Alma Mater Studiorum · Università di Bologna

SCUOLA DI SCIENZE

Corso di Laurea in Informatica per il management

STUDIO E ANALISI DI PROBLEMATICHE DI INTERDEPENDENT PRIVACY ONLINE

Relatore:
Dott. LUCA BEDOGNI

Presentata da:
LUCIA CARLOTTI

Sessione III
Anno Accademico 2017-2018

Introduzione

L'idea di questa tesiè nata dalla lettura dell'articolo *"The Eternal Value of Privacy"* in cui Bruce Schneier [1] risponde a coloro che, favorevoli alle misure di sorveglianza, di solito argomentano con "se non si sta facendo nulla di male, allora non si ha nulla da nascondere". In realtà l'autore richiama al fatto che la privacy non è semplicemente nascondere qualcosa di sbagliato: è un diritto umano non solo necessario a mantenere una condizione di dignità e rispetto ma anche a proteggerci dagli abusi di coloro che si trovano al potere. La privacy, infatti, ci tutela dall'uso improprio delle informazioni di sorveglianza che possono essere impiegate per osservarci, per essere vendute a società di marketing, per spiare nemici politici e per una moltitudine di altre azione lesive della libertà. Il dibattito non dovrebbe quindi focalizzarsi sul tema sicurezza vs privacy in quanto la vera scelta si può avere solo tra libertà vs controllo. E la libertà richiede sicurezza senza intrusione, una sicurezza con la privacy. Partendo da questa riflessione, ho quindi realizzato quanto le informazioni che quotidianamente diffondiamo, volontariamente o involontariamente, sul web possano essere potenzialmente molto pericolose per la nostra privacy. In un mondo interconnesso come il nostro, però, non dobbiamo preoccuparci solamente delle tracce che lasciamo ma anche dei dati che le persone vicine condividono riguardo a noi. Si tratta di un concetto, chiamato Interdependent Privacy, che è emerso soprattutto a causa dell'enorme diffusione dei social network, basati proprio sulla condivisione delle proprie attività e pensieri. Per cercare di capire quanto gli altri influiscano sulla nostra privacy online, ho analizzato le problematiche legate ai contenu-

ti divulgati da noi stessi e dagli altri nei social network, i permessi che gli sviluppatori richiedono con l'installazione di app di terze parti, i rischi alla privacy legati alle informazioni sulla posizione ed infine l'influenza dei collaboratori nelle app che si appoggiano a servizi di cloud. Questo documento è strutturato come segue: nel Capitolo 1 si riporta la normativa in materia di protezione dei dati personali con particolare riferimento al GDPR, nel Capitolo 2 si affrontano in modo specifico i rischi legati all'Interdependent Privacy sopra riportati, nel Capitolo 3 infine si illustrano le principali problematiche che gli studiosi di questo fenomeno si sono trovati ad affrontare.

Indice

Elenco delle figure

Capitolo 1

Background

1.1 La protezione dei dati personali e i Big Data

Nella società odierna la protezione dei dati personali e il loro corretto uso assume un ruolo sempre più importante. I nostri dati sono infatti indispensabili per utilizzare tutti gli strumenti tecnologici al massimo delle loro potenzialità e per questo i soggetti che li trattano sono in costante crescita da anni. Ancora maggiore è però l'offerta di dati personali derivanti principalmente dalla diffusione di applicazioni che sui dati fondano il loro successo, come ad esempio social networks, piattaforme di commercio elettronico e cloud computing [2]. La normativa europea di riferimento per la protezione dei dati personali, il GDPR (General Data Protection Regulation) [3], definisce all'art. 1 il dato personale come *"qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»)"*. Si considera identificabile la persona fisica che può essere riconosciuta, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale. La raccolta di tutti questi dati è caratterizzata da volume, velocità e varietà. Infatti il numero di dispositivi

oggi connessi alla rete fornisce una quantità di dati ed una facilità di raccolta senza precedenti nella storia umana. I dati vengono ormai prodotti con continuità, in maniera dinamica e non più statica, e le nostre abitudini vengono continuamente tracciate e aggiornate. Le informazioni di una persona possono poi arrivare da una estesa varietà di fonti e in formati diversi, come ad esempio pubblicando qualcosa su Facebook o Twitter, o attraverso sensori presenti nei nostri dispositivi. Questo ovviamente porta a una maggiore complessità di gestione e, a causa di questa eterogeneità di sorgenti, diventa più complesso accertarsi della veridicità dei dati. Ciascuno di questi dati poi, preso singolarmente, potrebbe non avere molto significato, ma l'estrazione di valore da questa raccolta, che richiede tecnologie e algoritmi sempre più sofisticati capaci di trovare relazioni significative all'interno di questi dati, riesce a fornire l'informazione di cui si necessita. Questo fenomeno si definisce Big Data e insieme a enormi potenzialità di sviluppo ci espone anche a notevoli rischi per la nostra privacy. Come detto, la nozione di dato personale ha un significato piuttosto ampio e include sia dati che permettono di identificare direttamente una persona, come ad esempio il nome, l'indirizzo o il codice fiscale, sia indirettamente identificativi, cioè riconducibili a dati direttamente identificativi attraverso codici o tabelle di passaggio intermedie come i codici utente o i nickname con cui veniamo identificati presso i vari fornitori di servizio con cui entriamo in contatto. Ogni volta in cui è possibile isolare un soggetto in un database, collegare il dato che si sta trattando a dati relativi allo stesso soggetto presenti in diversi database, o ancora dedurre, con probabilità significativa, una caratteristica di un soggetto del trattamento di un dato, per prendere decisioni riguardanti l'utente, anche senza che sia nota la sua identità, siamo in presenza di un trattamento di dati personali. Il titolare di questo trattamento è colui che determina le finalità e gli strumenti impiegati per il trattamento dei dati personali. Egli si assume anche notevoli responsabilità da cui derivano obblighi specifici come riscontrabile dai principi elencati nell'art.5 del GDPR. Il primo di questi è l'obbligo di rispettare i principi sulla qualità dei dati, per cui questi devono fornire una giusta

rappresentazione della persona, e quindi essere aggiornati, ed essere trattati solo se l'individuo a cui si riferiscono ne è realmente consapevole. A questo segue poi l'obbligo di rispettare le finalità del trattamento, le quali devono essere esplicite e definite prima dell'avvenuta di questo. Inoltre, secondo il principio di necessità, devono essere raccolti solo i dati che rispondono alle finalità del trattamento e questi devono essere trattenuti solo per il tempo necessario al compimento dei suddetti scopi. Presupposto necessario al trattamento dei dati da parte del titolare è poi la legittimità del trattamento. Per quanto riguarda i Big Data, il consenso, l'adempimento di obblighi contrattuali e l'interesse legittimo del titolare assumono un ruolo fondamentale. In particolare, il consenso rappresenta lo strumento di controllo migliore a disposizione delle persone: per essere valido deve essere dato liberamente, essere informato, specifico e inequivoco. Per "dato liberamente" si intende che colui che potrebbe dare il consenso deve avere la possibilità di accettare o rifiutare il trattamento dei suoi dati personali. Il consenso è poi *"informato"* se la persona ha tutte le informazioni necessarie a permettendogli di crearsi una propria opinione sul trattamento e *"specifico"* se riguarda gli scopi per cui i dati sono raccolti e trattati, inoltre è *"inequivoco"* se l'autorizzazione al trattamento è data senza ambiguità prima che il trattamento si svolga. Il titolare del trattamento ha infatti degli obblighi di trasparenza, come ad esempio la trasparenza sulle finalità del trattamento, ed è pienamente responsabile per la sicurezza dei dati come definito nell'art. 32 del GDPR.

1.2 GDPR

Il quadro giuridico europeo riconosce alle persone il diritto alla protezione dei propri dati personali tramite una serie di obblighi per chi vuole trattare i nostri dati. Questi principi sono contenuti nel General Data Protection Regulation (GDPR), un regolamento dell'Unione Europea volto a stabilire norme relative alla protezione delle persone fisiche con riguardo al trattamento dei dati personali e alla libera circolazione di tali dati (art. 1 comma

1 GDPR). Lo scopo principale di questa normativa, in vigore dal 25 maggio 2018, è quello di riformare, aggiornare e uniformare la legislazione in materia così da renderla più solida e coerente tra i vari Paesi membri dell'UE.
Gli articoli che seguono, sono forse i più rilevanti:

- **Art 7 – Condizioni per il consenso**
  L'articolo 7 sancisce che, nel caso in cui il trattamento sia basato sul consenso, deve essere sempre dimostrabile che l'interessato ha prestato il proprio consenso al trattamento dei suoi dati personali. Inoltre l'interessato ha il diritto di revocare il proprio consenso in qualsiasi momento.

- **Art 12 - Informazioni, comunicazioni e modalità trasparenti per l'esercizio dei diritti dell'interessato**
  Il titolare del trattamento deve adottare le misure appropriate per fornire all'interessato tutte le informazioni e le comunicazioni relative al trattamente in forma concisa, trasparente, intelligibile e facilmente accessibile, in un linguaggio semplice e chiaro. Le informazioni devono essere fornite per iscritto o con altri mezzi elettronici. In caso di richiesta dell'interessato possono anche essere fornite oralmente, ma deve essere comprovata l'identità dell'interessato con altri mezzi.

- **Art 13 - Informazioni da fornire qualora i dati personali siano raccolti presso l'interessato**
  In caso di raccolta presso l'interessato di dati che lo riguardano, il titolare del trattamento deve fornire all'interessato, nel momento in cui i dati personali sono ottenuti, le seguenti informazioni: l'identità e i dati di contatto del titolare del trattamento, i dati di contatto del responsabile della protezione dei dati, le finalità del trattamento a cui sono destinati i dati personali, gli interessi perseguiti dal titolare del trattamento o da terzi, i destinatari dei dati personali, se presente l'intenzione del titolare del trattamento di trasferire i dati personali a un altro paese o a un'organizzazione internazionale, il periodo di conservazione dei dati

personali, l'esistenza del diritto dell'interessato di chiedere al titolare del trattamento l'accesso ai propri dati personali e la cancellazione o rettifica degli stessi e la limitazione del trattamento o l'opposizione al loro trattamento, inoltre deve specificare l'esistenza del diritto di revoca del consenso in qualsiasi momento.

