informazione che descrive un insieme di dati. *Sono come le informazioni nutrizionali su una barretta di cioccolato.*
(fonte: Sideways Dictionary)

[27] *http://predictiveworld.watchdogs.com*

[28] *www.psychometrics.cam.ac.uk*

CASO STUDIO

Data Selfie

Data Selfie è un'estensione per browser[29] che traccia i propri utenti mentre sono su Facebook per poi restituire loro una visione di come le tracce lasciate sul social network potrebbero essere usate da un algoritmo per generare inferenze sulla loro personalità. L'estensione è un progetto sviluppato da Hang Do Thi Duc, studentessa di design alla Parsons School of Design, insieme a Regina Flores Mir, ricercatrice per la stessa università. Sponsorizzato dal NYC Media Lab Combine Program[30], Data Selfie sfrutta per le sue analisi le API di Apply Magic Sauce[31] e IBM Watson[link], due strumenti di machine learning rispettivamente dell'Università di Cambridge e della IBM.

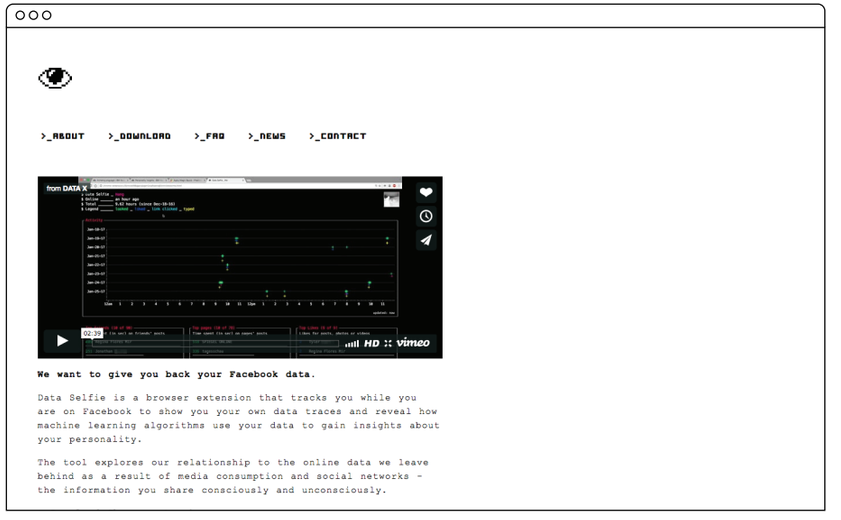
http://dataselfie.it

[29] ● È l'applicazione usata per navigare sul Web. *È come un'automobile, che permette di muoversi per lunghe distanze e in autostrada.*
(fonte: Sideways Dictionary)

[30] ■ *http://www.thecombine.nyc*

[31] ■ *http://applymagicsauce.com*

IMG. 14-15 Screenshot estratti dal sito di Data Selfie e da un articolo di The Next Web sullo strumento.

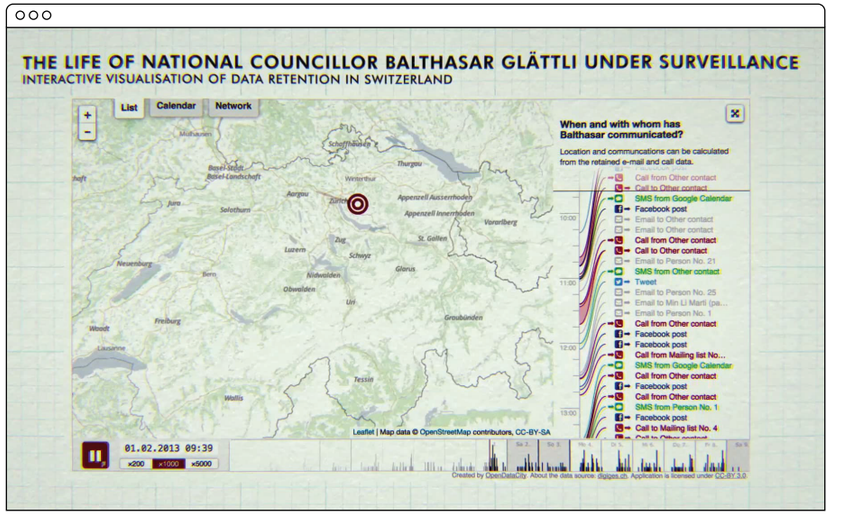


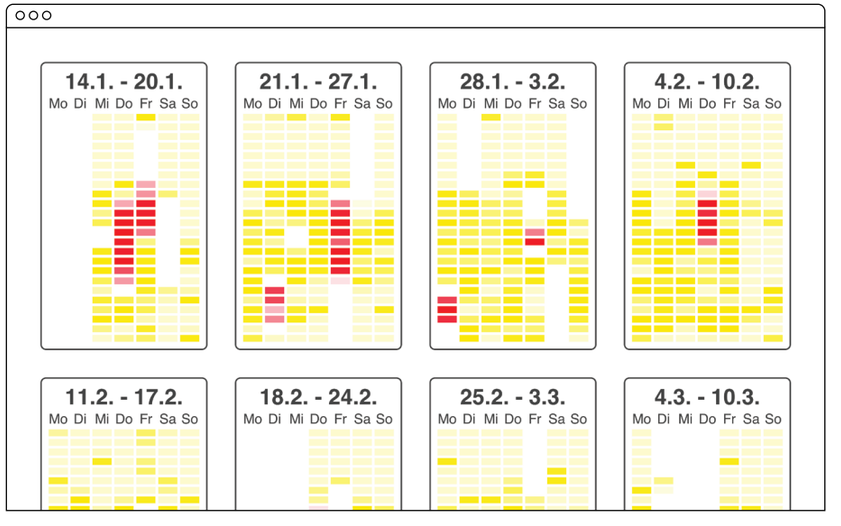
CASO STUDIO

Data retention in Switzerland

Nel 2013 Balthasar Glättli, membro del Consiglio Nazionale per il partito dei Verdi in Svizzera, aveva richiesto ed ottenuto 6 mesi di metadati tenuti dalla sua compagnia telefonica (chiamate, messaggi e connessioni a Internet) e dal suo provider di email. Senza dover sapere nulla del contenuto, queste informazioni sono sufficienti per seguire Balthasar in ogni suo movimento per tutto l'arco dei 6 mesi ed avere un'idea molto chiara delle persone con cui è stato a contatto, nonché delle sue abitudini, dove abita e dove lavora.

https://apps.opendatacity.de/vds/index_en.html





IMG. 16-17 Screenshot estratti dal sito di Open Data City relativo ai dati raccolti da Balthasar Glättli.

CASO STUDIO

Clickclickclick.click

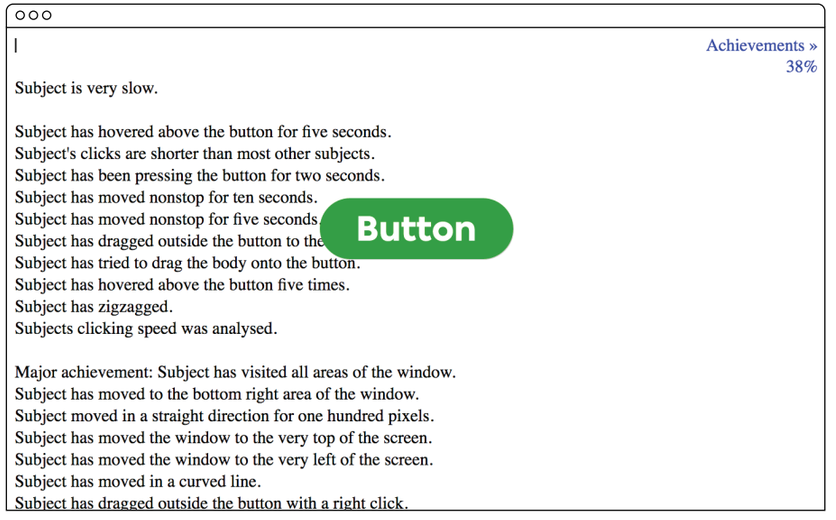
Nato dalla collaborazione di tre studi di design olandesi (Moniker[32], Studio Puckey[33] e VPRO Medialab[34]), il sito è una divertente provocazione che rivela però la quantità di informazioni che è possibile estrarre durante la navigazione online con un browser. Tutto il processo di richiesta ed estrazione dati sarebbero completamente invisibili all'utente se non ci fosse una fastidiosa voce che le elenca tutte. Numero di click, lunghezza di scroll, quantità di tempo passata senza muovere il mouse, passaggio da un tab all'altro del browser, tutto è registrato e l'utente non è mai lasciato solo.

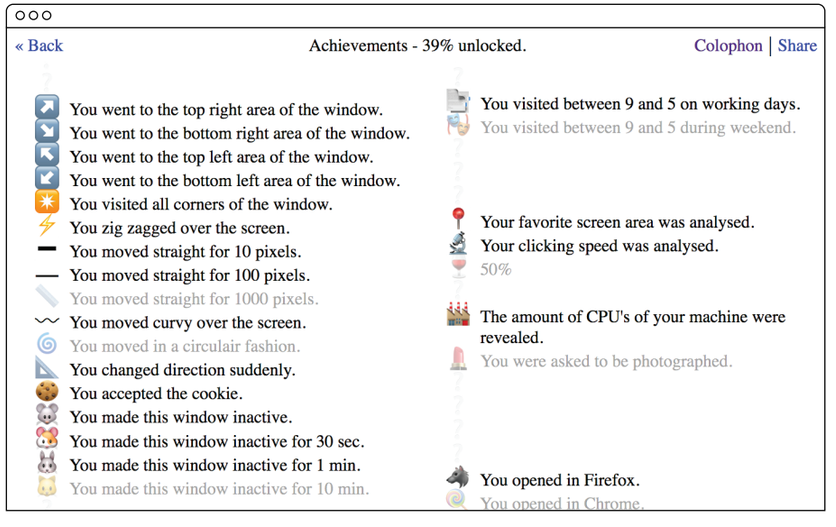
http://clickclickclick.click

32 *https://studiomoniker.com*

33 *http://puckey.studio*

34 *https://www.vpro.nl/medialab*





IMG. 18-19 Screenshot estratti dal sito di Clickclickclick.click.

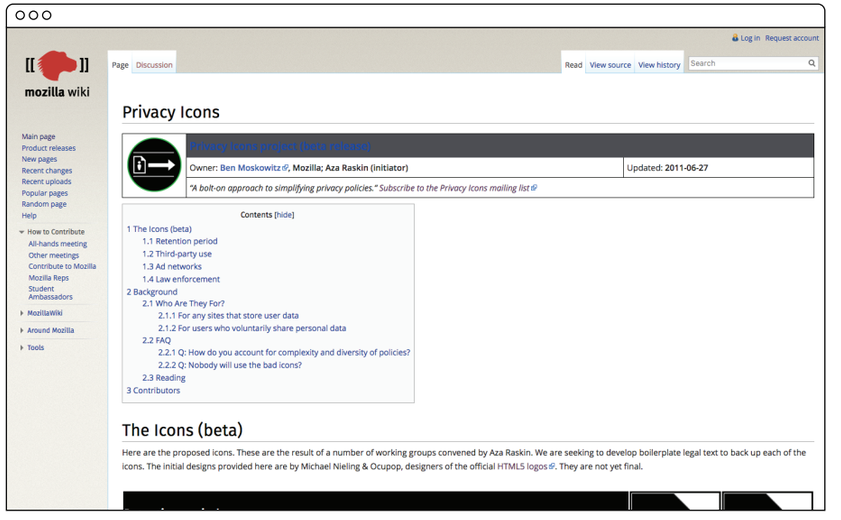
CASO STUDIO

Privacy Icons

Risultato di una serie di workshop organizzati da Mozilla, le Privacy Icons sono un set di icone disegnate da Michael Nieling & Ocupop in collaborazione con Aza Raskin e con l'obiettivo di categorizzare e semplificare varie voci che si trovano all'interno delle privacy policies di siti e applicazioni, al fine di renderle più facilmente comprensibili da parte degli utenti. Descrivono il periodo di conservazione dei dati, ovvero quanto tempo un'azienda intende mantenere una copia delle informazioni prima di eliminarle dai propri server, le modalità di condivisione con terze parti, la volontà di condividere i dati con le autorità quando richiesto, etc.

Anche se le icone non sono mai state implementate per vari problemi pratici oltre che alla resistenza delle aziende, sono un interessante tentativo di superare l'inefficacia delle privacy policies con un approccio che mette il Design come fulcro dell'iniziativa.

https://wiki.mozilla.org/Privacy_Icons



IMG. 20-21 Screenshot estratti dal sito relativo al progetto delle Privacy Icons.

CASO STUDIO

Am I unique?

Am I unique? è un progetto sviluppato da un team di ricercatori dell'Istituto Nazionale di Scienze Applicate di Rennes[35] con l'obiettivo di rendere gli utenti consapevoli dello stato dell'arte nel mondo del tracciamento online, in particolare su come è possibile riconoscere un individuo senza avere il suo nome e cognome ma semplicemente identificando univocamente il browser che utilizza. Questo metodo si chiama *fingerprinting*[36] ed è una pratica sempre più diffusa sul Web, nata per ovviare alla nascita di estensioni per browser che bloccavano o eliminavano i *cookies*[37] delle aziende.

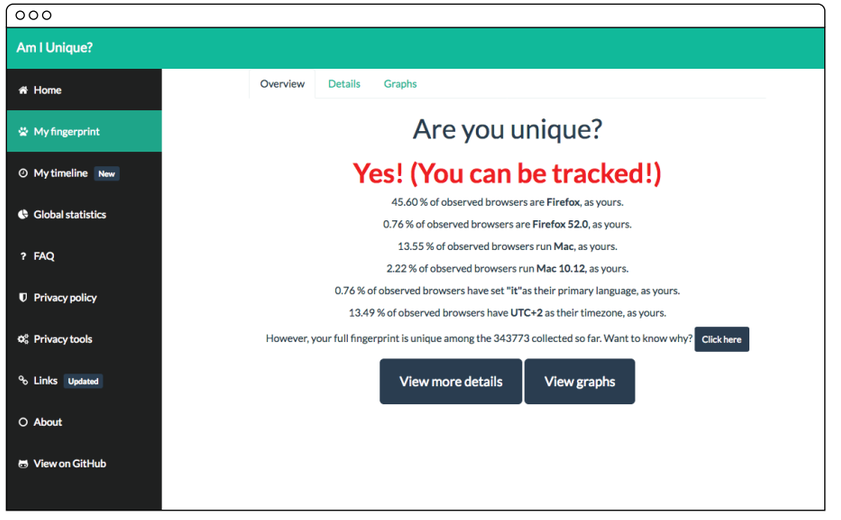
https://amiunique.org

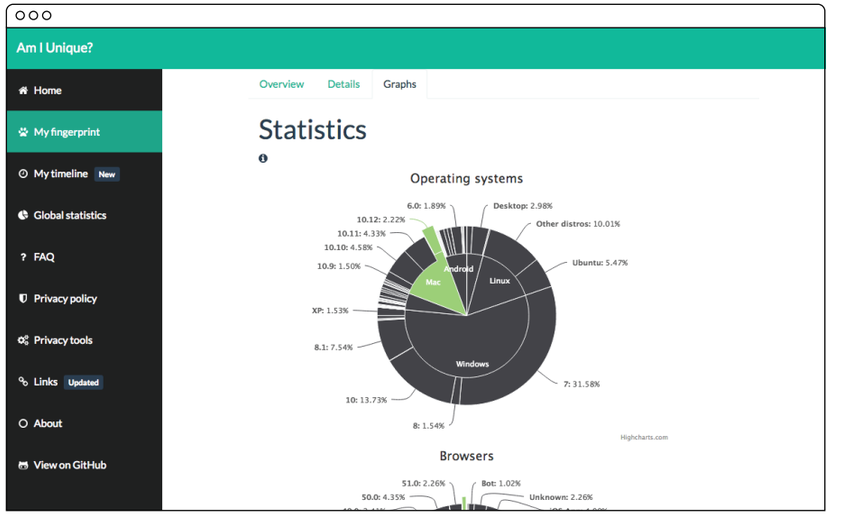
[35] *https://www.insa-rennes.fr/en.html*

[36] Letteralmente l'impronta digitale del dispositivo. Se si riescono a raccogliere abbastanza informazioni su un dispositivo, esse lo renderanno abbastanza unico da poter essere tracciato per tutto il Web. *Sono come le impronte sulla neve, non ti identificano personalmente ma possono dire tanto su chi sei - la fattura della scarpa, il numero, la lunghezza del passo, la strada che hai percorso.*
(fonte: Sideways Dictionary)

[37] Piccolo file di testo salvato nel browser per salvare le preferenze dell'utente. *È come unbarista con un'ottima memoria: ogni mattina quando entri ti chiederà 'Il solito?'.*
(fonte: Sideways Dictionary)

IMG. 22-23 Screenshot estratti dal sito Am I Unique?.





Raffigurare non vuol dire soltanto replicare il visibile, più o meno fedelmente, più o meno schematicamente. Raffigurare vuol dire anche mostrare l'invisibile. Mostrare l'invisibile a sua volta non vuol dire soltanto rendere visibile strumentalmente ciò che si presenta nell'esistente, ma vuol dire anche costruire figure e modelli visibili del possibile, del probabile, dell'ipotetico.

Giovanni Anceschi

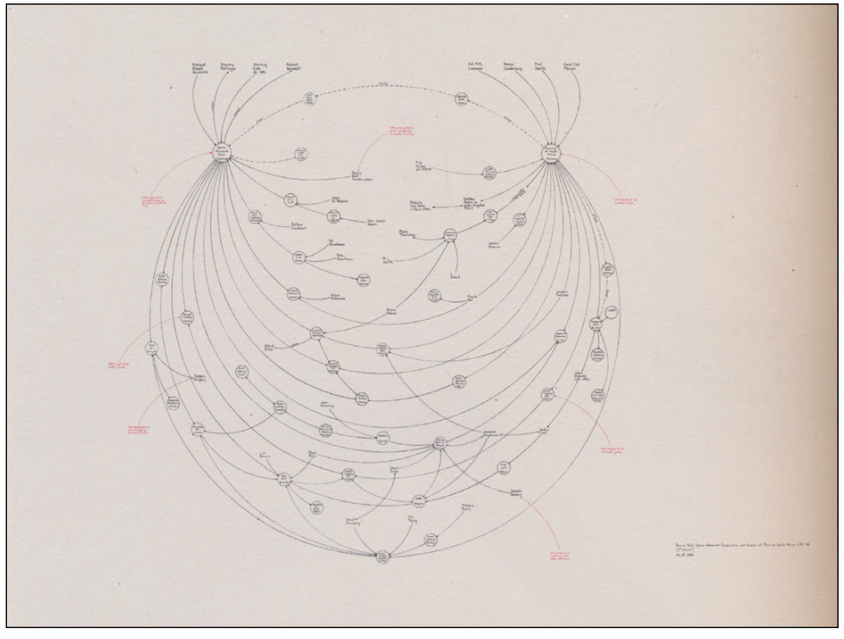
4 — Il ruolo del designer e della visualizzazione

In un contesto complesso e ricco di opacità, la figura del designer della comunicazione diventa un valido sostegno per il lavoro dei vari attori impegnati a coinvolgere un pubblico più ampio su tematiche rilevanti ma opache come quella della privacy. Sempre di più varie discipline e campi di ricerca si trovano a dover lavorare insieme perché i problemi affrontati necessitano competenze molto distanti tra loro: nello specifico dell'information privacy informatici, giornalisti, avvocati e molti altri devono integrare i loro background per generare nuova conoscenza (Buchanan, 1992). I confini delle varie discipline perdono di definizione e tendono a mischiarsi, in una sorta di fluidità professionale (Bertola & Manzini, 2004). Il designer però serve a connettere e unire i diversi ambiti disciplinari, mediando continuamente tra gli elementi del contesto e la diversità degli attori coinvolti (Baule & Caratti, 2016). Già nel 1992, Giovanni Anceschi scriveva ne "L'oggetto della raffigurazione":

> Il progetto è condannato alla multidisciplinarietà: anche quando non lo sa, è globale, è sempre design totale. (Anceschi, 1992)

Ora più che mai questa affermazione è vera ed il designer si colloca nella posizione chiave di traduttore (ibidem) non solo della resa finale ma dell'intero processo progettuale.

L'interdisciplinarietà è però in fondo una conseguenza della complessità dei fenomeni che ci si trova a dover affrontare e il designer è chiamato a ideare soluzioni per mappare le relazioni in gioco e rappresentarne visualmente le informazioni (Venturini, 2008). La "comunicazione di informazioni in modo significativo" (Meirelles, 2013) è il focus dell'information design e le visualizzazioni, o più in generale i diagrammi, sono gli strumenti con cui il designer può progettare la creazione di nuova conoscenza. Attraverso la visualizzazione si deve poter rivelare la complessità, rendendola trasparente, ma allo stesso tempo ridurla per renderla leggibile e l'abilità del progettista sta nel bilanciare queste due qualità (Schoffelen et al., 2015). Masud et al. descrivono inoltre come le visualizzazioni siano in grado di unire dati, informazioni o conoscenze, visualizzandoli sotto forma di informazioni in un artefatto comunicativo, per creare infine, se ben progettate, nuove conoscenze nel ricevente (Masud et al., 2010). Ogni visualizzazione, secondo gli autori, è una trasformazione nel continuum dato-informazione-conoscenza e insieme rappresentano il processo di trasformazione dei dati in conoscenza.



IMG. 24 Uno dei diagrammi di Mark Lombardi, creato per spiegare le connessioni in sistemi di relazioni che lui definiva "di uso e abuso di potere".
Fonte: Ben Fry

Nonostante la positività di queste premesse, ad oggi la figura del designer è quasi totalmente assente nei progetti che si occupano di privacy online: in parte per la relativa novità del discorso e in parte per via della forte complessità tecnologica, le soluzioni pratiche per far fronte a questo problema sono guidate da figure con una forte competenza tecnica, sociale o giuridica, trovandosi in una situazione spesso già vista in altri campi: il "design senza i designer" (Caviglia, 2013; Burdick, 2009). Dovendo fare i conti con analisi di traffici di rete, decifrazione di algoritmi e manipolazione di codice, è naturale che i primi progettisti siano ingegneri informatici (ricercatori e non), sviluppatori

software e attivisti con un solido background di programmazione.

La componente tecnologica è il fattore chiave e a volte anche l'unico considerato, finendo col progettare soluzioni che non tengono in considerazione un possibile utente finale ma solo i bisogni del progettista. Avviene così che, usando le parole di Anne Burdick:

> [...]"design" is not the discipline that we know and love—that is, it's not the province of design practitioners, researchers, and educators. Instead, "Design" is variably a value-add, an everyday event, a working method, a byproduct, a literacy, and a complete abstraction. And frequently designers are nowhere to be found. (Burdick, 2009)

È fondamentale perciò che i designer entrino come parte integrante del progetto fin dall'inizio del processo progettuale e per farlo devono comprendere il contesto specifico in cui operano e adattare i propri metodi affinché risultino efficaci. Questa tesi si prefigge quindi di testare l'adattabilità dei metodi propri della professione del designer al contesto della privacy online, sondandone le specificità.

SPERIMENTAZIONI PROGETTUALI

It is such an agreeable feeling to be busy with something one is only half-competent to do that nobody should criticize the dilettante for taking up an art he will never learn, or blame the artist who leaves the territory of his own art for the pleasure of trying himself in a neighbouring one.

Johann Wolfgang von Goethe

5 — Metodo e approccio sperimentale

La parte di progetto di questa tesi è partita con l'obiettivo di applicare i metodi conosciuti – ad esempio per la raccolta, manipolazione e trasformazione visuale dei dati – ad un contesto nuovo – quello della privacy – e testarne la validità. Come si possono mappare processi invisibili, coperti da un velo di complessità tecnologica e, a volte, di barriere legali? Come queste caratteristiche influenzano il processo progettuale? Gli strumenti e metodi di cui il designer dispone sono ancora adeguati o vanno riadattati? Ne servono di nuovi?

Per rispondere a queste domande verrà utilizzato un approccio sperimentale, che parte dall'osservazione della realtà, o di ciò che di essa si può vedere, permettendo allo stesso tempo di generare riflessioni sull'evoluzione di cultura e pratica del design (Bertola & Manzini, 2014).

Il contesto analizzato è peculiare perché qui non c'è un attore o cliente che è disposto a fornire i dati necessari alla progettazione, né è aperto ad un dialogo col designer. Viceversa l'accesso ai dati è spesso ostacolato, e i processi con cui le informazioni personali degli utenti sono raccolte e

utilizzate dalle aziende vengono protetti da numerose barriere tecnologiche e legali. Questo accade sia per questioni di sicurezza, trattandosi di dati sensibili, sia per difendere i propri asset economici – essendo dopotutto la risorsa indispensabile per mantenere in vita quest'industria. Un designer che vuole rappresentare fenomeni di questo tipo perciò, deve poter superare l'opacità generata da dati e processi tenuti volutamente nascosti. Il processo progettuale deve inevitabilmente adattarsi, sfruttando la collaborazione interdisciplinare con gruppi di impronta fortemente tecnologica e dovendo a volte anche confrontarsi con situazioni che richiedono soluzioni al limite della legalità e che portano a riflessioni di carattere etico.

Nei prossimi capitoli verranno perciò illustrate due differenti sperimentazioni progettuali sviluppate durante il periodo di tesi. Ogni esperimento indaga uno degli aspetti opachi introdotti nel capitolo 2, ovvero:

- L'impossibilità di capire la portata reale dell'accesso alle informazioni personali – come e quanto spesso viene richiesto nonché a chi viene venduto;

- La difficoltà di poter valutare come i dati personali vengono utilizzati per influenzare la vita delle persone e prendere decisioni in loro vece.

Cercando di proporre soluzioni pratiche a problemi reali, è possibile far emergere i limiti e i punti di forza della propria metodologia.

Jessica Fletcher: "Non capisco, come ha fatto a sapere che ero qui se non è stato Guzman a dirglielo?"

Agente FBI Bartles: "Oggi, signora Fletcher, facciamo tutti parte di un'unica e felice famiglia elettronica".

La signora in giallo –
Vagone letto con omicidio (1997)

6 — Esperimento n° 1: ReCon

La prima sperimentazione progettuale nasce dalla collaborazione tra il laboratorio di ricerca DensityDesign e il Data Transparency Lab[1] (abbreviato DTL), un centro interistituzionale fondato da vari partner tra cui Mozilla e Telefonica nel 2014 al fine di creare una comunità solida e attiva tra università, aziende ed istituzioni che lavorasse al miglioramento della trasparenza nell'uso dei dati personali online. DTL finanzia ogni anno 6–8 progetti provenienti dal mondo accademico per lo sviluppo di software che contribuiscano in qualche modo alla consapevolezza del pubblico sulla privacy, ad esempio quantificando e identificando i metodi di tracciamento online, individuando pratiche di discriminazione basate sulla posizione geografica, facendo *reverse engineering* sugli algoritmi che inseriscono pubblicità nei siti web.

Uno dei progetti finanziati è ReCon, un'app che permette di monitorare e controllare tutto il traffico di dati in uscita dal nostro smartphone in modo da poter avere un riscontro reale su quali informazioni personali vengono richieste dalle applicazioni installate, quante volte ven-

[1] *www.datatransparencylab.org*

gono prese e a chi vengono inviate (una spiegazione più dettagliata del suo funzionamento può essere trovata più avanti, → CAP. 6.2). L'obiettivo iniziale della collaborazione era quello di utilizzare i dati raccolti da ReCon durante i test di validazione del software per produrre alcune visualizzazioni su questo passaggio di informazioni. All'obiettivo pratico si aggiunge quello metodologico: testare quanto del fenomeno sia effettivamente possibile sondare con gli strumenti disponibili e supportati da competenze tecnologiche specifiche, dato che la trasmissione e ricezione dei dati sono solitamente invisibili e protette.

Dopo una breve descrizione del gruppo di persone che hanno partecipato al progetto e un'introduzione ad alcuni elementi necessari a comprendere meglio il contesto, verranno illustrati il processo progettuale e i risultati ottenuti, sempre nell'ottica di usarli come spunti di riflessione sulla disciplina del design applicata al tema della privacy online.

6.1 Il team di progetto

ReCon nasce da una collaborazione internazionale di un gruppo di ricercatori provenienti da diverse università: Northeastern University, INRIA, University of Helsinki e University of California. All'interno di questo però le persone che sono state più coinvolte durante il periodo di tesi sono state David Choffnes e Jingjing Ren, entrambi ingegneri informatici e membri del dipartimento di Computer and Information Science della Northeastern University[2]. Il progetto ha quindi già in partenza una forte impronta tecnologica e ingegneristica.

6.2 Applicazioni mobile e il controllo delle autorizzazioni

Un'applicazione per smartphone di default potrebbe solo accedere a processore e RAM[3], che insieme compongono il "cervello" del cellulare, e allo schermo, utile per restituire visivamente l'output all'utente. Per ottenere ulteriori informazioni, come ad esempio le foto salvate in memoria, la geolocalizzazione o l'accesso a Internet, l'app deve richiedere permessi specifici. Le modalità con cui queste autorizzazioni vengono comunicate all'utente cambiano tra i vari Sistemi Operativi, così come la possibilità di controllo dopo l'installazione. Verranno analizzati solo i due sistemi operativi principali, che si spartiscono insieme intorno al 96% del mercato: Android di Google e iOS di Apple.