- **Art 14 - Informazioni da fornire qualora i dati personali non siano stati ottenuti presso l'interessato**
  Se i dati non sono ottenuti dall'interessato ma presso un terzo, il titolare del trattamento deve fornire all'interessato tutte le informazioni sul trattamento, come ad esempio l'identità e i dati del titolare del trattamento e le finalità del trattamento a cui sono destinati i dati personali. Inoltre deve fornire all'interessato le informazioni necessarie per garantire un trattamento corretto e trasparente, come ad esempio il periodo di conservazione dei dati personali. In breve il titolare del trattamento deve fornire all'interessato tutte le informazioni per la salvaguardia dei suoi dati.

- **Art 15 - Diritto di accesso dell'interessato**
  L'interessato ha il diritto di ottenere dal titolare del trattamento la conferma che sia o meno in corso un trattamento dei suoi dati personali, se è in corso, ha il diritto di ottenere l'accesso ai suoi dati personali e ad alcune informazioni, quali: le finalità del trattamento, le categorie di dati personali, i destinatari a cui i dati personali sono stati o saranno comunicati, il periodo di conservazione previsto, il diritto dell'interessato di chiedere al titolare del trattamento la rettifica o la cancellazione dei propri dati personali, ect. Inoltre se i dati personali sono trasferiti a un paese terzo o a un'organizzazione internazionale, l'interessato ha il diritto di essere informato dell'esistenza di garanzie adeguate ai sensi dell'articolo 46 relative al trasferimento. In più il titolare del trattamento deve fornire una copia dei dati personali oggetto del trattamento.

- **Art 17 – Diritto alla cancellazione (diritto all'oblio)**
  L'interessato ha il diritto di ottenere dal titolare del trattamento la cancellazione dei dati personali che lo riguardano senza ingiustificato ritardo, e il titolare del trattamento ha l'obbligo di cancellare senza ingiustificato ritardo gli stessi. Anche se questi sono stati resi pubblici rimane l'obbligo di cancellazione. Il titolare del trattamento, tenendo conto della tecnologia disponibile e dei costi di attuazione, deve adottare misure ragionevoli, anche tecniche, per informare coloro che stanno trattando tali dati della richiesta dell'interessato di cancellare qualsiasi link, copia o riproduzione dei suoi dati personali. Le disposizioni di cui sopra non si applicano però nella misura in cui il trattamento sia necessario per l'esercizio del diritto alla libertà di espressione e di informazione, per l'adempimento di un obbligo legale che richieda questo trattamento, per motivi di interesse pubblico o per l'esercizio o la difesa di un diritto in sede giudiziaria.

- **Art 25 – Protezione dei dati fin dalla progettazione e protezione per impostazione predefinita**
  Il titolare del trattamento deve mettere in atto misure tecniche ed organizzative adeguate, tra cui la pseudonimizzazione, volte ad attuate in modo efficace i principi di protezione dei dati. Per impostazione predefinita devono essere trattati solo i dati personali necessari per ogni specifica finalità del trattamento.

- **Art 32 – Sicurezza del trattamento**
  Il titolare e il responsabile del trattamento devono mettere in atto misure tecniche e organizzative adeguate a garantire un livello di sicurezza adeguato al rischio, che comprendono anche la pseudonimizzazione e la cifratura dei dati personali.

Il GDPR spinge inoltre le organizzazioni a pseudonimizzare i dati in modo da limitare i rischi di violazioni di privacy e furti d'identità, garantendo la tutela dei dati sensibili di persone fisiche e/o giuridiche [4]. Si tratta di una

tecnica che consiste nel conservare i dati in una forma tale da impedirne l'attribuzione ad un soggetto specifico senza l'utilizzo di informazioni aggiuntive. Tali informazioni devono essere conservate separatamente, come ad esempio in differenti server, e soggette a misure tecniche e organizzative intese a garantire che i dati personali non siano attribuiti a una persona identificata o identificabile. Le informazioni aggiuntive vengono comunemente definite "chiavi" e consistono in un serie di bit impiegati nelle operazioni di codifica e decodifica. Al contrario dell'anonimizzazione, che consiste nel rendere un dato completamente anonimo rendendo impossibile associarlo ad un soggetto, nella pseudonimizzazione il legame tra il dato e la persona rimane ma richiede mezzi onerosi per il ricongiungimento. Il principale strumento attraverso cui questo diventa operativo è la crittografia o cifratura dei dati: una tecnica che consente di modificare un'informazione digitale rendendola impossibile da leggere per un osservatore esterno ma in una maniera tale per cui il destinatario la possa recuperare interamente attraversi l'uso delle specifiche chiavi.

1.3 La privacy

Il GDPR tenta quindi di proteggere la privacy dei cittadini europei in modo organico, ma cosa si intende per privacy?
Con il termine inglese *"privacy"*, che evoca significati a volte mutevoli traducibili in italiano con le parole *"riservatezza"* o *"privatezza"*, si indica il diritto alla riservatezza della propria vita privata [5], o secondo la formulazione dei giuristi statunitensi Louis Brandeis e Samuel Warren, che furono probabilmente i primi a formulare una legge sulla riservatezza, il diritto di essere lasciati in pace (*"the right to be let alone"*). L'invenzione del diritto alla privacy come concetto giuridico risale infatti ad un articolo pubblicato su una rivista giuridica in cui i due avvocati di Boston specificano questo diritto alla privacy come il diritto di scegliere se condividere o non condividere con gli altri le informazioni sulla propria vita privata, le proprie abitudini, i propri

atti e le proprie relazioni [6]. Questo poiché si erano resi conto che quando le informazioni sulla vita privata di un individuo vengono rese disponibili ad altri,questo può portare ad influenzare o ledere la parte più intima della sua personalità (*"his estimate of himself "*).
Nell'ordinamento italiano la sfera privata degli individui riceve un tutela ampia ma molto frammentaria, infatti viene tutelata sia dalla Costituzione (in particolare agli art. 2 in cui si incorpora la privacy nei diritti inviolabili dell'uomo,13,14,15,21) sia dalla carta dei diritti fondamentali dell'Unione Europea (artt.7-8), sia da numerose leggi ordinarie [7] . Per assicurare la tutela dei diritti e delle libertà fondamentali e il rispetto della dignità nel trattamento dei dati personali è stata istituita, dalla cosiddetta "legge sulla privacy" n. 675 del 31 dicembre 1996, un'autorità amministrativa indipendente *"Il Garante per la protezione dei dati personali"*. Il Garante è anche l'autorità di controllo designata ai fini dell'attuazione del Regolamento Europeo generale sulla protezione dei dati personali (GDPR) [8] .
A differenza della parola che è rimasta sempre la stessa, il significato di privacy si è evoluto nel tempo anche in relazione all'evoluzione tecnologica che dalla fine del XIX secolo (quando fu scritto [6]) ad oggi è intercorsa. Inizialmente era infatti riferita alla sfera della vita privata degli individui, nel corso degli ultimi decenni però ha subito un'evoluzione estensiva, arrivando ad indicare anche il diritto al controllo sui propri dati personali. Quindi, il significato attuale di privacy comprende anche il diritto dell'individuo di controllare che le informazioni che lo riguardano vengano guardate e trattate solo da soggetti autorizzati e in caso di necessità. Infatti con l'evoluzione delle nuove tecnologie si è reso molto più facile il reperimento di informazioni su un individuo, ad esempio tramite la tracciabilità dei dispositivi mobili o la scoperta degli indirizzi email delle persone, si può andare incontro al fenomeno dello spamming. Sul web ogni gestore/provider può acquisire i dati personali, come ad esempio lo stato di salute, il reddito e le preferenze d'acquisto degli utenti, questi dati vengono poi inferenzialmente connessi per mezzo di specifici algoritmi che "costruiscono" la personalità digitale dei singoli utenti

e ne predicono i comportamenti futuri. Infatti anche se ciascuno di questi dati, preso singolarmente, potrebbe non avere molto valore, se unito ad altri e messo nel contesto giusto potrebbe rivelare molto sull'utente, per questo è indispensabile tutelare le informazioni che circolano sulle persone. Anche se ormai sono proprio gli utenti stessi, nella maggior parte dei casi, a fornire spontaneamente i loro dati personali con la self-disclosure (auto-rivelazione di sé).

Mentre quindi esiste una vasta normativa atta a regolare la raccolta e il trattamento dei dati personali da parte dei fornitori di servizi, i quali devono mettere in atto specifici strumenti di tutela come per esempio la pseudonimizzazione, gli utenti non hanno alcun controllo o quasi sulle loro informazioni se queste non vengono fornite in prima persona. Gli amici, i colleghi e le persone vicine possono infatti mettere a rischio la nostra privacy condividendo informazioni online potenzialmente utilizzabili da chiunque ne entri in possesso senza che la persona a cui sono riferite le possa cancellare agevolmente (come nel caso di dati auto pubblicati, anche se comunque l'inserimento dei propri dati, dei propri commenti, delle proprie foto su un social network costruisce una memoria storica sulle attività e sulla personalità di un soggetto quasi impossibile da far scomparire anche quando questo lo vorrebbe) o anche solo venirne a conoscenza. Questo fenomeno è definito *"interdependent privacy"* e verrà discusso in modo approfondito nel prossimo capitolo.

Capitolo 2

Intedependent privacy

Le persone condividono ogni giorno in rete enormi quantità di informazioni personali e opinioni, tra di loro e con differenti fornitori di servizi. Non tutte queste li riguardano in prima persona: spesso infatti si condividono anche, magari involontariamente, dati che riguardano i propri amici o conoscenti, compromettendo così anche la privacy di questi ultimi. Clarke fornisce una definizione estremamente attuale di *"privacy"* [9] intesa come l'interesse che le persone hanno nel mantenere uno spazio personale, cioè libero dalle interferenze di altre persone e/o organizzazioni. Egli ne identifica 4 dimensioni: *privacy of the person*, *privacy of personal behaviour*, *privacy of personal communication* e *privacy of personal data*. Biczòk e Chia [10], partendo da questa definizione, strutturano i rischi della privacy online in 3 categorie:

- Personale: in cui si hanno perdite di informazioni sull'utente e sui dati legati alle suoi comportamenti.
- Relazionale: che riguarda il modo in cui un utente si relaziona e comunica con gli altri.
- Spaziale: che si riferisce all'invasione dello spazio virtuale di un utente. Anche in quest'ultimo caso i rischi sono notevoli poiché gli spazi virtuali, come blog e social media, vengono associati agli utenti con possibili ripercussioni anche nel mondo reale.