Per Android il punto di accesso delle app è il Google Play Store. Cercando nuove app, l'utente può consultare per ognuna di esse, prima dell'installazione, l'elenco delle autorizzazioni che verranno richieste al telefono comprese di una descrizione alquanto generica di cosa ciò com-

[2] ■ *http://www.ccis.northeastern.edu*

[3] ● Random Access Memory. È la memoria usata dal processore per, beh, processare calcoli. *È come la memoria a breve termine: aiuta a ricordarsi di compiere tutte quelle cose che ti eri prefissato di fare nell'arco della giornata.*

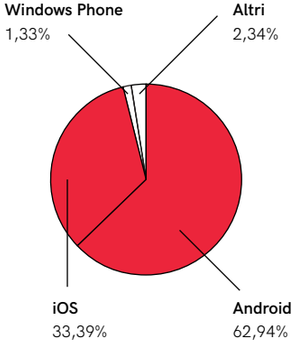
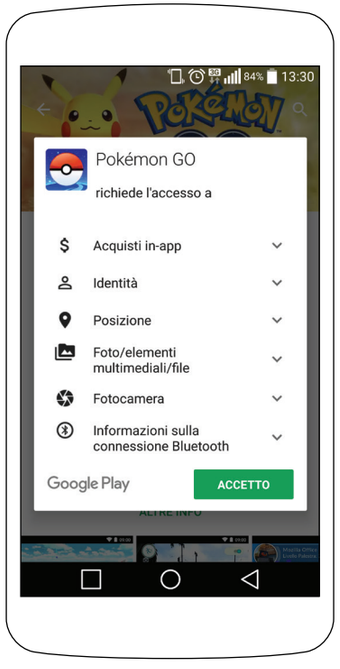
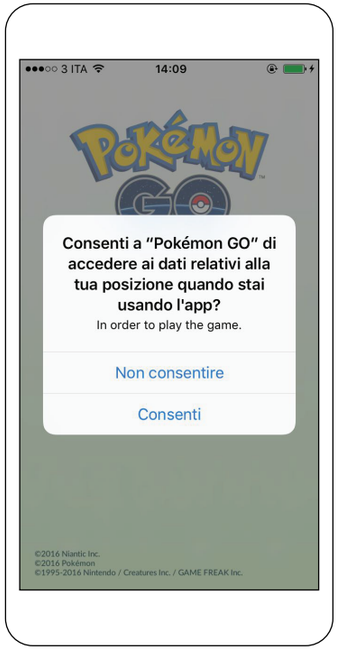


FIG. 20 Sistemi Operativi per smartphone: la divisione delle quote di mercato.
Fonte: https://www.netmarketshare.com





IMG. 25-26 Schermate di richiesta dei permessi di accesso a dati aggiuntivi rispettivamente su Android e iOS.

porti. Un elenco simile, ma non uguale, è riproposto come riassunto e accettarlo è condizione necessaria per poter installare l'applicazione desiderata. Su Android infatti, le autorizzazioni devono essere accettate in toto e l'utente non ha poi modo di modificare o limitare l'accesso a determinate informazioni. Questa limitazione funziona quando un eccessivo controllo da parte di un utente poco esperto potrebbe minare le funzionalità delle app installate ma diventa un ostacolo quando le informazioni richieste sono totalmente superflue ad un corretto funzionamento e l'unica soluzione per evitare un abuso nell'utilizzo dei dati personali è la disinstallazione dell'applicazione.

Apple ha invece un approccio quasi opposto. L'App Store non contiene alcuna descrizione dei permessi che verranno concesse all'applicazione – aldilà di un link alle privacy policies degli sviluppatori – e l'utente può installarla subito senza ulteriori passaggi. Dopodichè però, la prima volta che una delle applicazioni ha bisogno di particolari autorizzazioni per accedere a informazioni personali, un avviso chiede il consenso dell'utente che può decidere se acconsentire o meno. Inoltre in qualsiasi momento egli può rivedere le proprie decisioni in un'apposita sezione del menù "Impostazioni", avendo il pieno controllo dei permessi dati. Anche se è un ottimo passo avanti in termini di privacy rispetto alle politiche adottate da Google per Android, anche su iOS c'è margine di miglioramento: non tutte le autorizzazioni vengono infatti mostrate, ad esempio quelle che riguardano la condivisione dell'ID o del numero di telefono.

In nessuno dei due sistemi operativi c'è però modo di sapere quanto spesso i dati personali vengano richiesti al telefono e quali soggetti stiano raccogliendo informazioni su di noi. Una maggior trasparenza nel comunicare i permessi richiesti dalle applicazioni e un miglior controllo su di essi darebbero all'utente finale degli strumenti più efficaci per

fare scelte consapevoli ed informate.

Varie soluzioni sono state proposte per sopperire a queste forme di opacità, ognuna coi suoi pro e contro. Ai fini di questa tesi verranno presentati, con un'infarinatura superficiale, solo alcuni tra gli approcci più diffusi, entrando poi un po' più in dettaglio sulla soluzione specifica proposta dagli sviluppatori di ReCon. Per identificare l'invio di dati personali in uno smartphone possiamo seguire due strade: analizzare il flusso di informazioni all'interno del cellulare (in inglese *information flow analysis*) o analizzare il flusso nel network (*network flow analysis*)(Ren et al., 2016).

Nel primo caso si monitora ogni qualvolta un'applicazione chiede di accedere a dei dati personali al sistema operativo e tutta l'analisi avviene all'interno del dispositivo. Può avvenire dinamicamente, se è svolta mentre le app sono attive, oppure staticamente, se è una simulazione fatta prima di avviare le app. Entrambi i sistemi sono molto efficienti a identificare scambi di dati personali, ma hanno anche alcuni importanti limiti (ibidem). L'analisi dinamica richiede molte risorse computazionali, rallentando quindi percepibilmente il telefono, e necessita di apportare modifiche al sistema operativo (in gergo tecnico *rooting* o *jailbreaking*), limitandone la diffusione ai soli utenti esperti ed annullando la garanzia sul cellulare; l'analisi statica può risultare imprecisa essendo solo una simulazione e non è in grado di gestire codice caricato dinamicamente, pratica sempre più comune nelle applicazioni mobile (ibidem).

Analizzare il network sul quale avvengono i passaggi di informazione invece delle applicazioni stesse permette di non dover apportare modifiche al sistema operativo e di non dover accedere al software. Il traffico in uscita dallo smartphone viene intercettato utilizzando un VPN[4] che, interposto virtualmente tra il telefono e Internet, permette anche più facilmente di bloccare o modificare i pacchetti di informazioni prima che raggiungano la loro destinazione finale. L'inconveniente è che questo sistema diventa inefficace se le informazioni vengono appositamente mascherate

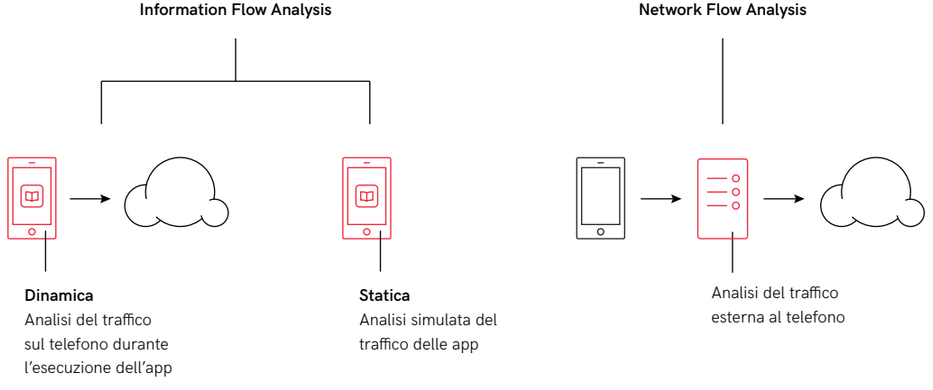


FIG. 21 Schema semplificato della differenza tra information e network flow analysis.

[4] • Virtual Private Network. Si tratta di una rete privata e cifrata ma che usa un protocollo di trasmissione pubblico (come Internet). *È come essere insieme ad un altra persona, gli unici due a parlare un determinato linguaggio. Potreste urlarvi cose segrete da una parte all'altra di una stanza affollata e nessuno capirebbe ciò che vi state dicendo. (fonte: Sideways Dictionary)*

dalle applicazioni prima di passare per il network.

ReCon fa parte di questo gruppo ma a differenza di progetti simili (AntMonitor[5], HayStack[6]) usa un VPN esterno al cellulare per monitorare il traffico sul network ed un algoritmo di machine learning per identificare i dati personali. Su un cellulare con installato ReCon, i dati non verranno inviati su Internet normalmente, ma tutto il flusso passa per un apparecchio esterno (il VPN) il quale decide se i pacchetti mandati contengono informazioni personali o no (senza che nessun essere umano debba controllare, grazie all'uso dell'algoritmo).

Si ottengono così tre grossi vantaggi: basandosi su tecnologia esterna al cellulare funziona su qualsiasi piattaforma; l'utilizzo di un algoritmo permette di non dover sapere a priori i dati personali dell'utente per riuscire a bloccarli; per lo stesso motivo è infine più flessibile rispetto a nuove o diverse modalità di invio di informazioni.

6.3 Idea e processo

Per il Data Transparency Lab la collaborazione aveva l'obiettivo di utilizzare i dati già in possesso dagli sviluppatori di ReCon per creare una serie di visualizzazioni che aiutassero a comunicare il problema ad un pubblico più ampio. La componente di design entrava alla fine del processo per rendere coinvolgente e fruibile il lavoro già svolto e le decisioni già prese. Il primo passo fondamentale era perciò quello di provare di persona ReCon in modo di avere una visione più completa dei dati raccolti, capendo come venissero estratti, cosa significassero realmente e facendosi un'idea approssimativa delle possibili lacune nel dataset (oltre alla curiosità nel poter avere un quadro più chiaro della propria situazione personale).

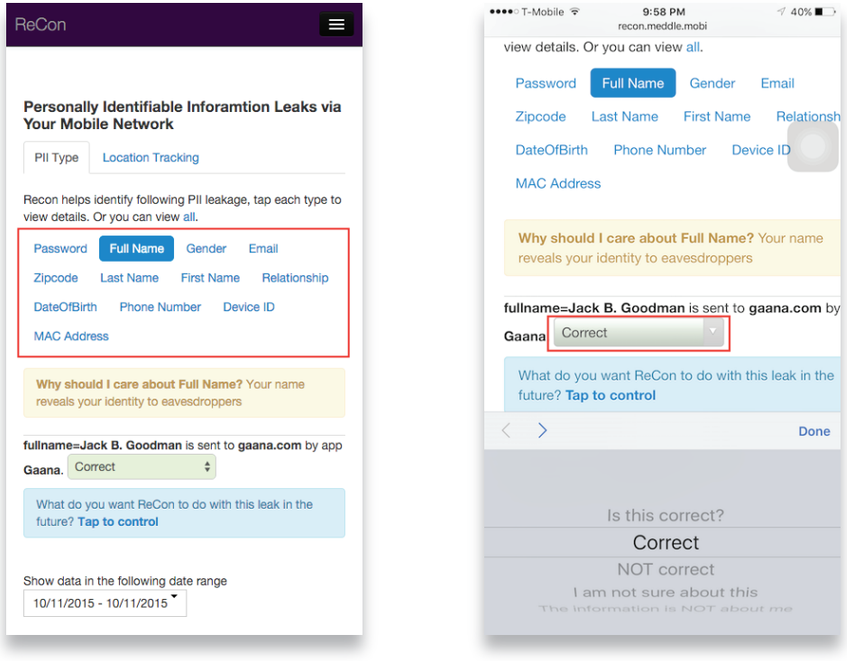
La natura sperimentale del progetto si è però subito concretizzata nella difficoltà di anche solo arrivare a poter dare uno sguardo ai dati. Essendo ancora un prototipo – seppur in fase avanzata – l'applicazione ha richiesto parecchio dispendio di energie e numerosi passaggi per poter essere installata. Non essendo finita, non era disponibile su Google Play Store e doveva essere scaricata seguendo una procedura particolare; un ulteriore processo è stato poi quello di configurare il cellulare perché si collegasse a Internet tramite il VPN dedicato. Anche una volta installata, l'app non ha dato segni di attività per lungo tempo, forse anche per problemi legati al telefono impostato in lingua italiana che poteva confondere l'algoritmo di riconoscimento delle informazioni personali. La scarsità di risorse economiche e umane e la distanza geografica (non tutto può essere risolto velocemente con Internet e il fuso orario è una barriera temporale per cui non è ancora stata trovata una facile so-

[5] *http://odysseas.calit2.uci.edu/doku.php/public:antmonitor*

[6] *https://haystack.mobi*

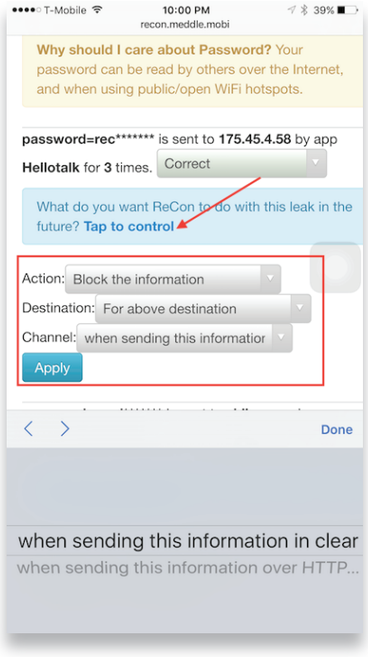
luzione) non hanno certo contribuito a velocizzare la risoluzione delle varie problematiche.

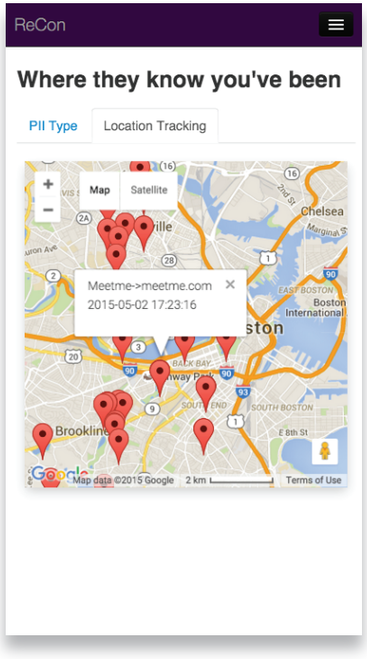
Una volta installata l'app ed utilizzata per qualche tempo, un ulteriore scoglio ha interrotto i lavori. L'obiettivo stabilito durante il primo incontro – progettare alcune visualizzazioni utilizzando i dati già estratti dagli sviluppatori – non sembrava più aver individuato la priorità in termini di efficacia comunicativa per il progetto. Il vedere dati aggregati, generali, difficilmente avrebbe aiutato gli utenti a trovare un punto di contatto con il loro vissuto, mentre la possibilità di dare uno sguardo ai propri dati personali avvicinerebbe il fenomeno alla loro esperienza quotidiana. D'altra parte ReCon si presenta con un'interfaccia confusa e senza gerarchia, che genererebbe probabilmente più domande che risposte nell'utente. Insieme con i vari stakeholder (non solo gli sviluppatori ma anche i rappresentanti del Data Transparency Lab) si è perciò deciso di rivedere il brief al fine di usare ReCon come caso studio per la riprogettazione di un'interfaccia che tenesse più conto dei bisogni dell'utente. Dare una visione più chiara, che mostrasse all'utente come le sue applicazioni usassero le sue informazioni personali, avrebbe sicuramente avuto un effetto più deciso sulla consapevolezza e la comprensione del problema rispetto ad una visualizzazione basata su dati generali, con applicazioni mobile magari sconosciute.



IMG. 27-28 Schermate attuali di ReCon che illustrano il processo di catalogazione e controllo dei dati personali.
Fonte: https://recon.meddle.mobi/app-pii.html

A questo obiettivo, legato all'esperienza specifica con ReCon, se ne aggiunge anche un secondo, funzionale a questa tesi ma proprio per questo centrale a tutto il lavoro svolto: il progetto diventa utile sperimentazione su come rendere visibili, attraverso i dati, processi nascosti e opachi. In quest'ottica esso diventa il punto di partenza per una riflessione più generale sulle caratteristiche che un designer si deve trovare ad affrontare lavorando in un contesto del





IMG. 29-30 Schermate attuali di ReCon che illustrano il processo di catalogazione e controllo dei dati personali. *Fonte: https://recon.meddle.mobi/app-pii.html*

genere, sugli ostacoli e sulle opportunità che si possono presentare.

Attualmente l'interfaccia di ReCon appare macchinosa, costringendo l'utente a premere diverse volte prima di raggiungere qualsiasi informazione desiderata. Manca infatti una gerarchia chiara delle funzionalità offerte a cui viene data a tutte la stessa importanza, risultando in un sovraccarico di informazioni che disorienta molto nei primi utilizzi. I dati sono restituiti sotto forma di elenco in formato tabulare, senza possibilità di avere una visione d'insieme. È stato necessario quindi riprogettare l'architettura dell'informazione di ReCon e assegnare una struttura gerarchica alle varie funzionalità, per far sì che l'utente sia incanalato in un processo che faciliti i primi utilizzi e che lo supporti nella gestione dell'app. Una serie di cicli iterativi che incorporassero i feedback di diversi tipi di stakeholder – i ricercatori della Northeastern University, alcuni componenti del Data Transparency Lab e di DensityDesign come utenti esperti, un campione ristretto di utenti totalmente esterni al progetto – ha portato alla creazione di un primo prototipo di interfaccia.

6.4 ReCon 2.0

Il primo punto critico da risolvere è stato il senso di disorientamento dovuto al sovraccarico di informazioni durante i primi approcci con ReCon. Al primo accesso perciò, l'utente è guidato in una serie di schermate introduttive che spiegano le funzionalità principali e iniziano a familiarizzare l'utente con alcuni termini che verranno riutilizzati nell'applicazione. Questa sorta di tutorial sarà poi sempre accessibile oltre il primo utilizzo, accedendovi dal menù laterale.

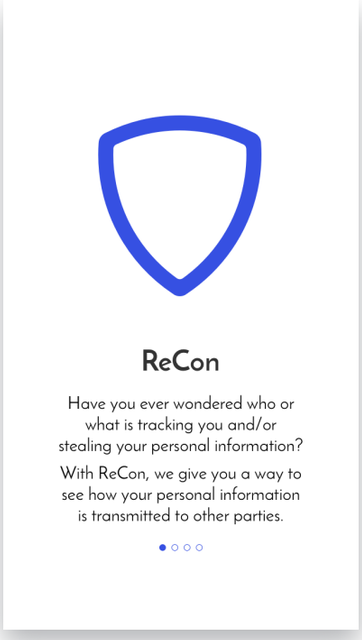


FIG. 22-25 Tutorial iniziale di spiegazione delle funzionalità di ReCon. Vengono introdotti i concetti principali e le possibilità offerte dall'app.

Al posto di un elenco con tutte le funzionalità, la schermata principale presenta solo una panoramica dei dati personali inviati dalle altre applicazioni quel giorno, raggruppati in categorie per somiglianza tematica. In questo modo l'utente ha subito una prima importante informazione che si aspettava di trovare, senza dover premere ulteriori bottoni. Espandendo le categorie può esplorare in maggior dettaglio quali dati sono stati inviati e poi scegliere di indagare quelli che trova più interessanti. Il resto delle funzionalità è stato spostato nel menù laterale, perché tutte secondarie e conseguenti ad una scelta di agire da parte dell'utente.

Entrando nelle specifiche tipologie di informazioni inviate (nome, data di nascita, posizione geografica, etc...) viene proposta innanzitutto una spiegazione sul perché è importante proteggere quell'informazione. Un visualizzazione riporta la quantità di dati personali inviati dalle applicazioni installate in un intervallo temporale definito, con la possibilità anche di vedere diversi momenti o periodi nel passato. La schermata presenta anche le app che ReCon pensa abbiano inviato in quel periodo informazioni di quella tipologia, riportando quante volte lo hanno fatto, se il destinatario è riconosciuto come tracker e il testo effettivo riconosciuto. Quest'ultimo serve per permettere all'utente di verificare che ReCon abbia rilevato correttamente

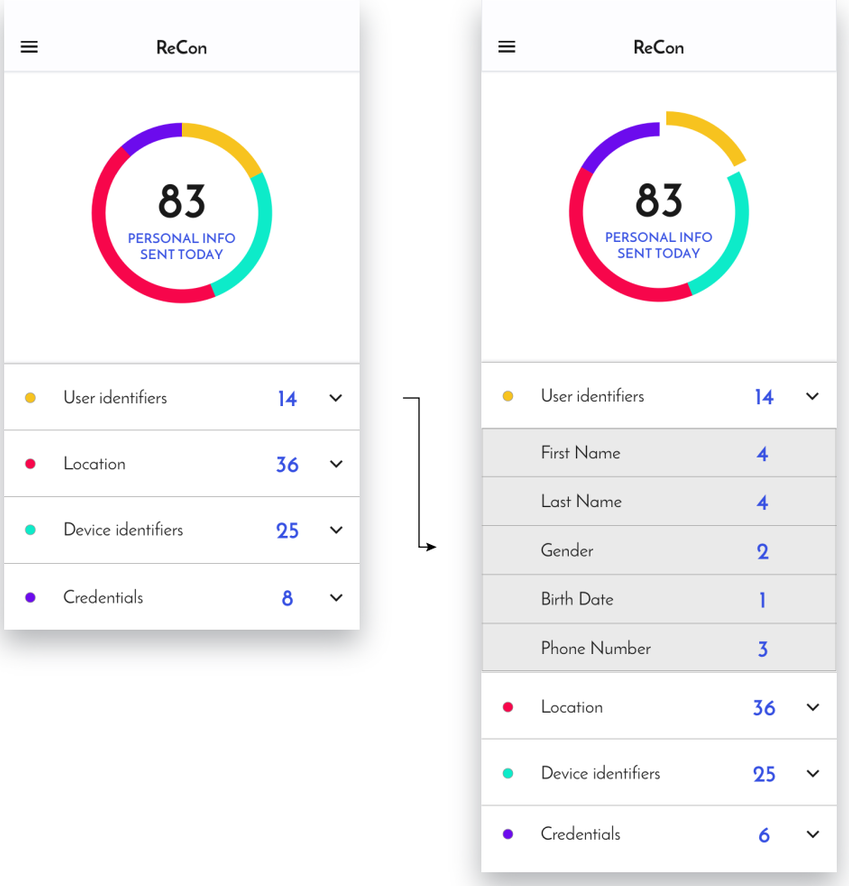
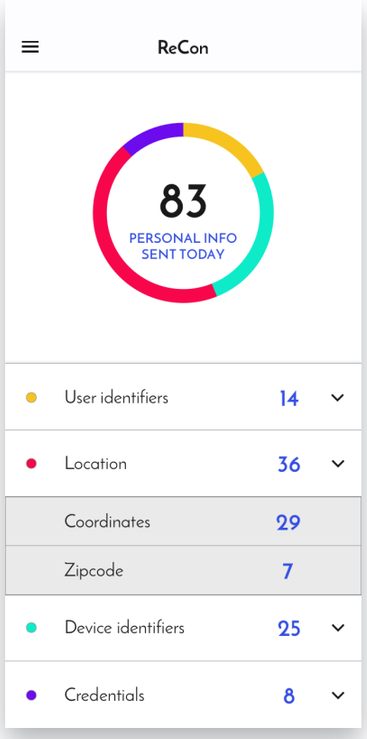


FIG. 26-27 Home page dell'applicazione. Ogni voce è espandibile con una categorizzazione più dettagliata.



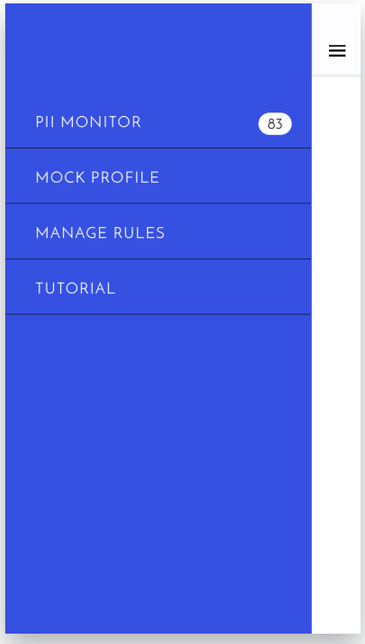
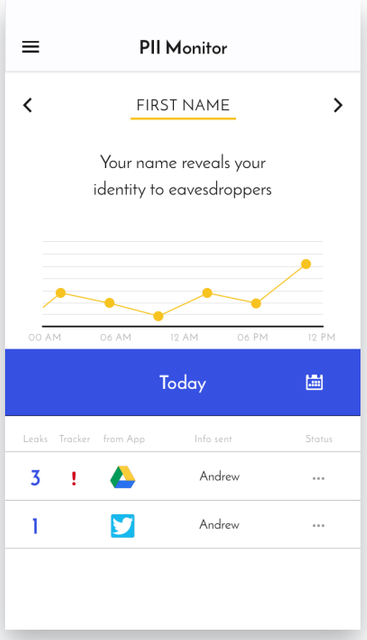


FIG. 28-29 Home page dell'applicazione e Menù laterale contenente le funzioni principali.



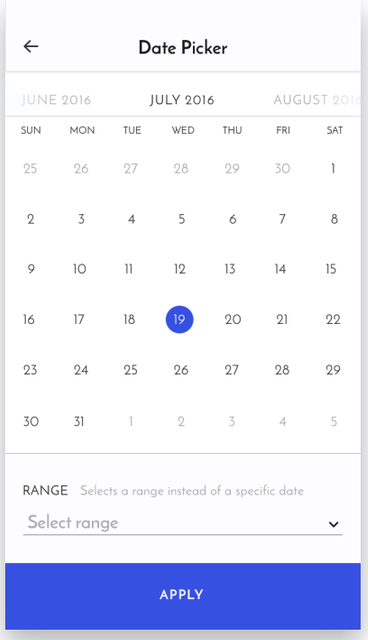


FIG. 30-31 All'interno di ogni categoria vengono elencate tutte le app che hanno inviato quella data informazione personale. L'utente può cambiare il periodo d'ispezione utilizzando il calendario.

l'informazione personale. Si è voluto dare maggior enfasi ai nomi delle applicazioni (o meglio alle icone) perché sono sicuramente il punto di contatto che l'utente conosce di più, avendo scelto lui stesso di installare quelle particolari app.

Visto che ReCon si basa su un algoritmo di machine learning, fornire nuove indicazioni sulla corretta individuazione delle informazioni inviate rendono l'algoritmo più preciso ed efficace. Per questo l'utente deve poter svolgere questa azione velocemente e senza troppi ostacoli. Si è pensato perciò di utilizzare un'interazione di *swiping* (trascinamento a destra o sinistra) per etichettare la correttezza di ReCon. Questa *gesture* è probabilmente già nota all'utente e usata in contesti simili, ad esempio l'archiviazione o eliminazione delle email sui client di posta elettronica.

Premendo poi su una singola voce, l'utente può decidere il comportamento di ReCon: ogni scelta è introdotta da una breve descrizione che però rende più chiaro il significato di voci che senza nulla richiederebbero delle competenze tecniche avanzate per essere comprese. L'utente può ad esempio scegliere se bloccare o meno l'invio di informazioni per una data applicazione, oppure inviarle ma mascherate, sostituite da parole inventate. ReCon offre infatti anche una sezione apposta per creare il proprio "profilo fasullo", una pratica di offuscamento già incontrata nella parte teorica (→ CAP. 03).

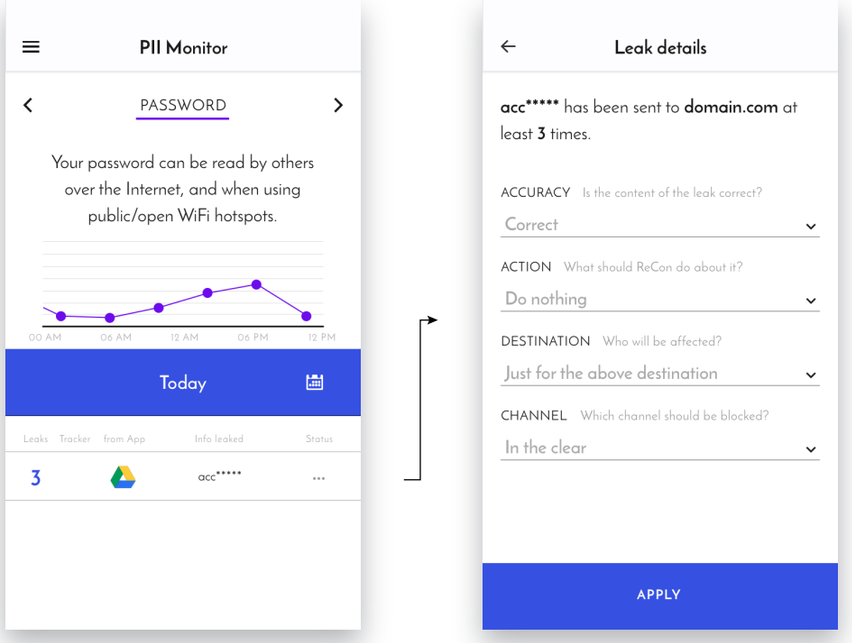


FIG. 32-33 Al fine di addestrare l'algoritmo a riconoscere correttamente le informazioni personali, l'utente può espandere ogni voce dell'elenco per segnare se è corretta o meno, nonché istruire ReCon a lasciarla andare, bloccarla o sostituirla con informazioni falsificate.

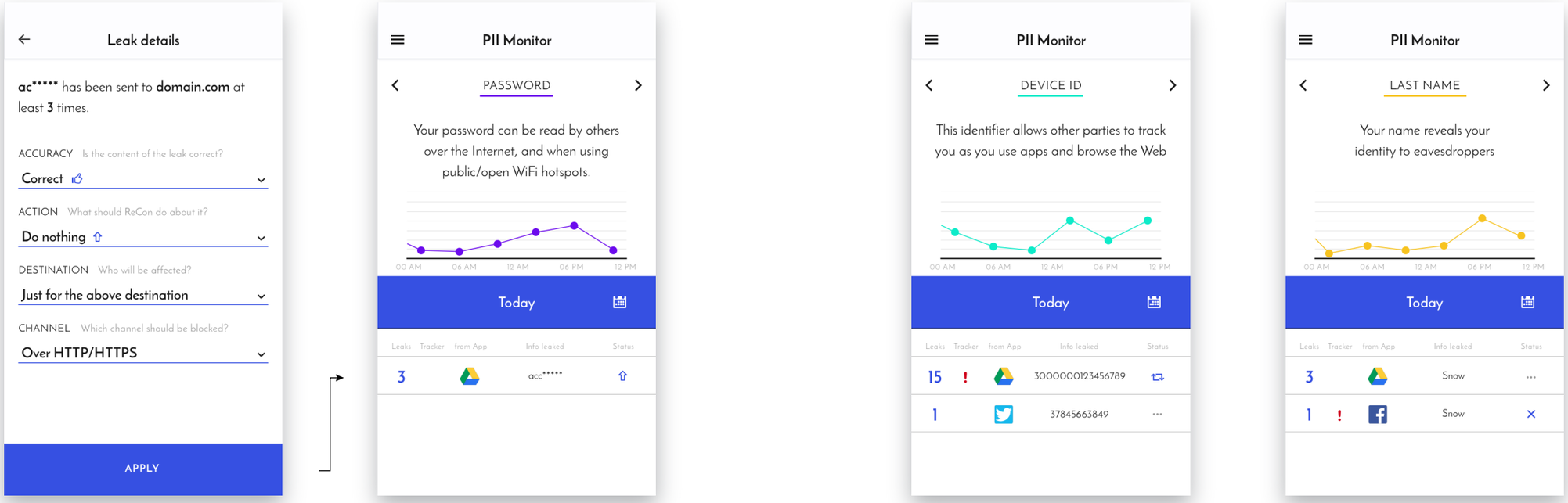


FIG. 34-37 La scelta dell'utente di lasciar passare, bloccare o offuscare i propri dati personali viene poi riflessa nell'interfaccia tramite tre diverse icone.