Tutte queste dimensioni sono a rischio a causa di quello che condividiamo online. In un contesto interconnesso come quello attuale, però, la privacy dei singoli utenti è influenzata anche dalle decisioni e dalle azioni altrui. Il termine che indica questo fenomeno, l'interdependent privacy, è stato coniato per la prima volta da Biczok et Chia nel loro studio sulle autorizzazioni concesse alle app di terze parti su Facebook [10]. Si poteva comunque parlare di interdependent privacy anche prima dell'avvento dei social network: si pensi all'esempio di Alice che mette in imbarazzo Bob affiggendo per tutta la scuola volantini che ritraggono il malcapitato in una foto "divertente" oppure pubblicandola sul suo blog. I social network hanno solo reso la condivisione dei dati estremamente più semplice: per mettere in imbarazzo Bob è sufficiente che Alice posti la suddetta foto sulla sua bacheca o su un gruppo e che vi taggi il protagonista perché la foto sia vista da un numero potenzialmente più vasto di persone con conseguente maggior imbarazzo. Mentre l'auto divulgazione può essere contenuta, la diffusione di informazioni da parte di altri, a volte anche senza che siano pienamente consapevoli dei rischi potenziali a cui espongono le persone a loro vicine, è al di fuori del controllo individuale. Nelle successive sezioni analizzeremo le implicazioni della interdependent privacy in tre situazioni: le informazioni condivise dai nostri "amici" con cui siamo collegati nei diversi social network online (OSN), le app di cloud e la geolocalizzazione dei dispositivi mobili .

2.1 Online Social Network

Sebbene vi sia un gran numero di servizi online in cui l'interdependent privacy è presente, tra cui blog, forum, portali di condivisione di foto e video, questo fenomeno è particolarmente evidente nei Social Network Online (OSN).
OSN come Facebook, Google+, Twitter e Instagram hanno attirato miliardi di utenti in tutto il mondo e sono sempre più coinvolti nelle attività quotidiane delle persone, grazie anche all'evoluzione degli smartphone che ha

permesso agli utenti di essere sempre connessi ovunque si trovino. Questa popolarità è dovuta al fatto che gli OSN si presentano come piattaforme multifunzionali (all-in-one) che offrono la possibilità di generare, pubblicare, commentare e condividere contenuti e di interagire con tutto il mondo gratuitamente. Alla base del successo dei social network online vi è il re-sharing di immagini e contenuti che può essere fatto sia all'interno della stessa piattaforma che su altre esterne, è il caso ad esempio di Facebook e Instagram i cui contenuti sono condivisibili su entrambe. Il fatto che contenuti generati da altri, come commenti, video, link e foto, che rivelano informazioni su altre persone vengano postati, visti e ri-condivisi senza il consenso di coloro a cui sono riferiti non è certamente infrequente sugli OSN.
Un interessante esempio di interdependent privacy su un OSN è appunto il tag delle foto: Alice tagga Bob in una foto e la condivide senza il consenso esplicito di quest'ultimo. Sia gli amici di Alice che quelli di Bob hanno accesso alla foto per le impostazioni predefinite. Se quest'ultimo non desidera che i suoi amici la vedano, e ad esempio rimuove il tag, la foto rimarrà comunque visibile sul profilo di Alice, a meno che lei non la elimini o che non violi le linee guida del social.
La maggior parte di social network prevede che l'utente riceva una notifica ogni qual volta venga taggato in un contenuto ma se la persona che lo posta non esegue il passaggio del tag, al momento, non vi è nessun modo di essere informati su media potenzialmente rilevanti per un soggetto.
Gli OSN sono diventati una grande risorsa di informazioni gratuite che possono essere facilmente accessibili, dedotte e utilizzate in qualsiasi momento. Poiché consentono agli utenti di condividere i contenuti in modo selettivo e limitare l'accesso alle informazioni all'interno degli spazi virtuali a loro destinati, con però poco o nessun controllo sui contenuti generati dagli altri, possono portare dei rischi alla privacy e sicurezza dei loro utenti. La gestione per un utente delle informazioni che condivide può infatti essere molto complessa, ma la gestione delle informazioni che gli altri condividono riguardo a lui diventa ancora più complicata e impegnativa. Al di fuori del proprio profi-

lo non vi è infatti praticamente alcun controllo riguardo i contenuti pubblicati da altri utenti. Ad esempio, se Alice pubblica un commento nello spazio di Bob, egli non può specificare quali utenti possono visualizzare il commento. In un altro caso, quando un utente carica una foto e tagga gli amici presenti, questi non possono limitare chi può visualizzarla, come invece potrebbero fare per un contenuto da loro pubblicato. Uno studio che evidenzia questo problema è [11] di Chutikulrungsee et al. che presenta i risultati preliminari di una ricerca sulla divulgazione di dati altrui nel social network Facebook. Infatti, come detto, un grande problema relativo agli OSN è la divulgazione di informazioni che riguardano altri da parte degli utenti e di terzi durante l'interazione o le attività sui social, in particolare tramite il tagging e il re-sharing. Per lo studio di questi fenomeni gli autori hanno deciso di sottoporre a degli utenti un sondaggio online e successivamente un'intervista semi-strutturata arrivando a dividerli in due gruppi: il *"discloser"* colui che divulga informazioni di amici, amici di amici o estranei, e il *"disclosed"*, il quale può essere utente o non utente di quel OSN ed è colui a cui si riferisce il dato. Attraverso questo lavoro è emerso che il 79,56% degli intervistati è stato taggato in almeno una foto e il 30,1% degli intervistati ha chiesto di rimuovere una foto di gruppo che lo ritraeva. Il 99,26% è stato taggato in un qualsiasi contenuto su Facebook, il 26,47% ha chiesto la rimozione di un tag mentre il 25,74% ha chiesto la rimozione di una foto. Solo al 16,65% è stato chiesto da altri di rimuovere una foto da loro postata.
Oltre al tag, anche il re-sharing è un'attività popolare di diffusione degli altrui contenuti su Facebook. La maggior parte degli intervistati ri-condivide i contenuti pubblicati da altri, infatti l'84,78% dichiara di aver ri-condiviso post durante la sua attività nel social network e l'85,78% di aver ri-condiviso links, mentre per le foto la percentuale di ri-condivisioni è del 75,36% e per i video del 69,56%. Questo ci fa chiaramente capire come generalmente la divulgazione dei dati superi il controllo del singolo utente e che per gestire questo fenomeno spesso una persona deve intervenire offline chiedendo al discloser di rimuovere o eliminare i contenuti a lui riferiti. Su quanto questo sia

efficace si può trovare conferma nel sondaggio prima citato, da cui è emerso che il 23,53% degli intervistati ha rimosso o eliminato una foto, il 23,13% un tag, il 17,78% un post e il 16,30% un commento su richiesta specifica di un amico.

2.1.1 Auto- e co-divulgazione delle informazioni personali

Viene quindi spontaneo cominciare a pensare alla privacy come a una questione comune, in quanto, grazie a questa natura così interattiva della comunicazione interpersonale sugli OSN, le informazioni private degli utenti vengono diffuse non solo dalle loro rivelazioni volontarie su se stessi, ma anche dalle attività dei loro "amici". Infatti, come già detto in precedenza, sebbene gli utenti siano liberi di selezionare quali informazioni personali divulgare, e quindi condividere con il mondo, spesso non sono in grado di controllare le divulgazioni che li riguardano da parte di altri utenti e l'utilizzo che verrà fatto di queste informazioni private. In [12] Alsarkal et al. hanno studiato la perdita della privacy di una persona causata dalla auto- e co-divulgazione, e quindi in che misura la divulgazione di informazioni da parte di terzi su un individuo possa portare a reali danni alla sua privacy, utilizzando come caso di studio l'OSN Twitter. Hanno selezionato un campione di utenti collocati negli Stati Uniti e recuperato i loro tweet insieme a quelli dei loro follower, identificando le informazioni sull'identità di un utente che possono essere dedotte dai suoi tweet, le informazioni auto-rivelate, e quelle rese note dai tweets dei suoi follower, ovvero le informazioni co-divulgate. Nel loro articolo hanno analizzano quindi questi due tipi di divulgazione di informazioni negli OSN: l'auto-divulgazione, vale a dire la divulgazione di informazioni private di un utente sull'OSN da parte sua, e la co-divulgazione, ovvero la divulgazione delle informazioni private dell'utente da parte di altri utenti. Gli autori si sono posti come quesiti quali sono le informazioni personali che sono più suscettibili ad essere rivelate attraverso l'auto-divulazionione rispetto alla co-divulgazione e viceversa, e se la co-divulgazione aumenti significati-

vamente il rischio di privacy di un individuo, ovvero quale sia la probabilità che una persona sia identificata nella vita reale in base alle attività che la riguardano nel OSN. Molto spesso sono gli utenti stessi a rivelare la propria identità nel profilo poiché, grazie a queste informazioni, la famiglia e gli amici possono identificarli e creare delle connessioni sul social network. Però altri elementi sull'identità possono essere divulgati attraverso le interazioni sociali tra l'utente OSN e le sue connessioni, avvenendo quindi sotto forma di co-divulgazione. Ad esempio, consideriamo un utente molto attento alla propria privacy che sceglie di non pubblicare mai dei post con la geo-localizzazione incorporata. Se uno dei suoi amici però lo tagga in un post geo-localizzato rivela la sua posizione attraverso la co-divulazione. Gli autori cercano di quantificare la perdita della privacy calcolando non solo il rischio complessivo per un utente sull'ONS di essere riconosciuto, ma anche l'effetto della rivelazione di singoli elementi di identità o la loro combinazione per ottenere l'identificazione. Hanno scelto Twitter come piattaforma OSN per due ragioni principali, la prima è che i suoi utenti spesso postano tweet che rappresentano delle interazioni dirette tra di loro, che possono essere liberamente accessibili da terze parti e che costituiscono quindi una fonte di raccolta di co-divulgazione di informazioni private. In secondo luogo Twitter fornisce delle API gratuite che facilitano l'estrazione e l'analisi di grandi quantità di dati sulle interazioni dell'utente. Dopo aver esaminato tutti i dati raccolti gli autori hanno identificato sei categorie di elementi di identità che sono spesso rivelati attraverso l'auto- o co-divulgazione su Twitter: nome, location, sesso, compleanno, età e relazioni familiari. Di ognuna di queste categorie le informazioni raccolte possono avere un diverso tipo di approfondimento, ad esempio il nome reso noto potrebbe essere solo il nome dell'utente, solo il cognome o entrambi. Anche le informazioni sulla posizione potrebbero essere sull'area metropolitana, sul codice di avviamento postale o sull'indirizzo di domicilio esatto. La tabella presente nella figura 2.1 riassume i risultati mostrando come spesso si verifichino l'auto- e la co-divulgazione su Twitter per ogni categoria individuata dagli autori. Inoltre dallo studio è emerso che per

Identity Elements	Self-Disclosures	Co-Disclosures
Name	The name could be extracted from the name attribute or inferred from the screen name of the user profile.	The name could be extracted from tweets that mention the user.
Location Information	The location could be extracted from the location attribute of the user profile or inferred from geo-enabled tweets.	The location could be inferred from geo-enabled tweets that tag the user.
Gender	The gender of the user could be inferred from the bio attribute of the user profile as some users describe themselves as a father or mother or wife or husband.	The gender could be inferred from tweets mentioning the user that include relational or gender specific data such as sister or bro.
Birthday	Disclosed by user tweets that mention birthday.	Birthday information could be inferred from co-owners' tweets of birthday wishes to the user, e.g., "Happy Birthday"
Age	Disclosed by user tweets that mention age, year of birth, etc.	Age information could also be inferred from birthday wishes, e.g., "Happy 43rd Birthday!"
Family Relationships	Relationships could be inferred from user tweets mentioning relatives such as siblings, spouses, and parents.	Relationships could be inferred from tweets coming from relatives such as "Miss you mom" or "Happy Birthday Dad", etc.