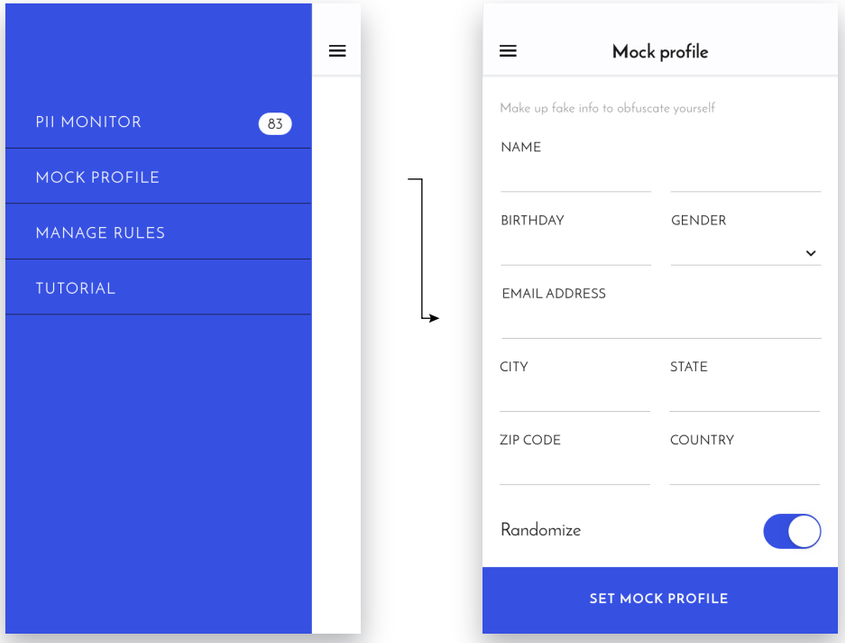


FIG. 38-39 La funzione che permette all'utente di sostituire i propri dati con alcuni fasulli.

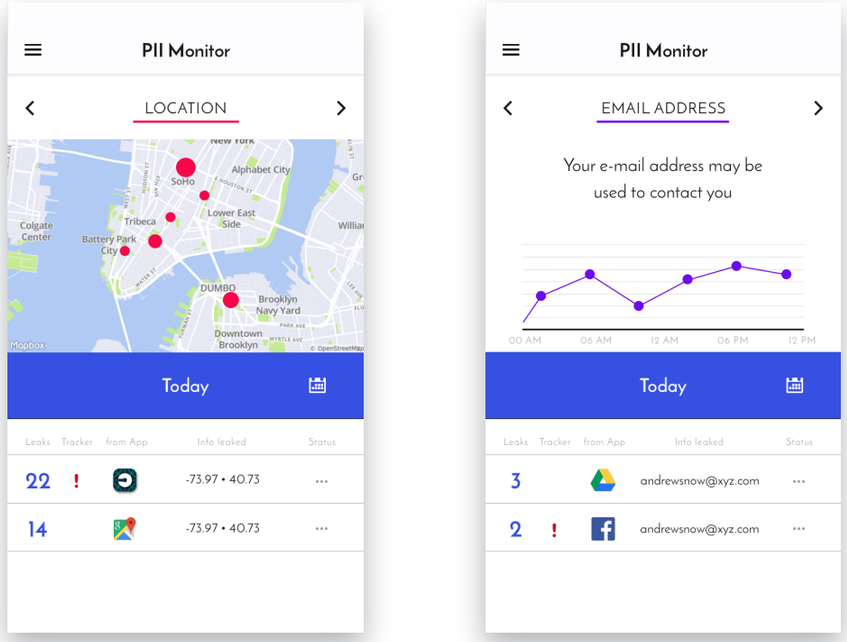


FIG. 40-41 Schermate di altre categorie di ReCon.

6.5 Limiti, opportunità, riflessioni

Una volta completato e rilasciato, ReCon potrebbe diventare un utile strumento a rafforzare la consapevolezza delle persone riguardo la gestione della privacy, aiutando gli utenti a fare scelte informate sulle applicazioni che installano sui propri smartphone. Durante il processo progettuale però sono emersi numerosi ostacoli che si interpongono tra l'app e la sua diffusione, vincoli tecnologici, contestuali ed etici che meritano una riflessione a parte.

In primo luogo, la complessità tecnologica che permea l'analisi di flussi d'informazione in entrata e uscita da uno smartphone, certamente anche per motivi di sicurezza, rende la realizzazione di soluzioni efficaci un processo tedioso e pieno di frustrazioni. Bisogna progettare lo strumento in modo che funzioni su tutte le piattaforme e su tutte le versioni dei vari sistemi operativi. Durante tutto il percorso della tesi, moltissimo tempo è andato perso per via di malfunzionamenti su una particolare versione di Android – o meglio, il server registrava i dati ma questi non comparivano sull'interfaccia – oppure perché il traffico di alcune app non veniva rilevato. Inoltre alcune applicazioni, ad esempio Uber, bloccano il traffico proveniente da particolari server, rendendo impossibile la raccolta di una parte delle informazioni. L'esperimento con ReCon ha quindi sottolineato come, prima ancora di verificare se i metodi conosciuti del designer per la trasformazione e visualizzazione dei dati siano applicabili al contesto della privacy online, bisogna innanzitutto trovare metodi adatti alla loro collezione e valutare se quanto raccolto sia sufficiente per rappresentare il fenomeno. Data la complessità tecnologica, le metodologie devono essere necessariamente sviluppate in stretta collaborazione con degli sviluppatori esperti: non basta più quella figura ibrida di designer-programmatore che può invece funzionare in altri contesti. È fondamentale quindi creare con gli sviluppatori un vocabolario comune che permetta una comprensione reciproca e un dialogo produttivo - per il quale una figura ibrida è sicuramente più preparata.

Un secondo punto di riflessione, uscito anche durante le sessioni di feedback con gli utenti, riguarda il target del progetto. Per persone senza o con limitate conoscenze del contesto tecnologico oppure ancora poco sensibilizzato al tema della privacy digitale, ReCon diventa uno strumento non molto immediato da usare, che ha bisogno di tempo perché si possa familiarizzare con i concetti chiave e le funzionalità. Questa è forse la barriera più grande contro una discussione pubblica più ampia e parte degli sforzi futuri dovrebbero concentrarsi a educare ad un uso più consapevole della tecnologia. Per utenti più esperti o per quelli ricettivi al tema privacy, l'app può essere invece un utile sup-

porto perché mostra informazioni su misura per ciascun individuo, avendo quindi l'opportunità di colpirlo più nel profondo.

Infine, è necessario spendere qualche parola sulle implicazioni etiche di strumenti come ReCon, soprattutto rispetto all'inviare informazioni manipolate o al bloccare totalmente l'invio dei propri dati. Brunton e Nissenbaum elencano una serie di accuse avanzabili contro le pratiche di *data obfuscation*. Esse potrebbero essere utilizzate con intenti malevoli per compiere atti illegali, minano la correttezza di database che potrebbero essere utilizzati per offrire benefici all'intera società, oppure danno la possibilità di usufruire dei servizi offerti dalle aziende senza pagarne il prezzo (Brunton & Nissenbaum, 2013).

L'eticità non è facilmente valutabile in astratto ma richiede di essere analizzata insieme ai dettagli che la contestualizzano. Distorcere i dati per organizzare frodi è un'azione con un fine chiaramente illegittimo e perciò ingiustificabile, ma spesso la legittimità del fine è una questione più complessa, non riassumibile in un "giusto/sbagliato". Se un individuo sente di non avere le giuste protezioni dalla legge, dalla tecnologia o dalle buone pratiche delle aziende, senza possibilità di poter conoscere l'affidabilità degli attori che accederanno alle sue informazioni personali né le modalità in cui queste verranno utilizzate, la questione sulla legittimità diventa molto più ardua da discernere (ibidem). Riflessioni simili nascono anche sugli altri punti sollevati: quanto è giusto obbligare a fornire informazioni veritiere per il bene di un'altra parte, senza assicurazioni o trasparenza di come esse verranno utilizzate, trasferite o protette? Quanto sono legittime eticamente le pretese delle aziende che raccolgono dati se manca una reale stima dei costi e dei benefici per la società derivanti dall'accesso e utilizzo dei dati delle persone? (ibidem)

Come per molti altri problemi legati all'etica, non ci sono forse ancora delle risposte a queste domande senza un ulteriore approfondimento del contesto in cui sono state poste. È importante però che queste domande vengano riconosciute ed entrino a far parte di un dibattito pubblico più ampio.

[Algorithmic personalization] moves us very quickly toward a world in which the Internet is showing us what it thinks we want to see, but not necessarily what we need to see.

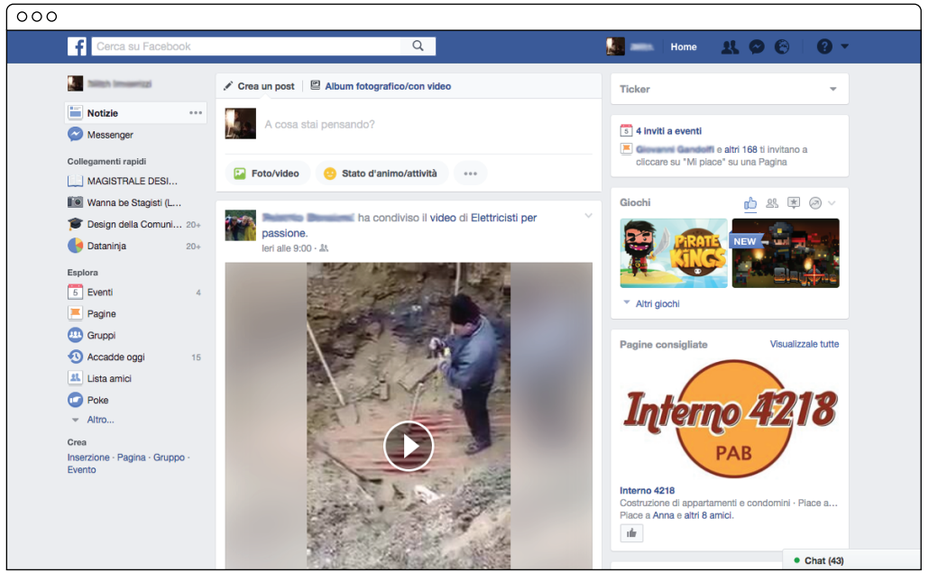
Eli Pariser

7 — Esperimento n° 2: facebook.tracking.exposed

Facebook.tracking.exposed è un progetto iniziato nel Giugno 2016 da Claudio Agosti, ingegnere del software e attivista che collabora con associazioni che si occupano di difesa dei diritti digitali come Tactical Tech Collective e CodingRights. Ora è portato avanti da un gruppo di volontari interdisciplinare che unisce competenze di ingegneria informatica, programmazione, diritto e management. Usando le parole degli sviluppatori, facebook.tracking.exposed ha come obiettivo quello di "[...] aumentare la trasparenza dietro gli algoritmi di personalizzazione, in modo che le persone possano avere un controllo più efficace della loro esperienza su Facebook oltre che una maggior consapevolezza delle informazioni a cui sono esposti"[1]. Il punto di partenza è perciò quello di voler mostrare ai propri utenti come l'algoritmo che gestisce il *News Feed*[2] di Facebook non sia un agente neutro, che restituisce imparzialmente la realtà intorno a noi, ma bensì come esso influisca in modo opaco sulla nostra percezione. Non avendo modo di poter vedere come funziona l'algoritmo al suo interno (come visto nel → CAP. 2.2), una soluzione è di sottolinearne

[1] ■ *https://facebook.tracking.exposed*

[2] ● È l'home page di Facebook, in cui appaiono i post degli amici con cui si è in contatto e le pagine che si seguono.



IMG. 31 Schermata del News Feed di Facebook.

gli output facendo emergere la sua presenza. Ciò è fatto tramite uno strumento – in fase alpha[3] un semplice script diventato poi durante la beta un'estensione per browser – che permette di raccogliere una serie di metadati mentre l'utente utilizza il social network.

La collaborazione, portata avanti durante la fase alpha del progetto, è nata con l'idea di realizzare una visualizzazione interattiva che aiutasse l'utente a comprendere i dati estratti dallo strumento. Contemporaneamente era un'altra opportunità per testare le metodologie proprie del designer in una situazione dove il dato non è facilmente accessibile ma al contrario strettamente difeso.

Come nel capitolo precedente, dopo una breve descrizione del gruppo di persone che hanno partecipato al progetto e un'introduzione ad alcuni elementi necessari a comprendere meglio il contesto, verranno illustrati il processo progettuale e i risultati ottenuti, sempre nell'ottica di usarli come spunti di riflessione sulla disciplina del design applicata al tema della privacy online.

7.1 Il team di progetto

Nella fase alpha i componenti di facebook.tracking.exposed erano quasi tutti programmatori: Claudio Agosti, Cristina Carnevali, Gilberto Conti e Alberto Granzotto da diverse parti del mondo investivano parte del loro tempo libero per dedicarsi al progetto. Gli unici provenienti da una disciplina diversa erano Luca Corsato e Andrea Raimondi, fondatori di Open Sensor Data[4], uno studio di consulenza specializzato in iniziative che fanno uso di open data[5]. Il gruppo riconosce i propri limiti nella mancanza di interdisciplinarità[6]:

> The team needs to grow diverse and multicultural; The whole goal of facebook.tracking.exposed can be reached only through a global and inclusive community able to critique the algorithms from multiple points of views. The current team suffer this limit because the project, so far, has been presented mostly in italian and german hacker events. This has been driven by opportunities and is not a self-segregating decision.

3 ● Nel ciclo di vita di un software le fasi di test e sviluppo sono comunemente chiamate alpha e beta e servono per testare velocemente il prodotto al fine di risolvere bug e migliorare le funzionalità.

4 ■ *https://osd.tools*

5 ● Dati liberamente accessibili a tutti con l'unico obbligo, nel caso di utilizzo, di citare la fonte e mantenere i dati ugualmente aperti.

6 ■ *https://facebook.tracking.exposed/about*

7.2 Facebook e il suo algoritmo

Già si è parlato di quanto spesso, a volte senza rendercene conto, interagiamo con degli algoritmi che scelgono, decidono e filtrano per noi, avendo sempre maggior influenza nella nostra vita (→ CAP. 2.2). Essi sono ormai una delle logiche fondamentali che regolano i flussi di informazione da cui dipendiamo (Langlois via Gillespie, 2014). Facebook, con i suoi 1.23 miliardi di utenti attivi giornalmente (di cui l'85.2% al di fuori di Canada e Stati Uniti)[7], è il social network più diffuso e presente nella vita di tutti i giorni.

Anche Facebook ha un algoritmo. Si potrebbe anzi arrivare a dire che interagiamo più con lui che con i nostri amici. Infatti la composizione della striscia di post che "scrolliamo" quando usiamo il social network, denominata *News Feed*, è generata dando più importanza a ciò che l'algoritmo pensi sia più rilevante per noi e filtrando via tutto ciò che non ritiene lo sia abbastanza. Il rischio è che questa selezione e personalizzazione porti alla creazione di una "bolla" isolata, una campana di vetro dove ciò che una persona vede è solo una parte della realtà e lei non ha voce in capitolo per decidere quale questa parte sia (Pariser, 2011). Il resto è tagliato fuori, impercettibile, perché non si ha modo di vedere la bolla degli altri né tanto meno rendersi conto di essere dentro ad una. Ci si trova di nuovo di fronte ad un processo nascosto, opaco.