Figura 2.1: Elementi d'identità di auto- e co-divulgazione [12]

nome, età e codice postale l'auto-divulgazione è più ricorrente rispetto alla co-divulgazione, mentre invece per il compleanno e il genere la co-divulgazione è più frequente. Per rispondere al quesito su quanto la co-divulgazione aumenti il rischio per la privacy, gli autori hanno misurato la quantità di perdita della privacy risultante dalla divulgazione di elementi di identità, calcolando prima la quantità di perdita causata dall'auto-divulgazione aggiungendo poi quella dovuta alla co-divulgazione. Ad esempio: se un utente rende noto il proprio nome, ma non il genere e un suo follower lo co-divulga attraverso un tweet in cui dice di essere a una "serata tra ragazze" insieme a lei, la perdita totale di privacy aumenta a causa del fatto che entrambe le informazioni sono trapelate. Dallo studio emerge inoltre che meno un utente auto-rivela informazioni su se stesso più i suoi follower sono propensi a co-divulgare informazioni su di lui. Anche se questo risultato sembra contrario a ciò che si potrebbe intuire, questo comporta in realtà che, per lo stesso comportamento di co-divulgazione, minori sono le informazioni che un utente auto-rivela, maggiore sarà il danno alla sua privacy che si avrà in caso di co-divulgazione.

Infatti, se una persona ha già condiviso il proprio nome e cognome, la città in cui risiede e la sua data di nascita, un eventuale follower che faccia un tweet il giorno del suo compleanno taggandolo in un locale della sua città durante la festa che ha organizzato gli arreca un danno di privacy minore, rispetto al caso in cui tutti questi dati fossero stati mantenuti privati, poiché tutte queste informazioni erano già note. Twitter consente agli utenti di modificare le loro impostazioni sulla privacy per nascondere i loro twitters e l'elenco dei followers e followees. Però gli autori hanno scoperto che rendere il proprio account Twitter privato non è sufficiente ad eliminare la co-divulgazione, infatti le informazioni rivelate da altri restano pubblicamente accessibili sui profili di coloro che le hanno create, sono solo più difficili da trovare per un potenziale interessato. Quindi, ad esempio, anche se Alice imposta il proprio account Twitter come privato, si può ancora trovare il suo nome nell'elenco dei follower di Bob, se Bob imposta il proprio account come pubblico. Allo stesso tempo, il tagging di Alice rimane pubblicamente visibile nei tweet postati da Bob. Le politiche di Twitter rendono quindi impossibile per l'utente bloccare la co-divulgazione attraverso la regolazione delle sue impostazioni sulla privacy, anche se impostando il proprio account come privato si rende più difficile scoprire le informazioni co-divulgate, poiché bisognerebbe trovare i follower dell'utente senza accedere a un elenco completo.
Esistono infine due importanti osservazioni associate sia all'auto che alla co-divulgazione. La prima è che la divulgazione delle informazioni potrebbe derivare da contenuti oppure da metadati di attività OSN. Un esempio del primo caso è quello della Figura 2.2 in cui la posizione è descritta come testo di un post (quindi è nel suo contenuto), mentre un esempio del secondo è quando la posizione viene rivelata attraverso un geotag associato al post (ovvero nei metadati associati) come nella Figura 2.3. La seconda osservazione è che la divulgazione della privacy potrebbe avvenire attraverso l'esposizione esplicita o inferenze implicite. Divulgare il nome della città in un post chiamato "Città natale" è un esempio di esposizione esplicita, mentre l'inferenza implicita si verifica quando un utente ha numerosi post geotaggati

Figura 2.2: Esempio di divulgazione delle informazioni derivanti da contenuti: la posizione viene rivelata attraverso la descrizione presente nel post

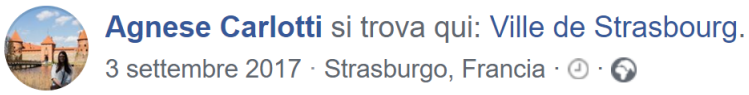

Figura 2.3: Esempio di divulgazione delle informazioni derivanti da metadati: la posizione viene rivelata attraverso il geotag associato al post

nella stessa città. In quest'ultimo caso, anche se l'utente non ha mai rivelato direttamente la città in cui vive, si può facilmente dedurre tale informazione. Tutto ciò che abbiamo appena esposto riguardo a Twitter è ovviamente riferibile anche alla maggior parte degli altri OSN che si basano sul tagging e il re-sharing come Facebook, Instagram e molti altri magari meno conosciuti come Netlog.

2.1.2 Il caso di Facebook e le app di terze parti

Gli utenti però, sui social network, non mettono a rischio la loro privacy e quella dei loro contatti solo attraverso i contenuti che pubblicano online ma anche con le autorizzazioni che concedono. Un esempio è Facebook e la sua piattaforma per app di terze parti come dimostrato dallo studio di Biczòk et Chia [10]. Questi autori si chiedono quanto sia appropriato, per un servizio OSN, permettere ad un utente di condividere informazioni riguardanti un'altra persona per ottenere un'esperienza migliore ed arrivano alla conclusione che questo può portare all'emergere di esternalities.
Come sappiamo da Mankiw [13], esse sorgono quando un'entità si impegna

in un'attività che influenza il benessere di un terzo e tuttavia non paga né riceve alcun compenso per quell'effetto. Queste esternalità possono essere positive o negative in base all'effetto che hanno su chi le riceve ma anche nel caso in cui questa condivisione di informazioni altrui porti effetti benefici (come per es. un'esperienza personalizzata) si perde comunque una parte di privacy.
Quindi l'interdependent privacy si può ritrovare nella piattaforma applicativa di Facebook per quanto riguarda la protezione della privacy dell'utente da sviluppatori di app di terze parti. Il Centro assistenza di Facebook [14] specifica che alcune app di terze parti possono richiedere informazioni prima di poter essere utilizzate. Queste vengono usate per operazioni che possono aiutare l'utente a individuare gli amici che utilizzano l'app o il gioco, per personalizzare i contenuti offerti nell'app rendendola più interessante, per semplificare la condivisione dei contenuti su Facebook e per facilitare la creazione di un account così da poter utilizzare subito l'app. Le autorizzazioni più comunemente richieste sono l'accesso al profilo pubblico (nome, immagini del profilo, nome utente, ID utente, qualsiasi informazione che l'utente scelga di rendere pubblica), l'accesso alla lista di amici, il genere, la fascia d'età e la lingua. Tutti questi permessi vengono richiesti con lo scopo di personalizzare l'esperienza sull'app. Viene evidenziato inoltre che le app non sono autorizzate a utilizzare le informazioni raccolte per gli annunci pubblicitari o a trasferirle senza il consenso dell'interessato. Le app di terze parti devono sempre chiedere il consenso dell'utente, che dovrà concedere le autorizzazioni [15]. Tra le più comuni, riguardanti la privacy personale, troviamo:

- `basic`: include tutte le informazioni di base su una persona come `id`, `first_name`, `last_name`, `middle_name`, `name`, `name_format`, `picture`, `short_name`

- `email`: concede all'app l'autorizzazione per accedere all'indirizzo e-mail principale dell'utente ma non consente l'invio di messaggi di pubblicità o spam.

- `user_age_range`: per accedere alla fascia d'età dell'utente.
- `user_birthday`: per accedere al compleanno dell'utente.
- `user_events`: per l'accesso di sola lettura agli eventi di cui un utente è organizzatore o a cui ha risposto.
- `user_gender`: per accedere al genere dell'utente.
- `user_hometown`: per accedere al luogo della città di origine dell'utente impostato nel proprio profilo.
- `user_likes`: per accedere alla lista delle Pagine Facebook a cui un utente ha messo "Mi piace".
- `user_link`: per accedere all'URL del profilo Facebook dell'utente della tua app.
- `user_location`: fornisce l'accesso alla città attuale di un utente tramite il campo location nell'oggetto utente.

Dimension	Dependency (Affecting)	# app	% app
Personal	Self	18204 [4634]	67.35 [17.15]
	Friends	518	1.92
Relational	Both self and friends	18204 [480]	67.35 [1.78]
Spatial	Self	494	1.83
	Friends	6249	23.12

Figura 2.4: Dimensioni della privacy online, dipendenza del controllo della privacy e il numero di app che presentano i rispettivi rischi. Le cifre in [parentesi] escludono app che richiedono solo la singola autorizzazione di base [10]

Biczòk et Chia nel 2012 rilevano che Facebook ha un totale di 65 permessi, classificati in 5 tipi: basic, user or friend information, extended, open graph, and page permissions. [15]
La figura 2.4 che contiene la Tabella 1 riportata in [10] mostra le tre dimensioni dei rischi della privacy online che danneggiano l'utente e i suoi amici

ed evidenzia come il controllo della privacy con le autorizzazioni app su Facebook dipenda non solo dall'utente, ma anche dalle azioni dei suoi amici. Mostra inoltre il numero e la percentuale di app che presentano un rischio per la privacy, derivate da un dataset di 27.029 app costruito da Chia et al. [16]. Questo è stato realizzato scaricando l'elenco di tutte le applicazioni di Facebook su socialbakers.com, e analizzando ciascuna app per salvare l'elenco delle autorizzazioni richieste al momento dell'installazione.

Permission	# app	% app
basic	18204	67.35
email	3766	13.93
user_about_me	284	1.05
user_activities	67	0.25
user_birthday	914	3.38
user_checkins	24	0.09
user_education_history	67	0.25
user_events	27	0.10
user_games_activity	5	0.02
user_groups	35	0.13
user_hometown	204	0.75
user_interests	94	0.35
user_likes	314	1.16
user_location	412	1.52
user_notes	12	0.04
user_online_presence	67	0.25
user_photos	574	2.12
user_questions	-	-
user_relationships	77	0.28
user_relationship_details	21	0.08
user_religion_politics	50	0.18
user_status	131	0.48
user_subscriptions	-	-
user_videos	187	0.69
user_website	12	0.04
user_work_history	107	0.40

Permission	# app	% app
friends_about_me	25	0.09
friends_activities	23	0.09
friends_birthday	162	0.60
friends_checkins	15	0.06
friends_education_history	30	0.11
friends_events	7	0.03
friends_games_activity	5	0.02
friends_groups	8	0.03
friends_hometown	44	0.16
friends_interests	33	0.12
friends_likes	51	0.19
friends_location	62	0.23
friends_notes	3	0.01
friends_online_presence	89	0.33
friends_photos	256	0.95
friends_questions	-	-
friends_relationships	19	0.07
friends_relationship_details	8	0.03
friends_religion_politics	20	0.07
friends_status	16	0.06
friends_subscriptions	-	-
friends_videos	75	0.28
friends_website	2	0.01
friends_work_history	29	0.11

Figura 2.5: Autorizzazioni di Facebook con implicazioni per la privacy personale, il controllo sta con l'utente stesso (a sinistra), o dipende dalle decisioni dei suoi amici (a destra) [10]

La perdita di privacy personale degli utenti, che comprende dati come opinioni politiche, storia dell'istruzione, data di nascita, può dipendere quindi sia dalle decisioni dell'utente, che da quelle dei suoi "amici" di installare un'app di terze parti. Facebook ha infatti diverse autorizzazioni che consentono a un'app di ottenere non solo le informazioni personali dell'utente ma

anche quelle dei suoi amici come mostrato nella Figura 2.5 [10]. L'1,92% delle app richiede infatti le informazioni personali degli amici ,come si può vedere dalle voci della tabella a destra. Sebbene si tratti di una somma inferiore rispetto al 17,15% delle app che richiedono le informazioni personali dell'utente (escluse quelle che richiedono solo il permesso `basic`), gli autori ritengono che la maggior parte degli utenti non sia consapevole di queste esternalities della privacy derivanti dalle autorizzazioni alle app terze parti di Facebook. Il tipo di privacy più a rischio è quello relazionale, che nel caso dei social network riguarda gli eventi che coinvolgono due "amici" (come tag, chat e messaggi) rispetto ai semplici collegamenti di amicizia.

Permission	# app	% app
basic	18204	67.35
read_friendlists	114	0.42
read_mailbox	1	0.00
read_requests	5	0.02
read_stream	356	1.32
rsvp_event	12	0.04
xmpp_login	14	0.05
manage_friendlists	1	0.00
manage_notifications	7	0.03

Figura 2.6: Autorizzazioni di Facebook con implicazioni sulla privacy relazionale. Il controllo dipende sia dallútente che dai suoi amici. [10]

Nella Figura 2.6 sono indicati i permessi che un'app di terze parti può richiedere per il suo funzionamento e che possono rivelare la relazione tra l'utente e i suoi amici. L'autorizzazione basic dà accesso all'elenco degli amici dell'utente, `read_friendlists` rivela gli elenchi personalizzati di amici che l'utente ha fatto (come ad esempio famiglia, colleghi, etc.), `manage_friendlists` consente addirittura all'app di modificare tali elenchi, `xmpp_login` fa accedere ai messaggi di chat privati mentre `read_stream` permette di leggere i messaggi meno privati come i post sul diario dell'utente. Se si esclude il permesso `basic`, richiesto da più della metà delle app prese in considerazione e non particolarmente pericoloso, l'1,75% di queste app pone un serio rischio

di violazione di privacy relazionale. Questa, agendo su entrambi gli utenti coinvolti, può essere protetta solo dalle azioni congiunte di tutte le parti in causa.
La tabella 3 qui riportata risale ad uno studio del 2012. Una versione più aggiornata delle possibili autorizzazioni che minano la privacy relazionale è:

- `user_friends`: concede all'app l'autorizzazione per accedere alla lista di amici che usano l'app. Perché un utente venga visualizzato nella lista di amici di un altro, entrambi devono condividere le proprie liste di amici con l'app senza aver disabilitato l'autorizzazione durante l'accesso.

- `user_photos`: fornisce l'accesso alle foto che un utente ha caricato o in cui è stato taggato.

- `user_posts`: fornisce l'accesso ai post sul diario di un utente e includono i propri, quelli in cui è stato taggato e quelli pubblicati sul suo diario da altri utenti.

- `user_tagged_places`: fornisce l'accesso ai luoghi in cui un utente è stato taggato nelle foto, nei video, negli stati e nei link.

- `user_videos`: fornisce l'accesso ai video che un utente ha caricato o in cui è stato taggato.

- `groups_access_member_info`: concede all'app l'autorizzazione per le informazioni pubbliche dei membri di un gruppo.

Anche la protezione dello spazio digitale dell'utente, la sua privacy spaziale, dipende sia dalle sue decisioni che da quelle dei suoi amici. Le autorizzazioni che permettono all'app di pubblicare a nome o sul diario dell'utente sono:

- `publish_pages`: concede all'app l'autorizzazione per pubblicare post, commenti e mettere "Mi piace" alle Pagine gestite dagli utenti che usano l'app. L'app deve avere anche `manage_pages` per pubblicare come Pagina.

- `publish_to_groups`: concede all'app l'autorizzazione per pubblicare contenuti in un gruppo per conto dell'utente che ha concesso l'autorizzazione.

- `publish_video`: concede l'autorizzazione a un'app per pubblicare video in diretta sul diario dell'utente,

Nella ricerca del 2012 erano citate inoltre le autorizzazioni `publish_actions` e `publish_streams`, che consentivano alle app di terze parti di postare sia sul diario dell'utente che su quello degli amici di quest'ultimo. Dallo studio emergeva che l'autorizzazione `publish_streams` era stata richiesta dal 23.12% delle app dello studio e che era la principale causa di post non richiesti e spesso imbarazzanti o non pertinenti sul diario dell'utente o dell'invio di messaggi osceni e spammer agli amici. Attualmente queste due autorizzazioni non figurano più tra quelle che Facebook concede agli sviluppatori.
Biczòk et Chia [10] hanno proposto un modello teorico di gioco chiamato Interdependent Privacy Game (IPG), con lo scopo di evidenziare gli effetti inter-utente delle installazioni di app su Facebook. I risultati a cui sono arrivati sono che l'esternality positiva si rafforza con il crescere del numero di utenti che installano la stessa app, ma allo stesso tempo l'esternality negativa risulta già presente quando un singolo utente decide di installare un'app, causando una perdita di privacy notevole per sé stesso e i suoi amici. Risultati simili sono stati ottenuti anche da Harkous et al. in [17] nello studio sulle app basate sui servizi di cloud che approfondiremo più avanti. Un altro fattore di rischio per la privacy di una persona sono le applicazioni, solitamente presenti sugli smartphone, che utilizzano per impostazione predefinita servizi basati sul GPS e la posizione e che la rivelano tramite l'uso degli OSN spesso senza che l'utente che posta il contenuto sia completamente consapevole dei rischi in cui può incorrere.

2.2 Interdependent privacy e co-location

Grazie all'evoluzione degli smartphone e alla rapida diffusione delle reti mobili ad alta velocità si è sviluppata una nuova cultura di condivisione dei propri contenuti sul web, in particolare di foto e video attraverso i social network. Un numero crescente di dispositivi è ormai in grado di incorporare le informazioni sulla posizione e altri metadati nel contenuto creato. Tuttavia, attualmente non vi è molta consapevolezza delle possibili conseguenze sulla privacy di tali dati. Inoltre, mentre nella maggior parte dei casi le persone caricano i propri media consapevolmente, il numero di quelli caricati da altri è invece talmente enorme che è quasi impossibile per gli utenti essere consapevoli di tutti quelli che potrebbero essere rilevanti per loro. Come già sottolineato dagli studi su Facebook [10] e Twitter [12], gli attuali servizi di social network e i siti di condivisione di contenuti offrono meccanismi di tutela per la riservatezza dei contenuti pubblicati da un utente, come ad esempio il controllo degli accessi su chi può vedere quel materiale, mentre offrono poche possibilità per la gestione delle implicazioni sulla privacy create dalle azioni degli altri. Le informazioni rivelate dalle immagini e dagli amici di un utente nei social network possono infatti essere utilizzate per dedurre informazioni private, tra cui quelle sulla posizione. Queste ultime possono poi anche essere rivelate dagli utenti mobili che si connettono a servizi basati sulla posizione da uno stesso indirizzo IP, compromettendo in questo modo la privacy di coloro che vogliono mantenere privata la loro posizione.

2.2.1 Quali sono i metadati?

Abbiamo gli accennato ai metadati parlando delle co-divulgazioni sui social network, citando in particolare l'esempio del geotag. Proprio per quanto riguarda le informazioni sulla posizione infatti questi possono causare notevoli perdite di privacy. Si ritiene quindi opportuno parlarne in modo più approfondito. I metadati memorizzati nei media, in particolare quelli delle foto, possono includere vari tipi di informazioni di contesto che isolate non so-

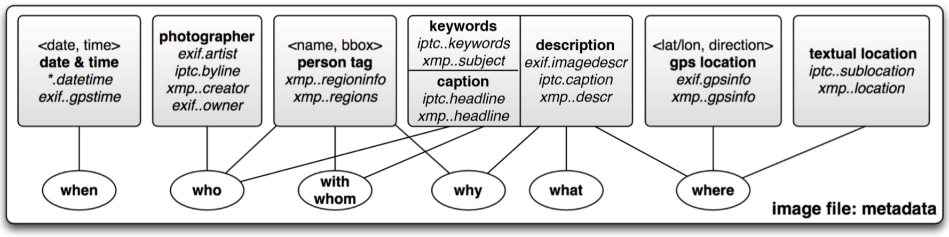

Figura 2.7: Metadati che possono essere inclusi nelle informazioni[18]

no per lo più degne di nota ma se vengono collegate ad altre possono portare a trarre conclusioni che potrebbero danneggiare la privacy personale. Si tratta di dati che rivelano informazioni come data, ora, artista, id telecamera, posizione GPS. Spesso questi metadati vengono generati automaticamente dalle fotocamere digitali e da molti smartphone nei file da loro creati. Se questo avviene, è necessario prestare molta più attenzione per proteggerli in quanto la loro estrazione, che spesso richiede tempo e specifiche conoscenze tecniche, è possibile e ha molto probabilmente un intento malevolo.