[7] *http://newsroom.fb.com/company-info*

Se il rischio di venire isolati da alcuni amici piuttosto che altri può sembrare poco rilevante, la questione diventa più problematica nel momento in cui, ad esempio, Facebook diventa la fonte di informazione principale di una crescente fetta della popolazione[8-9], con 4.75 miliardi di pezzi di contenuto condivisi giornalmente[10]. In questo contesto un algoritmo di personalizzazione tenderà a fornire all'utente solo notizie provenienti da fonti che egli segue, su cui ha cliccato o che piacciono ai suoi amici (sono sempre supposizioni, non si può avere la certezza di cosa l'algoritmo prenda come input e che peso assegni ad ogni variabile), col pericolo che si crei una sorta di camera d'eco che ripeta all'infinito un'unica prospettiva, senza possibilità di incontrare mai opinioni differenti e narrative trasversali. Questo pericolo aumenta esponenzialmente se le fonti che compongono la "dieta informativa" di una persona sono link a siti pieni di mala-informazione, faziosità, bufale, senza che la persona in questione sia abituata a selezionarle con uno sguardo critico. Schmidt et al. hanno mostrato come gli utenti di Facebook (analizzando 376 milioni di utenti e 920 agenzie di news) tendono a limitare la loro attenzione ad un numero molto limitato di pagine, creando una forte polarizzazione in termini di consumo di notizie (Schmidt et al., 2017). Dalle elezioni di Donald Trump nel Novembre 2016, il dibattito sulle "fake news" (notizie create con

[8] *http://www.journalism.org/2015/06/01/facebook-top-source-for-political-news-among-millennials*

[9] *http://www.franzrusso.it/condividere-comunicare/facebook-e-una-delle-principali-fonti-di-informazione-in-italia*

[10] *https://zephoria.com/top-15-valuable-facebook-statistics*

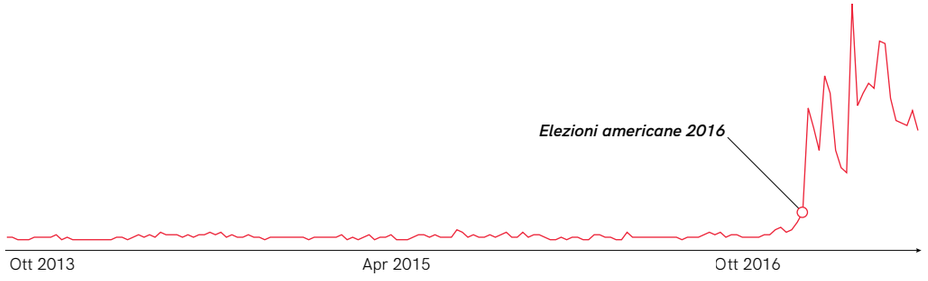


FIG. 42 Google Trend relativo alla parola chiave "Fake News". Dopo le elezioni americane, il dibattito ha, tra le altre cose, sollevato dubbi sul ruolo di Facebook e la sua responsabilità nello sviluppo del fenomeno.

l'intento malevolo di fornire, appunto, falsa informazione) è diventato uno dei temi più discussi online e nel mondo giornalistico.

Oltre agli sforzi volti a cercare di limitare queste fake news e a sensibilizzare il pensiero critico degli utenti, è fondamentale progettare soluzioni che facciano emergere e rendano visibile la presenza dell'algoritmo di Facebook, a ricordare che online, come nella vita reale, nulla è neutrale. Facebook.tracking.exposed è un primo tentativo di fare proprio questo: una volta installato, lo strumento inizia a raccogliere metadati mentre l'utente naviga l'interfaccia di Facebook, raccogliendo informazioni sulla posizione che un post occupa nella Timeline, sul momento in cui è stato creato, sulla sua tipologia (se è un post sponsorizzato, se proviene dalla cerchia di amici o da amici di amici). In questo modo l'utente può in un secondo momento rivedere, in una sorta di panoramica generale, l'evoluzione nel tempo delle sue permanenze su Facebook e in certo senso avere anche la possibilità di monitorare la presenza dell'algoritmo nella propria Timeline. Anche se al momento della scrittura di questa tesi facebook.tracking.exposed ha aggiunto altri metadati raccolti e ha cambiato approccio rispetto alle idee iniziali, prenderemo in analisi il suo stato durante la collaborazione nella fase Alpha del progetto.

7.3 Idea e processo

L'obiettivo generale della collaborazione per Claudio e il suo gruppo consisteva nel creare una visualizzazione che aiutasse a comunicare meglio il progetto ad un pubblico meno competente di tecnologia. Come per l'esperimento precedente, il design era visto in funzione di "impacchettamento finale", anche se la fase embrionale del progetto lasciava qui maggior margine ad un intervento sostanziale. Nella fase iniziale di facebook.tracking.exposed, i metadati a disposizione per una visualizzazione non erano molti, per via del modo in cui lo strumento li estraeva da Facebook (nello specifico, usava uno script lato client per prendere le informazioni dei post man mano venivano caricati). Perciò si è deciso di focalizzarsi sul mostrare tre tipi di relazioni: quanto l'ordine dei post in una data Timeline si allontana dall'ordine cronologico; come cambiasse la posizione di un

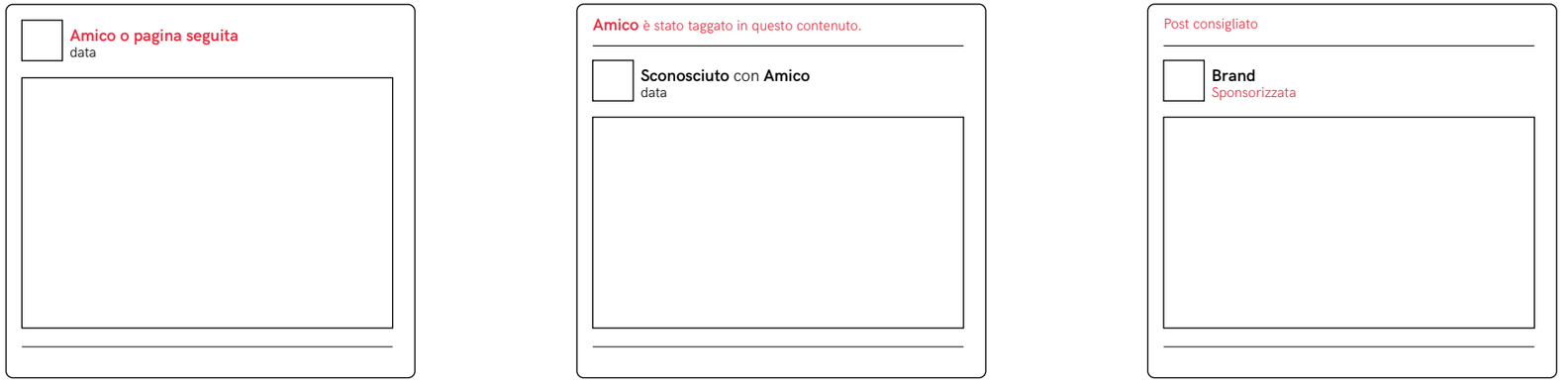


FIG. 43 Tipologie di post che facebook.tracking.exposed distingue. Nell'ordine dall'alto in basso: *Feed, Friends Feed, Promoted*.

dato post in diverse Timeline consecutive; quale fosse la composizione in termini di tipologia di post in una Timeline, differenziandoli a seconda che fossero attività di amici con cui siamo in contatto o pagine che seguiamo, attività che coinvolgono i nostri amici ma che normalmente non comparirebbero nel nostro Feed e infine post inseriti da Facebook per scopi pubblicitari (i post "sponsorizzati").

Altre relazioni interessanti come il numero di amici presenti nella maggior parte dei post rispetto al numero di amici totali oppure il rapporto tra la posizione nella Timeline e il numero di like e di condivisioni non erano strade percorribili per via dell'instabilità dello strumento. Facebook cambia infatti molto frequentemente il suo codice sorgente, a volte in sezioni più periferiche ma a volte negli elementi chiave, rendendo eventuali sforzi poco efficaci e uno spreco delle poche risorse umane disponibili. Bisognava quindi sfruttare al meglio le informazioni raccolte stabilmente che fornivano comunque un ottimo punto di partenza per rivelare l'algoritmo.

L'idea è stata quindi quella di creare una visualizzazione interattiva che mettesse in relazione le ultime tot volte che un utente sia stato su Facebook, comparando una vicino all'altra le diverse Timeline, mostrando i dati dei singoli post, sottolineando le volte che quest'ultimi si ripetono più volte e dando la possibilità di riordinarli per il loro reale ordine cronologico o di raggrupparli per tipologia. Volendo dare all'utente la possibilità di comparare diversi momenti in cui aveva utilizzato il social network, si è deciso nella visualizzazione finale di limitare ad un massimo di 6 il numero di Timeline mostrate e a 25 il numero di post per ognuna di esse. Ciò permette all'artefatto finale di rimanere entro i limiti di uno schermo desktop o tablet ed essere leggibile senza ridurre eccessivamente la rappresentazione del fenomeno.

Anche se inizialmente la visualizzazione avrebbe dovuto mostrare sul sito di riferimento del progetto i dati di un utente rappresentativo (dell'autore o di Claudio Agosti), si è poi ritenuto più significativo dare la possibilità a chiunque avesse installato lo strumento di poter vedere i propri metadati, affinché l'artefatto comunicativo svolgesse la sua funzione più efficacemente. È stato perciò necessario fare in modo che esso venisse creato dinamicamente prendendo i dati dal server del progetto rispetto alle informazioni dell'utente (semplicemente il suo ID su Facebook).

Durante il processo progettuale, la visualizzazione ha trovato anche un secondo ma altrettanto importante impiego: dare feedback visuali agli sviluppatori per aiutarli a trovare bug[11] nascosti nel codice, aiutandoli a renderlo più stabile. Da puro output comunicativo finale a valido supporto durante tutto il processo.

7.4 MetaTimeline: a timeline of timelines

Una volta installato lo strumento sul proprio browser e averlo utilizzato per qualche tempo, l'utente può rivedere i propri dati andando su *facebook.tracking.exposed/realitycheck* (non più funzionante, ora *https://fbtrexvisual.github.io*).

Al caricamento della pagina, egli viene accolto da una breve spiegazione che introduce il progetto e gli aspetti principali della visualizzazione, insieme ad un campo dove può inserire il proprio Facebook Id o usarne uno di esempio. Questo serve per comporre la richiesta al server e caricare dinamicamente i dati sempre aggiornati.

La visualizzazione riporta attraverso una semplificazione visuale la composizione delle ultime visite di quell'utente su Facebook: ogni colonna è un momento temporale ed ogni casella è un post, nell'ordine in cui comparivano in quella specifica Timeline. Ad ogni post è assegnato un colore per identificarne la tipologia, che è poi spiegata più in dettaglio nella legenda. Sono mostrati anche i post che facebook.tracking.exposed non è riuscito a catalogare, per mantenere l'ordinamento originale. Questi ultimi hanno anche la molteplice funzione di feedback visuale per gli sviluppatori, che possono poi risalire più velocemente alla causa del problema.

Attraverso l'interfaccia è poi possibile isolare un post per vedere quanto spesso e con che posizionamento si è ripetuto in diversi momenti temporali, oltre a poter poi riordinare i post per tipologia o per ordine cronologico reale. In questo modo si può osservare le Timeline da diversi punti di vista, sottolineando ad esempio la quantità di pubblicità inserita tra i post da Facebook oppure analizzando possibili pattern nell'ordinamento dei post in diverse Timeline.

[11] • Errore nella scrittura del codice che ne previene il corretto funzionamento. *È come la Torre di Pisa, che è sembrata dritta per i primi 5 anni ma ha poi iniziato a inclinarsi, rivelando un bug nel progetto originale. Le fondazioni erano profonde solo 3 metri in un terreno notoriamente friabile e instabile.*
(fonte: Sideways Dictionary)

META-TIMELINE

A timeline of timelines

Facebook is so present in most people's lives that we tend to forget some of the logics that run it. How does it decides what we get in our newsfeed?
We have no idea, but this is an attempt to shed a small light on it.

If you have the facebook.tracking.exposed script installed, you can choose to see some of the past times (aka *refreshes*) you have been scrolling Facebook.
In order to get any insight, insert your Facebook ID. Each column represents a specific time you have been on Facebook scrolling, up to a maximum of 25 posts per session. Hovering or clicking on a post (*a rectangle*) will highlight whenever that particular post appeared in other occasions.

Insert Facebok ID or leave blank to load example

Load timelines

FIG. 44 Schermata iniziale della visualizzazione, ha il compito di introdurre i concetti chiave che appariranno in seguito.

FIG. 45 Visualizzazione delle Timeline dell'utente. Ogni colonna è un momento temporale in cui si è stati su Facebook.

FIG. 46 Muovendosi sopra la legenda, maggiori informazioni vengono fornite sul significato delle diverse categorie.

FIG. 47 Le Timeline possono essere riordinate secondo diversi criteri, ad esempio per tipologia di post.

FIG. 48 Le Timeline possono essere riordinate secondo diversi criteri, ad esempio per ordine cronologico reale.

FIG. 49 Muovendosi sopra un qualunque post, esso verrà evidenziato attraverso tutte le Timeline.

FIG. 50 La caratteristica principale dei post sponsorizzati è che non hanno una data di creazione.

FIG. 51 Alcuni post cambiano tipologia tra una Timeline e un'altra.

FIG. 52 Alcuni post non compaiono in tutte le Timeline.

FIG. 53 La visualizzazione è stato anche un'utile feedback visivo per gli sviluppatori per trovare velocemente degli errori nel codice.

7.5 Limiti, opportunità, riflessioni

L'esperimento progettuale è stato un primo tentativo di far emergere la presenza di uno degli algoritmi che è ormai parte integrante della vita di tutti i giorni. Anche se l'artefatto ha avuto riscontri positivi nella – seppure piccola – base iniziale di utenti ed è servito a diffondere la fase Alpha del progetto durante varie conferenze e presentazioni, esso è servito ad evidenziare alcuni punti chiave di riflessione sull'accesso ai dati necessari a sviluppare questo tipo di progetti, sia da un punto di vista tecnologico che legale ed etico. È stato anche uno spunto per considerare nuovi ruoli emergenti per la visualizzazione dati e per l'information design, che verranno esposti più approfonditamente nell'ultima parte di questo capitolo.