2.2.2 L'influenza dei metadati sugli altri

Come detto precedentemente la maggior parte dei dispositivi è in grado di raccogliere informazioni sulla posizione tramite GPS o tracciamento Wi-Fi. Queste informazioni possono essere utilizzate dai servizi basati sulla posizione e vengono spesso incorporate nei media creati dal dispositivo, come ad esempio nelle foto. Questi metadati possono causare serie implicazioni per la privacy del creatore dell'immagine, ma spesso si trascura la loro influenza sull'interdependent privacy. In [18], Henne et al. si concentrano sullo studio dei problemi di interdependent privacy causati dalla condivisione di media, in particolare foto, in cui sono presenti informazioni sulla posizione e altri metadati, e di come questi possano influire sugli altri. Le minacce per la privacy da loro individuate vengono divise in due categorie in base all'origine: informazioni auto-divulgate o problemi interni, causati dagli utenti stessi che caricano i media a loro riferiti con protezioni insufficienti (un classico esem-

pio è la condivisione su Facebook di immagini compromettenti sulla propria bacheca anziché l'invio delle stesse per messaggio privato ad un amico), e informazioni co-divulgate o problemi creati da altri che risultano particolarmente difficili da gestire poiché la persona danneggiata, non essendo coinvolta nel processo di caricamento, non può impedire che ciò accada né prendere alcuna precauzione efficace. Al momento infatti, oltre alla richiesta offline di rimozione del contenuto, le uniche contromisure possibili sono quelle per vie legali che spesso richiedono tempi molto lunghi e costi non indifferenti. Si deve tener conto anche della persistenza dei media e dei dati associati su internet che costituiscono un'altra preoccupazione per la privacy: mentre alcuni tipi di contenuti dopo un certo tempo vengono rimossi dal database, altri rimangono visibili a tempo indeterminato e possono quindi danneggiare la persona a cui sono riferiti anche molto tempo dopo la loro pubblicazione.

2.2.3 Co-location e offuscamento

Il metadato che forse più causa perdita di privacy è la posizione, la quale rivela l'ubicazione di un utente in un determinato momento. Una localizzazione frequente può portare alla scoperta delle abitudini di una persona rendendola un soggetto vulnerabile. Con il termine co-location si indica il fatto che due utenti si trovano nella stessa posizione in un determinato momento [19]. Le informazioni sulla co-location delle persone sono sempre più semplici da reperire online. Ad esempio, gli utenti che si interfacciano con dispositivi mobili segnalano frequentemente le loro posizioni nei messaggi e nelle immagini che pubblicano sui siti web di social network taggando i nomi degli amici con cui si trovano. Le informazioni sulla co-location possono però essere ottenute in molti altri modi diversi, come con il riconoscimento facciale automatico sulle immagini (che contiene l'ora e il luogo in cui è stata scattata la foto nei suoi metadati) oppure tramite dispositivo abilitato Bluetooth che rileva e segnala dispositivi vicini. Allo stesso modo, è probabile che gli utenti che si connettono dallo stesso indirizzo IP siano collegati al medesimo punto di accesso a Internet, fornendo così la prova della loro co-location. Secondo

la ricerca di Olteanu et al. [19] anche nel caso in cui un utente non divulghi alcuna informazione sulla propria posizione, la sua privacy può comunque diminuire fino al 21% a causa delle informazioni riportate da altri utenti. Le persone infatti perdono parzialmente il controllo sulla privacy della propria posizione in quanto le co-location e le informazioni sulla posizione individuale rivelate da altri utenti influiscono sostanzialmente su questo tipo di privacy. Gli attacchi che utilizzano sia la posizione che le informazioni sulla co-location possono essere piuttosto potenti anche se sfruttare questo tipo di dati in modo corretto può essere complesso perché richiede la considerazione congiunta delle informazioni raccolte riguardo un numero potenzialmente elevato di utenti (la posizione di un utente è correlata a quella dei suoi amici, che a sua volta sono correlate con quelle dei propri amici e così via). In combinazione con le informazioni sulla posizione, tali co-locations possono essere utilizzate per migliorare l'inferenza delle posizioni degli utenti, minacciando ulteriormente la loro privacy sulla posizione: man mano che vengono prese in considerazione le informazioni sulla co-location, non solo le posizioni segnalate dall'utente e i suoi modelli di mobilità possono essere utilizzati per localizzarlo, ma anche quelli dei suoi amici (e degli amici dei loro amici e così via). Quindi le loro posizioni non sono indipendenti ed un eventuale attacco alla posizione di un utente prende potenzialmente in considerazione quelle di tutte le persone a lui collegate. Questo è un numero limitato di persone che si restringe ad esempio ad alcuni colleghi durante l'orario di lavoro e alla famiglia e gli amici intimi durante il resto del tempo. C'è poco che un utente può fare per proteggersi da un'eventuale attacco, a parte cercare di nascondere le informazioni sulla co-location o impedirne l'inferenza. Nella pratica questo si attua nascondendo gli indirizzi IP (utilizzando un proxy, una rete VPN o una rete di anonimato peer-to-peer come Tor), disattivando il Bluetooth o sfocando i volti nelle immagini pubblicate sui social network e soprattutto generalizzando il tempo e le informazioni dell'utente nelle co-location pubblicate (per es. scrivendo "sono a Bologna con amici" invece di "Sto facendo aperitivo da @PincoPallino con @Alice @Bob). Un'altra contro-

misura efficace, ma estremamente difficile da realizzare nella pratica, sempre proposta da Olteanu et al. in [19], è il coordinamento tra gli utenti nel segnalare in modo offuscato le proprie informazioni, compresa la posizione in cui si trovano insieme. Una possibile soluzione, nel caso di co-location esplicitamente segnalate, è che un utente che pubblichi questa informazione incorpori la sua posizione offuscata in modo che tutti gli altri utenti la riportino allo stesso modo. Ciò non significa che gli utenti co-localizzati debbano segnalare le singole posizioni, ma piuttosto che, se vogliono farlo, accettano di renderle uguali. L'offuscamento consiste in un meccanismo che seleziona una posizione adiacente casuale rispetto alla posizione effettiva e su cui l'utente non ha alcun controllo. Una già citata tecnica di protezione della privacy è anche l'offuscamento attraverso la generalizzazione (ad esempio, segnalare un'area ampia in cui è contenuta la posizione effettiva dell'utente). Allo stesso modo, si consiglia che gli utenti generalizzino le informazioni sulla co-location, in un modo coarse-grained: ciò implica la generalizzazione della componente temporale di una co-location e/o la componente degli utenti co-localizzati. Generalizzare la componente orario in un'informazione di co-locazione significa riportare un intervallo di tempo anziché l'ora esatta (ad esempio, utilizzare "morning" anziché "10 a.m.") mentre generalizzare la componente utente significa escludere i nomi degli amici con cui ci si trova e riportare solo il numero di amici (ad esempio, invece di riferire di essere con Alice, un utente riferirebbe semplicemente di essere con un amico). Riportare singole co-locations offuscate in modo coordinato dovrebbe quindi dare ad un malintenzionato meno informazioni, poiché massimizza l'insieme delle possibili posizioni in cui gli utenti co-localizzati potrebbero effettivamente trovarsi. Inoltre, diventa più dispendioso il processo di individuazione delle persone co-locate con un soggetto in quanto rende necessario esplorare tutte le possibili combinazioni di utenti reali con cui è insieme e assegnare una probabilità a ciascuno di essi per ottenere informazioni utili. Questa contromisura protegge la privacy degli utenti rendendo l'inferenza computazionalmente molto costosa. L'offuscamento della componente temporale delle

co-locazioni porta anche ad un drastico aumento della complessità perché si devono considerare tutte le combinazioni di istanze temporali esatte quando gli utenti sono co-localizzati. La collaborazione potrebbe anche essere raggiunta tramite tecnologie di comunicazione ad hoc a corto raggio come Wi-Fi Direct o Bluetooth, in quanto gli utenti co-localizzati sono fisicamente vicini.

2.2.4 Altre soluzioni: l'informatica Context Aware e il rumore

Come argomentato nei paragrafi precedenti, la sola localizzazione dà già molti indizi se presa in considerazione insieme ad altre risorse disponibili, come ad esempio Google Maps, ma spesso gli utenti rendendola pubblica aggiungono anche informazioni sulle attività che stanno svolgendo. I sensori presenti nei dispositivi mobili sono poi una delle principali fonti per il rilevamento della posizione e dell'attività. In alcuni casi, gli utenti stessi forniscono volontariamente o involontariamente informazioni che connettono diverse funzionalità inviando immagini geo-taggate o mappe dettagliate che illustrano le loro prestazioni di fitness (come quanti km hanno percorso attraverso un app contapassi) e il check-in in luoghi visitati, taggando anche le altre persone con cui sono. Ad esempio, le informazioni sulla posizione sono sicuramente una delle funzionalità più utilizzate tra quelle generate dagli smartphone e sono spesso esposte pubblicamente dalle applicazioni.
Studi recenti hanno evidenziato come questi dati, oltre che essere usati individualmente, possono anche essere impiegati per prevedere il contesto di un utente, cioè prevedere il suo futuro contesto in base ai dati attuali o passati da lui condivisi.
Un contesto può essere definito come qualsiasi informazione utilizzabile per caratterizzare la situazione di un'entità [20] . Partendo dalla teoria dei processi di Markov e dalla programmazione dinamica, Bedogni e Levorato in [21], propongono una metodologia che non tenta di nascondere le informazioni per impedire la stima del contesto (come attraverso la generalizzazione o l'offuscamento di cui si parla in [19]) ma che è basata sull'approccio di iniettare

"noise" (rumore) per minimizzare l'accuratezza della previsione preservando al tempo stesso le funzionalità e l'attività degli utenti sui social network.
Un malintenzionato potrebbe infatti voler prevedere per quanto tempo un persona si troverà fuori dalla propria abitazione, e quindi vorrebbe scoprire una caratteristica che è associata a un sottoinsieme di contesti nello spazio di tutti i possibili scenari. Con il modello proposto dagli autori, l'utente costruisce un modello simile e inietta dinamicamente il rumore, inteso come aggiornamenti contestuali, per alterare il modello stimato dall'aggressore. Sempre in [21] viene sottolineata la differenza tra privacy e privacy predittiva basata sul contesto dell'utente. Mentre la prima, se violata, potrebbe essere recuperata più facilmente ad esempio cambiando la password in caso di e-mail trapelate, la seconda è più difficile da modificare, poiché spostarsi in un'altra casa o stravolgere la propria routine quotidiana è assai più complesso. Per spiegare quest'ultimo tipo, su cui si basa lo studio, vengono proposti due esempi molto significativi: *"the Prudent Burglar"* e *"the Futurist stalker"*. Il primo considera lo scenario in cui un utente malintenzionato (lo scassinatore prudente) per eseguire con successo un attacco, in questo caso un furto con scasso, deve prevedere quando la sua vittima, Alice, non si trova nella sua abitazione per un tempo sufficiente da permettergli di portare a compimento il suo piano. Supponendo che Alice abbia un abitudinario schema giornaliero che prevede una passeggiata nel parco, una visita al negozio e infine il recupero dei suoi figli all'uscita da scuola e che renda disponibile sui social network sufficienti informazioni che permettono al ladro di dedurre il suo stato attuale, se il ladro raccoglie i contesti nel tempo e costruisce un modello probabilistico di sequenze tipiche potrebbe iniziare l'attacco nel momento in cui Alice comincia ad andare verso il parco se il tempo che gli serve è minore di quello che normalmente impiegerebbe la vittima a tornare. Al contrario, se non si basasse sul modello che ha creato e considerasse solo il contesto istantaneo individuale, non avrebbe abbastanza informazioni sul tempo che gli servirebbe per derubare la casa e quindi avrebbe molte più probabilità di essere colto sul fatto. Per quanto riguarda *"The Futurist Stalker"*

viene preso in considerazione uno scenario che non è direttamente correlato alle informazioni sulla posizione. Infatti, in questo caso, lo stalker pianifica i suoi attacchi quando presume che la vittima sia sola in quanto il beneficio che ne trae aumenta all'aumentare della loro durata. Si pensi ad esempio ai casi di cyber-bulli o di pedofili di Internet, che cercano di entrare in contatto con le vittime quando sono più vulnerabili.
Per pianificare e migliorare l'attacco, lo stalker dovrebbe quindi considerare, oltre alle informazioni che ha a disposizione sulla vittima, anche quelle sulle persone con cui questa trascorre il suo tempo. Se per esempio il contesto dei genitori di Alice non è protetto, non lo sarà neanche quello della vittima. Sapendo infatti che il padre ha un lavoro full-time e riconoscendo che la madre sta guidando in un determinato intervallo di tempo come la prima mattinata (utilizzando i dati dell'accelerometro presenti nel suo dispositivo mobile), l'aggressore saprà che quel giorno la donna lavorerà dall'ufficio invece che da casa e potrà quindi dedurre che, al ritorno da scuola, Alice sarà sola in casa per un periodo di tempo maggiore. Se la madre non guida alla mattina invece, lo stalker saprebbe che al ritorno a casa di Alice da scuola un genitore sarà presente e quindi il suo attacco durerebbe meno e la sua ricompensa sarebbe quindi minore. Per evitare questi attacchi, gli autori propongono due algoritmi, uno greedy e uno conservative, per limitare la capacità degli aggressori di prevedere i futuri contesti degli utenti. Per fare ciò iniettano il rumore, che riduce la prevedibilità del contesto futuro degli utenti in base a una sequenza di osservazioni passate. I risultati che hanno ottenuto indicano che si tratta di una soluzione praticabile in grado di preservare le funzionalità sociali per cui un utente utilizza un OSN e permettere al contempo di garantirne la privacy.
L'informatica context aware offre importanti spunti di riflessione sull'interdependent privacy e su come le informazioni che condividiamo online espongano noi e le persone a noi vicine a rischi concreti anche nella vita reale. Sebbene infatti gli esempi sopra riportati richiedano competenze informatiche avanzate per essere messi in atto proficuamente, si potrebbero prevedere i futuri

contesti di una persona anche solo se questa o i suoi amici pubblicano spontaneamente le informazioni relative ai loro spostamenti sui social network. L'esempio classico è quello del teenager che, prima di partire per una vacanza in famiglia, pubblica l'intero itinerario con le date su Facebook rendondo così noto che lui e la sua famiglia non si troveranno a casa per un determinato periodo di tempo. Questo espone la sua abitazione al rischio di essere derubata facilmente se un malintenzionato è presente tra i suoi contatti.
Oltre alle persone fisiche però anche le varie applicazioni che scarichiamo e utilizziamo quotidianamente sono interessate alla raccolta di dati sulle nostre attività in modo da utilizzarli per i loro scopi, che potrebbero essere pubblicitari o anche solo di miglioramento del servizio offerto. Per esempio, le app che utilizzano i servizi di cloud richiedono l'accesso alle informazioni con la promessa di realizzare un servizio migliore esponendo però l'utente e i suoi collaboratori a perdite di privacy notevoli come dimostrato in [17].

2.3 La perdita di privacy attraverso le app che accedono al Cloud

Servizi di cloud storage (CSP) come Google Drive, Dropbox e OneDrive sono diventati sempre più popolari nel corso degli ultimi anni accumulando centinaia di milioni di utenti. Questi spazi di archiviazione personale online sono sempre più utilizzati dalle persone poiché rendono accessibili agli utenti i propri file in qualsiasi momento ed in ogni luogo utilizzando semplicemente una connessione Internet. Sollevano però innumerevoli preoccupazioni sulla privacy, in quanto svolgono anche il ruolo di piattaforme che consentono a una miriade di app di terze parti di lavorare sui dati degli utenti. Infatti, per attrarre ulteriormente le persone, i CSP nel tempo sono passati da semplici fornitori di servizi a ecosistemi di app. Ora offrono API per gli sviluppatori che permettono di importare ed elaborare i file degli utenti archiviati nel cloud. Un esempio può essere l'app PandaDoc, che consente di creare, modificare e firmare documenti online. Quando un utente utilizza PandaDoc

dal browser del suo computer portatile può importare i file memorizzati nel suo Google Drive al posto che quelli presenti nel suo disco rigido. Dropbox afferma che centinaia di migliaia di app sono state integrate con la sua piattaforma. Nell'impostazione aziendale poi, le app cloud di terze parti sono sempre più in aumento e molte imprese stanno adottando in modo efficace Dropbox Business, OneDrive for Business e Google Drive for Work. Queste app di terze parti forniscono determinate funzionalità per le quali richiedono l'accesso ai dati degli utenti, che sacrificano così parte della loro privacy per ottenerne i servizi. Come si evince da [22] però i due terzi delle app analizzate dagli autori sono troppo privilegiate, ciò significa che acquisiscono più dati di quelli realmente necessari per il proprio funzionamento. Gli utenti, per ottenere determinati servizi, alle volte finiscono quindi per esporre più dati del necessario alle varie app. Oltre a divulgare i propri dati in modo eccessivo, possono però fornire anche quelli di altri utenti, come ad esempio i propri collaboratori. Sempre in [22] è emerso che circa il 76% delle app di terze parti di Google Drive presenti nel Google Chrome Store richiede l'accesso completo ai dati degli utenti e che il 64% di queste app ha privilegi eccessivi. Quindi, ogni volta che un utente concede l'accesso ai propri documenti a una nuova app, ovvero a un nuovo fornitore di servizio, oltre ad arrecare una perdita di privacy a se stesso la infligge anche ai suoi collaboratori, vale a dire agli utenti con cui ha dei file condivisi. Quando infatti le persone concedono i permessi di accesso a queste app, non solo condividono i propri dati personali, ma anche quelli degli altri. Allo stesso modo, i collaboratori dell'utente possono installare app che espongono i file con lui condivisi a nuovi fornitori. I provider di cloud storage per loro natura sono intrinsecamente piattaforme collaborative in cui gli utenti condividono e cooperano su file condivisi la cui protezione non è solo nelle mani del singolo utente. Il fornitore di software di sicurezza cloud, Skyhigh Networks, segnala che il 37,2% dei documenti (su 23 milioni di utenti) sono condivisi con almeno un altro utente. Dallo studio di [17] è emerso poi che i collaboratori infliggono una perdita di privacy che è almeno del 39% superiore rispetto a ciò che gli utenti stessi causano (39%

in più con il 5% dei file condivisi e 523% in più con il 60%), quindi che le decisioni di adozione delle app dei collaboratori hanno un impatto significativo sulla perdita della privacy dell'utente stesso e, se gli utenti vogliono ridurre al minimo la perdita di privacy, non dovrebbero ignorarle. Gli autori propongono un nuovo metodo per ridurre questa perdita introducendo il concetto di decisioni basate sulla storia, in cui si informano gli utenti in tempo utile sui fornitori ai quali è già stato concesso precedentemente l'accesso ai propri dati riducendo così la propria perdita di privacy non installando app da nuovi fornitori quando possibile. Ogni decisione che un utente compie, fidandosi del fornitore del servizio e quindi autorizzando un'app ad accedere a determinati dati, aumenta potenzialmente la probabilità di perdere parte della sua privacy e di quella dei suoi collaboratori. Siamo di nuovo di fronte ad un problema di interdependent Privacy. A differenza di quanto detto però per le autorizzazioni nel contesto delle app di terze parti nel social network Facebook [10] in cui l'1,92% delle app richiedeva informazioni personali degli amici, il problema è molto più pronunciato nella app cloud di terze parti, in cui tutte le app che accedono ai file dell'utente, ottengono anche la parte condivisa. A causa della natura collaborativa delle app di cloud, i CSP non offrono agli utenti nessuna opzione per controllare se le app che i loro collaboratori utilizzano possono accedere a dei dati di cui sono proprietari. In [22] il metodo per ridurre il rischio collegato a questo tipo di app è stato quello di individuare tutte quelle con privilegi eccessivi per dissuadere gli utenti dall' installarle, anche se è risultato che molti utenti continuerebbero a farne uso poiché privilegiano la loro utilità nel breve termine rispetto al possibile rischio futuro di privacy. In [17] invece, per ridurre al minimo le possibili perdite di privacy, non si è tenuto conto del fatto che le app fossere troppo privilegiate, ma si è cercato di guidare gli utenti a prendere decisioni migliori per ridurre al minimo le perdite di privacy nella fase di installazione di una nuova app. Questo attraverso decisioni basate sulla storia (*History-based*) in cui si prediligono i fornitori che hanno già precedentemente ottenuto l'accesso ai dati dell'utente, sia direttamente tramite il suo consenso, sia tramite i