Si è già parlato delle problematiche dovute alla complessità tecnologica che è parte intrinseca del contesto e che spingono anche qui ad una stretta collaborazione interdisciplinare (→ CAP. 6.5), però facebook.tracking.exposed ha esposto una serie di difficoltà – se non impossibilità – che ostacolano una progettazione "ottimale" dell'informazione con cui un designer si deve scontrare lavorando in situazioni simili. Specificatamente a Facebook, ad esempio, la raccolta dati dovrebbe tener conto di tutti i device usati dall'utente nell'interazione con la piattaforma, soprattutto quando gli utenti attivi mensilmente sono 1.86 miliardi in totale ma 1.74 miliardi usano anche o solamente l'applicazione su smartphone[12]. C'è perciò una dimensione di incompletezza del dato dovuta alla raccolta tramite browser che limita la possibilità di restituire all'utente una maggior trasparenza sul come le sue informazioni personali influiscano sul risultato finale dell'algoritmo, potendo solo avere accesso ai dati prodotti via sito internet. Una seconda dimensione di incompletezza è legata alla volatilità della struttura del social network: il codice con cui funziona è costantemente modificato e aggiornato, richiedendo continui cambiamenti alle modalità di estrazione delle informazioni. Il processo può arrivare ad essere talmente costoso in termini di risorse da essere costretti a scartare dati potenzialmente molto utili (→ CAP. 7.2). Certamente le modifiche al codice sono in parte progettate per scoraggiare i tentativi di automatizzare la collezione di informazioni sulla piattaforma, fonte di profitto principale del gigante informatico. La raccolta di dati in maniera automatizzata viola infatti esplicitamente i termini e condizioni d'uso (in inglese *Terms of Service* o ToS) della piattaforma.

È chiaro quindi come in realtà progetti come facebook.tracking.exposed vivano in una sorta di "illegalità". Anche se nessuna legge sta venendo infranta, in qualsiasi momento Facebook avrebbe tutto il diritto di bloccare il progetto, in quanto infrange non uno ma ben tre punti del ToS[13]:

[12] *http://newsroom.fb.com/company-info*

[13] *https://www.facebook.com/terms*

- non raccogliere contenuti o informazioni degli utenti, né accedere in altro modo a Facebook usando strumenti automatizzati (come bot di raccolta, robot, spider o scraper) senza previa autorizzazione da parte nostra.
- non intraprendere azioni che possano impedire, sovraccaricare o compromettere il corretto funzionamento o l'aspetto di Facebook, ad esempio un attacco di negazione del servizio o altre azioni di disturbo che interferiscano con il rendering delle pagine o con altre funzioni di Facebook.
- non favorire o incoraggiare l'inottemperanza della presente Dichiarazione o delle nostre normative.

Finché il progetto rimane piccolo difficilmente Facebook farà qualcosa per intervenire, ma con abbastanza pubblicità il rischio di un'interruzione improvvisa cresce esponenzialmente. È ciò che è successo a B.S. Detector[14], estensione Open-Source per browser creata per identificare ed etichettare articoli di Fake News su Facebook, che ha smesso di funzionare appena aveva guadagnato una discreta popolarità subito dopo le elezioni americane. Visto che le norme di accesso e utilizzo di tutti i servizi online sono ad oggi frutto di un'autoregolamentazione delle aziende, qualsiasi lavoro di enti, associazioni ed attivisti che metta in discussione lo status quo è condannato a scontrarsi con questa realtà. Nonostante questo, ci sono delle violazioni – come l'accesso a Facebook con strumenti automatizzati – che possono considerarsi eticamente giustificabili, specialmente se fatte nell'interesse della società: rendere evidenti e visibili gli algoritmi che prendono decisioni per noi in base ai nostri dati personali è un modo per aiutare a tutelare il diritto alla privacy degli utenti, una sorta di cartina tornasole verso intrusioni eccessive.

[14] *http://bsdetector.tech*

Queste considerazioni portano alla luce un nuovo ruolo della visualizzazione di dati. Fino ad ora si è dimostrata un supporto formidabile quando si trattava di dover aiutare un agente umano a prendere una decisione o a suggerire delle relazioni nascoste, trasformando visualmente una quantità di dati ingestibile e apparentemente incomprensibile per la mente umana. La visualizzazione è dunque lo strumento che traduce la complessità del dato in conoscenza usabile dall'uomo. In un mondo dove gli algoritmi prendono sempre più decisioni al posto nostro, questo tramite indispensabile di traduzione sembra venire a mancare: la macchina non ha bisogno di semplificazioni per poter comprendere vaste quantità di dati, seppur multidimensionali ed eterogenei. Si sta forse avvicinando la fine dell'information design? Certamente no. Con la diffusione degli algoritmi ci sarà sempre più bisogno di poter valutare e monitorare il loro operato, traslando quindi semplicemente la necessità di un momento di traduzione più avanti nel processo. Da strumento di decisione la visualizzazione di dati diventa strumento di supervisione.

CONCLUSIONI

8 — Riflessioni finali

Questa tesi nasce con l'obiettivo di testare l'efficacia delle metodologie progettuali note all'information designer nell'ambito della privacy digitale, partecipando a progetti che mirassero alla creazione di una maggior consapevolezza dell'opinione pubblica riguardo al tema. Con privacy digitale si intendono tutte le dinamiche di controllo dell'accesso, elaborazione e utilizzo delle informazioni personali su Web e servizi collegati. Tramite le due sperimentazioni, sono emersi alcuni punti chiave che è bene tenere a mente per eventuali sviluppi futuri:

- La componente di design è ad oggi in gran parte assente e la maggior parte dei progetti viene da gruppi con una forte componente tecnica (ingegneri informatici, sviluppatori);
- I dati non vengono forniti apertamente dalle aziende ma anzi nascosti attraverso processi tecnologici che li rendono opachi. La collezione di questi dati è quindi il primo punto critico da superare per il designer e mancano degli strumenti e dei metodi adatti;

- Dati opachi non significa solo che essi non siano facilmente accessibili ma anche che, una volta ottenuti, non sia chiaro il rapporto che hanno col fenomeno che li ha generati, rendendoli di conseguenza difficilmente interpretabili;
- Dovendo sondare l'opacità di dati e processi, molto spesso i progetti hanno una forte componente sperimentale che li rende instabili. Inoltre la mancanza di risorse umane e di investimenti rallenta la risoluzione degli ostacoli che man mano si presentano.
- È possibile che per portare avanti alcune soluzioni sia necessario violare alcuni termini di servizio, entrando in una zona legale "grigia";
- Ogni scelta progettuale in questo contesto porta con sé delle considerazioni etiche che non sono trascurabili e che vanno valutate volta per volta;
- La visualizzazione di dati trova un altro importante ruolo come strumento di supervisione di processi dove l'uomo ha ceduto il proprio potere decisionale ad un algoritmo.

La tesi è stata un primo tentativo di analizzare dalla prospettiva del designer un contesto relativamente nuovo e in continua evoluzione – un'evoluzione dettata dall'incessante progresso tecnologico – com'è quello della privacy online. Si è visto come la crescente richiesta di dettagli e informazioni sulla vita e le abitudini delle persone, e le modalità con cui questo processo viene portato avanti soprattutto su Internet e nei servizi digitali presenti nella vita quotidiana, non sempre rispetti gli interessi e i diritti dei consumatori, degli utenti del web, dei cittadini. In particolare i processi di collezione e utilizzo dei dati personali da parte di aziende e governi sono spesso avvolti da un alone di opacità che previene le persone dall'avere una comprensione completa del fenomeno e la possibilità di avere voce in capitolo nel controllo e nella gestione delle proprie informazioni.

Il tema della privacy è estremamente interessante da un punto di vista metodologico perché permette all'information designer di mettere alla prova i propri strumenti in un contesto dove il dato non è facilmente visibile ed accessibile e dove i processi che si vogliono mappare sono nascosti, opachi. I due esperimenti progettuali hanno fornito la base per stimolare delle riflessioni proprio su questo. La scelta di chiamarli "esperimenti" vuole infatti sottolineare l'obiettivo ultimo di questa tesi: essi non sono la conclusione definitiva di un percorso di progetto, studiati per dare una soluzione concreta ai problemi evidenziati nell'analisi, quanto piuttosto dei tentativi utili – anzi, imprescindibili – per vedere se si è dotati degli strumenti necessari per

affrontare il progetto stesso. Un po' come quando, nei primi giorni caldi d'estate, si va al lago per rinfrescarsi e non sapendo se l'acqua sia troppo fredda per fare un bel bagno, la si sfiora appena con la punta del piede.

8.1 Sullo sperimentare

Se è vero che i due esperimenti non possono essere considerati il punto di arrivo di questo lavoro, non è giusto neanche affibbiargli il ruolo di semplici "prodotti secondari". È soprattutto attraverso l'osservazione della realtà e la risoluzione di problemi pratici che il designer genera oppure consolida la maggior parte delle considerazioni sulla propria disciplina e la ricerca è per definizione esplorativa.

La natura iterativa del processo di design li rende inoltre il punto di partenza di collaborazioni future e certamente produttive: entrambi i partner di progetto sono rimasti piacevolmente soddisfatti dall'introduzione della componente di design all'interno del proprio flusso di lavoro. Il Data Transparency Lab ha immesso nuovi risorse nell'applicazione ReCon per creare un prototipo funzionante ed affinare l'interfaccia, mentre Facebook.tracking.exposed è in continua evoluzione e necessita di elaborare nuove visualizzazioni per poter sondare l'algoritmo di Facebook da nuove prospettive. Ogni nuova iterazione è terreno fertile per testare gli strumenti e i metodi disponibili al designer.

8.2 Sul metodo

Il risultato principale di questa tesi è l'aver sottolineato la mancanza di strumenti adeguati per poter raccogliere i dati necessari a mappare i processi opachi di collezione e utilizzo delle informazioni personali. Le metodologie di cui dispone l'information designer necessitano di essere ripensate e riadattate al nuovo contesto: non vi è un cliente che fornisce apertamente i dati – per cui si potrebbe passare direttamente ad usare strumenti di analisi e trasformazione del dato – né sono sufficienti le conoscenze informatiche che ha mediamente un figura di designer-programmatore e che lo supportano nella raccolta di informazioni non direttamente disponibili (accesso a *API*[1], scrittura di *scraper*[2]).

In questo caso è indispensabile che si crei una profonda collaborazione interdisciplinare con figure provenienti dall'ingegneria informatica e dallo sviluppo software, in modo da trovare un vocabolario comune e per il designer, comprendere più a fondo le caratteristiche dei dati a cui si sta accedendo.

Una criticità emersa durante le sperimentazioni e da tenere in considerazione quando ci si vuole approcciare al tema della privacy, è il trovarsi a volte a dover operare in zone legalmente "grigie". Questo non vuol dire ovviamente compiere azioni illegali, ma può essere che esse violino i termini del servizio e le condizioni accettate nell'utilizzo.

[1] • Interfacce che permettono a due programmi di comunicare. *Sono come i connettori sul retro della TV. Ti lasciano collegare ad essa dispositivi di altre marche e sia la TV che il dispositivo sanno che fare.*
(fonte: Sideways Dictionary)

[2] • Letteralmente "raschietto". Sono dei programmi che permettono di estrarre delle informazioni da uno o più siti web, in modo automatizzato.

Eticamente questo non è necessariamente sbagliato, ma è bene che il progettista ne sia conscio. Le riflessioni sull'etica del proprio operato hanno anche portato ad una sorta di esame di coscienza sulle modalità in cui un designer dell'informazione dovrebbe approcciarsi alla gestione del dato. Nel processo di astrazione che porta alla visualizzazione, i passaggi per assicurarsi il più alto grado di anonimizzazione possibile dei soggetti rappresentati devono essere posti al centro della progettazione.

8.3 Prospettive future

L'augurio è che questo lavoro possa servire ad altri studenti o progettisti come punto di partenza per lavorare in contesti simili a quello analizzato, facendoli partire con una visione un po' più chiara di ciò che li aspetta. Tanto ancora manca da esplorare e il miglior modo per farlo è intraprendere nuovi esperimenti, che risolvano problemi pratici.

Un'ultima riflessione, che va oltre i propositi della tesi ma che è piano piano emersa durante tutto il percorso, risiede nella necessità di educare le persone all'ambiente tecnologico in cui vivono. Prima di poter creare consapevolezza sui rischi in cui essi incorrono condividendo le proprie informazioni personali online, prima di qualsiasi tentativo di coinvolgerli in un dibattito che miri a difendere i loro diritti e interessi, è fondamentale fornirli di un vocabolario con cui essi possano comprendere realmente la tecnologia che li circonda quotidianamente e che spesso finisce per essere mistificata o data per scontata. Questo è certamente un problema a cui un designer dell'informazione – ma più in generale un designer della comunicazione – può proporre soluzioni.

Un esempio in questa direzione è proprio appena uscito (Marzo 2017). Il The Washington Post in collaborazione con l'incubatore tecnologico di Alphabet, Jigsaw[3], ha sviluppato il Sideways Dictionary[4]: un particolarissimo vocabolario dove al posto di indicare i termini tecnologici attraverso le loro definizioni, essi vengono spiegati tramite analogie. Proprio da questo vocabolario sono state tratte la maggior parte delle analogie inserite nelle note della tesi. ■

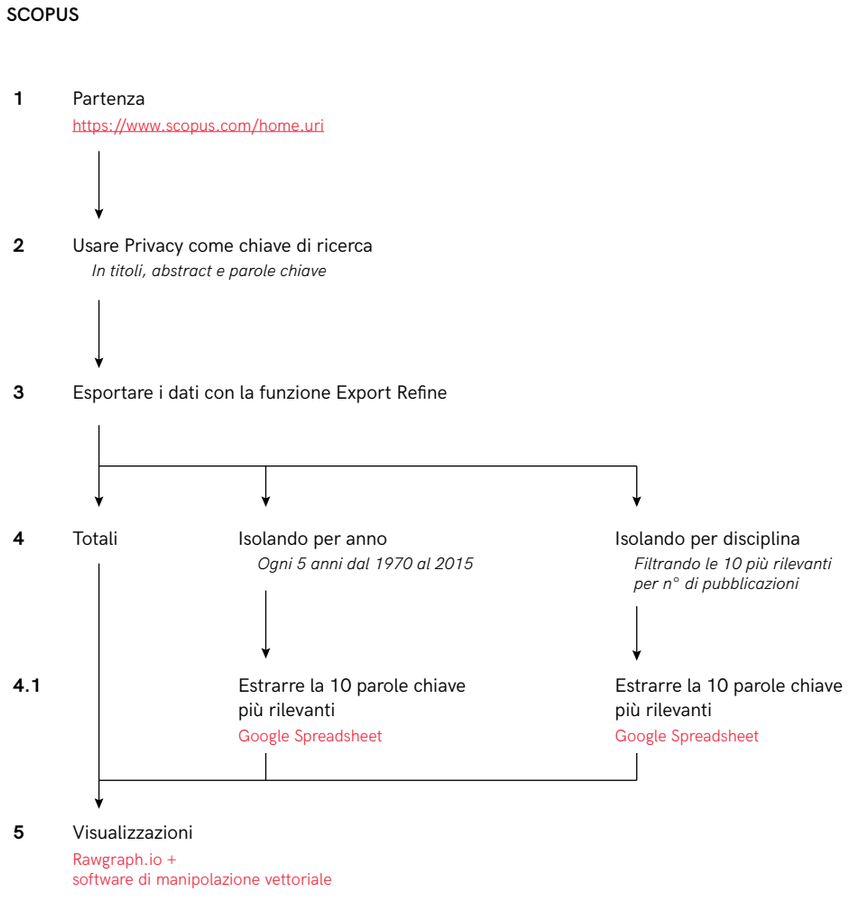
[3] ■ *https://jigsaw.google.com*

[4] ■ *https://sidewaysdictionary.com*

Compendio metodologico

In questa sezione è presente una descrizione schematica dei protocolli seguiti per portare avanti l'analisi su i *device* incontrati durante la tesi. Essi riflettono il metodo utilizzato per estrarre, trasformare e visualizzare i dati necessari da Scopus, Wikipedia, Medium, ArsTechinca e Google Takeout. L'obiettivo è far in modo che il lettore possa, volendo, riprodurre le stesse metodologie usate in questa tesi, applicandole anche a contesti differenti.

Gli strumenti e gli script presenti nei diagrammi sono tutti gratuiti e open source. L'unico database che necessita di autorizzazione è quello di Scopus.



WIKIPEDIA

1 Partenza
https://en.wikipedia.org/wiki/Privacy
https://fr.wikipedia.org/wiki/Vie_privée
https://it.wikipedia.org/wiki/Privacy
https://de.wikipedia.org/wiki/Privatsphäre

2 Creare le reti di temi correlati
Seealsology

3 Visualizzare
Gephi +
software di manipolazione vettoriale

Categorie calcolate con k-means + revisione qulitativa

MEDIUM

1 Partenza
https://medium.com/tag/privacy

2 Estrarre le parole chiave correlate
Mediumology (Github repository)

Estrarre le parole chiave negli articoli relativi al tema
Filtrando i 100 più rilevanti
Medium-topArticles (Github repository)

2.1 Catalogare qualitativamente le parole chiave

3 Visualizzare
Gephi +
software di manipolazione vettoriale

ARSTECHINICA

1 Partenza

https://arstechnica.com/tech-policy

https://arstechnica.com/search

1.1 Usare Privacy come chiave di ricerca

2 Estrarre le parole chiave
negli articoli della sezione Policy
Filtrando i 100 più rilevanti
ArsTechinca-topPolicy (Github repository)

Estrarre le parole chiave
negli articoli relativi al tema
Filtrando i 100 più rilevanti
ArsTechinca-topArticles (Github repository)

3 Visualizzare
Gephi +
software di manipolazione vettoriale

GOOGLE TAKEOUT

1 Partenza
https://takeout.google.com/settings/takeout

2 Esportare i dati di localizzazione
In formato JSON

2.1 Pulire il file esportato
Sublime Text

3 Visualizzare
D3.js

3.1 Valori giornalieri
Takeout-averageLocation
(Github repository)

Andamento orario
Takeout-locationByHour
(Github repository)

Andamento orario - giorno singolo
trasformando la data in Epoch Time
Epochconverter.com +
Takeout-locationByHour02
(Github repository)

Riferimenti bibliografici

Acquisti, A., Taylor, C., & Wagman, L. (2016). *The economics of privacy*. Journal of Economic Literature, 54(2), 442-492.

Altman, I. (1977). *Privacy regulation: culturally universal or culturally specific?*. Journal of Social Issues, 33(3), 66-84.

Anceschi, G. (1992). *L'oggetto della raffigurazione*. Etaslibri.

Baule, G., & Caratti, E. (2016). *Towards Translation Design: A New Paradigm for Design Research*. Proceedings of DRS2016: Design+ Research+ Society-Future-Focused Thinking, 1047-1060.

Bertola, P., & Manzini, E. (Eds.). (2004). *Design multiverso. Appunti di fenomenologia del design*. edizioni polidesign.

Brunton, F., & Nissenbaum, H. (2013). *Political and ethical perspectives on data obfuscation*. Privacy, due process and the computational turn: The philosophy of law meets the philosophy of technology, 164-188.

Buchanan, R. (1992). *Wicked problems in design thinking*. Design issues, 8(2), 5-21.

Burdick, A. (2009, April). *Design without designers*. In Keynote for a conference on the future of art and design education in the 21st century. University of Brighton, England.

Burrell, J. (2016). *How the machine 'thinks': Understanding opacity in machine learning algorithms*. Big Data & Society, 3(1), 2053951715622512.

Cate, F. H. (2010). *The limits of notice and choice*. IEEE Security & Privacy, 8(2).

Caviglia, G. (2013). *The design of heuristic practices. Rethinking communication design in the digital humanities*. Diss. Politecnico di Milano.

Cranor, L. F. (2012). *Necessary but not sufficient: Standardized mechanisms for privacy notice and choice*. J. on Telecomm. & High Tech. L., 10, 273.

DeCew, J. (2006). *Privacy* In Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/spr2015/entries/privacy

Diakopoulos, N. (2014). *Algorithmic accountability reporting: On the investigation of black boxes*. Tow Center for Digital Journalism, Columbia University.

Gillespie, T. (2014). *The relevance of algorithms*. Media technologies: Essays on communication, materiality, and society, 167.

Holvast, J. (2008, September). *History of privacy*. In IFIP Summer School on the Future of Identity in the Information Society (pp. 13-42). Springer Berlin Heidelberg.

Iaselli, M. & Gorla, S. (2015). *Storia della privacy*. Edizione Lex Et Ars - Editoria Professionale.

Johnson, E. J., Bellman, S., & Lohse, G. L. (2002). *Defaults, framing and privacy: Why opting in-opting out*. Marketing Letters, 13(1), 5-15.

Kang, J. (1998). *Information privacy in cyberspace transactions*. Stanford Law Review, 1193-1294.

Masud, L., Valsecchi, F., Ciuccarelli, P., Ricci, D., & Caviglia, G. (2010, July). *From data to knowledge-visualizations as transformation processes within the data-information-knowledge continuum*. In Information Visualisation (IV), 2010 14th International Conference (pp. 445-449). IEEE.

Mauri, M. (2015). *Progettare il non-finito. Diagrammi per la mappatura di fenomeni sociali attraverso il web*. Diss. Politecnico di Milano.

Meirelles, I. (2013). *Design for information: an introduction to the histories, theories, and best practices behind effective information visualizations*. Rockport publishers.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). *The privacy paradox: Personal information disclosure intentions versus behaviors*. Journal of Consumer Affairs, 41(1), 100-126.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.

Ren, J., Rao, A., Lindorfer, M., Legout, A., & Choffnes, D. (2016, June). *Recon: Revealing and controlling pii leaks in mobile network traffic*. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services (pp. 361-374). ACM.

Rogers, R. (2013). Digital methods. MIT press.

Schmidt, A. L., Zollo, F., Del Vicario, M., Bessi, A., Scala, A., Caldarelli, G., ... & Quattrociocchi, W. (2017). *Anatomy of news consumption on Facebook*. Proceedings of the National Academy of Sciences, 201617052.

Schoffelen, J., Claes, S., Huybrechts, L., Martens, S., Chua, A., & Moere, A. V. (2015). *Visualising things. Perspectives on how to make things public through visualisation*. CoDesign, 11(3-4), 179-192.

Solove, D. J. (2006). *A taxonomy of privacy*. University of Pennsylvania law review, 477-564.

Srnicek, N. (2016). *Platform Capitalism*. Polity Press.

Tufekci, Z. (2015). *Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency*. J. on Telecomm. & High Tech. L., 13, 203.

Venturini, T. (2008). *Piccola introduzione alla cartografia delle controversie Introducing the cartography of controversies*. Etnografia e ricerca qualitativa, 3, 369-394.

Indice delle tavole

FIGURE

01 Classifica delle aziende con la maggior capitalizzazione azionaria. La capitalizzazione è il valore delle azioni di una compagnia ed è usata per stimarne il valore totale. In rosso sono segnate le aziende che si occupano di IT e il valore è espresso in dollari ($). *Fonte: http://dogsofthedow.com Data: 29 Marzo 2017* *.19*

02 Evoluzione temporale del numero di articoli scientifici inerenti al tema della privacy. La crescita d'interesse appare legata agli sviluppi tecnologici dagli anni '70 ad oggi. Nella parte bassa della tavola sono riportate la parole chiave più rilevanti associate a privacy estratte da Scopus ogni 5 anni. *.26-27*

03 Evoluzione temporale della ripartizione per disciplina degli articoli scientifici inerenti al tema privacy (in percentuale). Nella parte bassa della tavola sono riportate le parole chiave più rilevanti per ogni area. *.28-29*

04 Mappa di correlazione tra pagine di Wikipedia, versione inglese. Ogni nodo è un articolo di Wikipedia e due pagine sono collegate tra loro se una cita l'altra nella sezione "Voci correlate". *.36-37*

05 Mappa di correlazione tra pagine di Wikipedia, versione francese. Ogni nodo è un articolo di Wikipedia e due pagine sono collegate tra loro se una cita l'altra nella sezione "Voci correlate". *.38-39*

06 Mappa di correlazione tra pagine di Wikipedia, versione italiana e tedesca. Ogni nodo è un articolo di Wikipedia e due pagine sono collegate tra loro se una cita l'altra nella sezione "Voci correlate". *.40-41*

07 Schema che illustra il principio delle tematiche correlate e della profondità di ricerca. *.45*

08 Rete che mostra le parole chiave associate a privacy su Medium, prendendole dalla sezione "related tags". L'intensità di colore rappresenta la rilevanza del tema sulla piattaforma. *.46-47*

09 Nuvola di parole chiave associate ai primi 100 articoli legati al tema privacy su Medium. La grandezza del carattere rappresenta la frequenza di apparizione negli articoli e il colore raggruppa le parole per categoria. *.48*

10 Mappare visualmente la co-occorrenza tra parole chiave su diversi articoli attraverso una rete. *.49*

11 Rete di co-occorrenza della parole chiave nei primi 100 articoli relativi al tema privacy su Medium, raggruppate per categoria (colore). *.50-51*

12 Le variabili che nascondono le parole chiave in un qualsiasi articolo su ArsTechnica. Per l'analisi è stata usata solo la seconda, contenente i termini inseriti dall'autore. *.53*

13 Rete di parole chiave estratte dai primi 100 articoli della sezione policy di ArsTechnica. *.54-55*

14 Rete di parole chiave estratte dai primi 100 articoli riguardanti il tema privacy su ArsTechnica. *.56*

15 Lunghezza della privacy policy di Facebook dalla sua fondazione ad oggi. La riduzione degli ultimi anni è sicuramente un passo positivo, ma la difficoltà nella terminologia usata rimane comunque un grande ostacolo. Fonti: https://web.archive.org/web/*/http://www.facebook.com/policy.php; http://niram.org/read *.69*

16 Rappresentazione da diverse prospettive del numero di accessi alla geolocalizzazione da parte di Google, dal 19 Luglio 2016 al 05 Febbraio 2017. *.71-75*

17 Schema semplificato della differenza tra algoritmi statici e dinamici. *.78*

18 Mappa che rappresenta la libertà su Internet nel mondo, secondo l'analisi svolta nel 2015-2016 da Freedom House. Fonte nazioni vettoriali: Free Vector Maps.com *.87*

19 Rappresentazione schematica della pratica dell'onion routing su cui si basa Tor. Ogni nodo ha contatti solo col nodo precedente e con quello successivo e non ha quindi modo di sapere l'intero percorso. Ispirazione: https://www.bof.nl/ons-werk/internetvrijheid-toolbox *.97*

20 Sistemi Operativi per smartphone: la divisione delle quote di mercato. Fonte: https://www.netmarketshare.com *.135*

21 Schema semplificato della differenza tra information e network flow analysis. *.139*

22-25 Tutorial iniziale di spiegazione delle funzionalità di ReCon. Vengono introdotti i concetti principali e le possibilità offerte dall'app. *.146-147*

26-27 Home page dell'applicazione. Ogni voce è espandibile con una categorizzazione più dettagliata. *.149*

28-29 Home page dell'applicazione e Menù laterale contenente le funzioni principali. *.150*

30-31 All'interno di ogni categoria vengono elencate tutte le app che hanno inviato quella data informazione personale. L'utente può cambiare il periodo d'ispezione utilizzando il calendario. *.151*

32-33 Al fine di addestrare l'algoritmo a riconoscere correttamente le informazioni personali, l'utente può espandere ogni voce dell'elenco per segnare se è corretta o meno, nonché istruire ReCon a lasciarla andare, bloccarla o sostituirla con informazioni falsificate. *.153*

34-37 La scelta dell'utente di lasciar passare, bloccare o offuscare i propri dati personali viene poi riflessa nell'interfaccia tramite tre diverse icone. *.154-155*

38-39 La funzione che permette all'utente di sostituire i propri dati con alcuni fasulli. *.156*

40-41 Schermate di altre categorie di ReCon. *.157*

42 Google Trend relativo alla parola chiave "Fake News". Dopo le elezioni americane, il dibattito ha, tra le altre cose, sollevato dubbi sul ruolo di Facebook e la sua responsabilità nello sviluppo del fenomeno. *.168*

43 Tipologie di post che facebook.tracking.exposed distingue. Nell'ordine dall'alto in basso: Feed, Friends Feed, Promoted. *.170-171*

44 Schermata iniziale della visualizzazione, ha il compito di introdurre i concetti chiave che appariranno in seguito. *.174-175*

45 Visualizzazione delle Timeline dell'utente. Ogni colonna è un momento temporale in cui si è stati su Facebook *.176-177*

46 Muovendosi sopra la legenda, maggiori informazioni vengono fornite sul significato delle diverse categorie. *.178*

47 Le Timeline possono essere riordinate secondo diversi criteri, ad esempio per tipologia di post. *.179*

48 Le Timeline possono essere riordinate secondo diversi criteri, ad esempio per ordine cronologico reale. *.180*

49 Muovendosi sopra un qualunque post, esso verrà evidenziato attraverso tutte le Timeline. *.181*

50 La caratteristica principale dei post sponsorizzati è che non hanno una data di creazione. *.182*

51 Alcuni post cambiano tipologia tra una Timeline e un'altra. *.183*

52 Alcuni post non compaiono in tutte le Timeline. *.184*

53 La visualizzazione è stato anche un'utile feedback visivo per gli sviluppatori per trovare velocemente degli errori nel codice. *.185*

IMMAGINI

01 Schermata di Medium che mostra la posizione delle tematiche correlate. *.45*

02 Struttura del JSON relativo alla Location History. Il testo in bianco sono un'interpretazione delle voci, alcune delle quali alquanto ambigue. *.72*

03 Il testo assegnato ad ogni immagine caricata su Facebook viene creato automaticamente proprio usando un algoritmo simile a quello descritto. *.80*

04-05 Screenshot estratti dal sito di security in-a-box. *.98-99*

06-07 Screenshot estratti dal sito di disconnect.me. *.100-101*

08-09 Screenshot estratti dal blog della Walker Art Center relativi al progetto ZXX. *.102-103*

10-11 Screenshot estratti dal sito di Ruin My Search History e da un articolo su Motherboard relativo all'estensione per browser. *.104-105*

12-13 Screenshot estratti dal sito di AdNauseam. *.106-107*

14-15 Screenshot estratti dal sito di Data Selfie e da un articolo di The Next Web sullo strumento. *.110-111*

16-17 Screenshot estratti dal sito di Open Data City relativo ai dati raccolti da Balthasar Glättli. *.112-113*

18-19 Screenshot estratti dal sito di Clickclickclick.click. *.114-115*

20-21 Screenshot estratti dal sito relativo al progetto delle Privacy Icons. *.116-117*

22-23 Screenshot estratti dal sito Am I Unique?. *.118-119*

24 Uno dei diagrammi di Mark Lombardi, creato per spiegare le connessioni in sistemi di relazioni che lui definiva "di uso e abuso di potere". Fonte: Ben Fry *.123*

25-26 Schermate di richiesta dei permessi di accesso a dati aggiuntivi rispettivamente su Android e iOS. *.136*

27-28 Schermate attuali di ReCon che illustrano il processo di catalogazione e controllo dei dati personali. Fonte: https://recon.meddle.mobi/app-pii.html *.143*

29-30 Schermate attuali di ReCon che illustrano il processo di catalogazione e controllo dei dati personali. Fonte: https://recon.meddle.mobi/app-pii.html *.144*

31 Schermata del News Feed di Facebook. *.164*

Stampato a Colnago (MB) –Aprile 2017