permessi dati dai suoi collaboratori. Sostanzialmente gli autori suggeriscono che selezionare l'app del fornitore che ha già accesso alla maggior percentuale di file dell'utente è la strategia ottimale per minimizzare la perdita di privacy. Per fare questo introducono degli indicatori di privacy nelle interfacce di permessi, che aiutano gli utenti a prendere decisioni basate sulla storia e quindi a minimizzare il numero di fornitori di servizi che hanno accesso ai loro dati. Prendendo come caso di studio Google Drive, decisione motivata dal fatto che ha uno degli ecosistemi di app terze parti più popolari, e compiendo un'indagine statistica in diversi scenari è emerso che, se sono presenti gli indicatori di privacy ideati dagli autori, questi portano gli utenti a prendere decisioni che preservano maggiormente la privacy, portandoli a scegliere più frequentemente l'app di un fornitore che già hanno autorizzato o che già possiede la maggiore percentuale di accesso ai dati. Gli scenari presi in considerazione sono 3 e si riferiscono rispettivamente al caso in cui un utente abbia già installato un'altra app di quel fornitore (*"Self-History"*), in cui decida di scaricare invece quella utilizzata da un suo collaboratore (*"Collaborator's App - Collaborator's Vendor"*) e infine quello in cui l'utente decida di installare, tra le possibili opzioni, l'app del fornitore che ha già il maggiore accesso ai suoi dati (*"Multiple Collaborators Scenario"*). Lo studio sottolinea come man mano che gli utenti e i loro collaboratori installano app di diversi provider la perdita di privacy aumenta e quindi prendere decisioni basate sulla storia, propria o altrui, come negli scenari descritti porta significativi benefici per l'interdependent privacy. Un altro dato risultante è quello quindi che indica come i collaboratori di un utente possano essere molto più dannosi per la sua privacy rispetto alle sue stesse decisioni. Si arriva allora alla conclusione che tener conto delle decisioni dei collaboratori dovrebbe essere una componente chiave dei futuri indicatori di privacy nelle app cloud di terze parti in quanto gli autori dimostrano l'impatto degli Insights basati sulla storia come tecnologie che migliorano la privacy. In particolare, in base allo studio sugli utenti, è emerso che questi sono meno propensi a ricordarsi delle precedenti decisioni in modo autonomo e per questo, al momento dell'installazione di

una nuova app, sarebbe utile rendere esplicite quali fornitori hanno già un accesso totale o parziale ai loro dati per fare in modo che questi ne tengano realmente conto. Tutte le conclusioni a cui sono arrivati, non sono ovviamente valide solo per la piattaforma Google Drive ma sono applicabili anche a tutti gli altri servizi di cloud. Come nel caso di Facebook, anche Google Drive concede agli sviluppatori di app terze parti la possibilità di richiedere autorizzazioni relative ai contenuti dell'utente. Come individuate in [23] si tratta di:

- `DRIVE`, che permette di visualizzare e gestire i file del Google Drive dell'utente
- `DRIVE_APPDATA`, che permette di visualizzare e gestire i dati di configurazione
- `DRIVE_FILE`, che concede di visualizzare e gestire i file e le cartelle di Google Drive che l'utente ha aperto o creato con l'app.
- `DRIVE_METADATA`, che permette di visualizzare e gestire i metadati dei file
- `DRIVE_METADATA_READONLY`, che consente solo di visualizzare i metadati dei file
- `DRIVE_PHOTOS_READONLY`, che permette di visualizzare le foto, i video e gli album
- `DRIVE_READONLY`, che consente di visualizzare i file
- `DRIVE_SCRIPTS`, che consente di modificare il comportamento degli script dell'utente di Google Apps Script

Concedere però alcuni di questi permessi può risultare pericoloso in quanto,come già ricordato, molte app sono troppo privilegiate e gli utenti non dovrebbero esporre i propri dati ad un numero di fornitori eccessivo in modo da tutelare la propria privacy e quella dei loro collaboratori.

Capitolo 3

Research Challenges

La sfida principale che ho riscontrato nella stesura di questa tesi è stata soprattutto la poca disponibilità di studi specifici relativi all'interdependent privacy e al fatto che questi si occupassero del fenomeno da punti di vista tra loro molto differenti (social network, app di terze parti, localizzazione). Gli articoli disponibili erano spesso indicati dai loro stessi autori come il primo lavoro di quel tipo ad essere pubblicato e quindi, essendo questa una tesi compilativa, il punto di vista su ogni sezione è limitato anche se probabilmente, a causa del fatto che il fenomeno è emerso soprattutto in anni recenti, si amplierà esponenzialmente nel breve periodo con altri studi. Inoltre in [11] e [12], ampiamente citati nella parte sui social network, e [17] i dati sono stati raccolti attraverso sondaggi online: il campione utilizzato è però limitato sia per nazionalità che fascia d'età e quindi poco rappresentativo rispetto alla reale utenza degli OSN. Sebbene questo non invalidi i risultati ottenuti, i problemi di campionamento rendono sicuramente la ricerca meno accurata. Un altro problema relativo all'utilizzo di questionari è poi la possibilità che le risposte dei partecipanti non indichino con precisione il comportamento messo effettivamente in atto in una situazione reale, ad esempio in [17] gli autori stessi si chiedono quanto l'efficacia degli indicatori di privacy da loro proposti sarebbe stata confermata se, invece di un esperimento web basato sul role-playing, gli utenti avessero autorizzato app reali ad accedere ai

loro dati. In [12], nella raccolta e analisi dei dati, ci si è anche scontrati con un limite al numero di richieste che potevano essere presentate all'API di Twitter per cui è possibile che la perdita di privacy rilevata sia in realtà sottostimata. Inoltre, le proposte per la risoluzione dei problemi di interdependent privacy come l'offuscamento della posizione [19] o l'aggiunta di rumore [21] richiedono, per poter essere utilizzate, un certo livello di conoscenze tecniche. L'utente medio raramente è consapevole non solo dei rischi che i metadati arrecano alla privacy ma addirittura della loro esistenza nei media che condivide sul web, per cui è possibile pensare che difficilmente si sforzi di contrastarli: le soluzioni proposte sono quindi poco praticabili per la maggior parte delle persone. Olteanu et al. [24] suggeriscono l'implementazione di un sistema innovativo per la condivisione consensuale di contenuti chiamato ConsenShare. Il caso di studio che prendono in considerazione è quello delle foto e la soluzione che propongono è quella di creare un meccanismo che sia in grado di riconoscere i volti delle persone presenti e di inviargli una richiesta di consenso ogni qualvolta questi vengano rilevati sul social. In attesa che l'utente interessato fornisca il consenso alla pubblicazione della sua immagine, il suo volto rimane oscurato. Un meccanismo analogo potrebbe essere creato anche per il riconoscimento del nome dell'utente nei post o nei commenti quando questo non viene taggato. Si tratta comunque di una tecnologia non ancora disponibile che ha però buone possibilità di essere implementata in tempi brevi, soprattutto visti i sempre più numerosi investimenti nei software di riconoscimento facciale creati principalmente nell'ambito della sicurezza e della lotta alla criminalità. Lo stesso Facebook comunque possiede già uno strumento simile, in quanto quando si posta una foto la piattaforma suggerisce gli amici che potrebbero essere presenti anche se con un margine di errore piuttosto alto. Si tratta quindi di una misura automatica che tutelerebbe anche le persone meno attente alla protezione della privacy, purtroppo però al momento questo tipo di tecnologia non è ancora stata perfezionata.

Capitolo 4

Conclusioni

La privacy online è un tema sempre più rilevante nella società odierna caratterizzata dalla vasta diffusione di internet. La normativa europea riguardo l'uso che i provider di servizi online possono fare dei dati che raccolgono, racchiusa principalmente nel GDPR, è ampia ed esaustiva. Vi è invece un notevole vuoto normativo riguardo la pubblicazione online di dati riguardanti altre persone da parte di soggetti privati come amici o familiari. Un utente infatti potrebbe voler mantenere privati il proprio nome, la città di residenza o altri dati sensibili che lo riguardano ma, come evidenziato in [12] nello studio su Twitter, sono spesso i suoi contatti ad infliggergli una perdita di privacy attraverso la co-divulgazione. Allo stesso modo sono i collaboratori con cui si hanno file condivisi su siti di cloud a portare alle maggiori perdite di privacy autorizzando le app di terze parti ad accedere a tutti i loro files, condivisi o meno [17]. Anche per quanto riguarda le informazioni sulla posizione, sono spesso le persone a noi vicine a rivelarle attraverso sensori presenti nei dispositivi mobili o tramite il geotag sui social network, senza contare poi tutte le informazioni condivise attraverso i metadati dei media pubblicati online [19] [21]. Tutti questi scenari sono accomunati dal fatto che le informazioni siano condivise da altre persone e che gli utenti a cui sono effettivamente riferite possano fare ben poco per arginare le loro perdite di privacy. Ad oggi infatti, salvo richieste offline di rimozione dei

contenuti come quelle analizzate in [11] o segnalazioni al OSN, non esistono altre modalità per proteggere i propri dati dall'imprudenza altrui. Le possibili soluzioni proposte nei vari studi già citati, in particolare l'offuscamento, la generalizzazione o l'aggiunta di rumore, sono infatti difficili da utilizzare per una persona che non abbia specifiche competenze tecnologiche: non tutti potrebbero, per esempio, essere in grado di nascondere il proprio indirizzo IP utilizzando un proxy, una rete VPN o una peer-to-peer per evitare di rivelare la propria posizione [19]. La ricerca riguardo possibili tecnologie che risolvano il problema dell'interdependent privacy online senza che l'utente debba prestarvi eccessiva attenzione, per esempio attraverso l'uso di un'app specifica che nasconda i metadati nei media che vengono pubblicati su un social, è sicuramente un campo di innovazione molto interessante che conoscerà nei prossimi anni una notevole espansione.

Bibliografia

[1] B. Schneier. *The eternal value of privacy.* Wired, May 2006

[2] D'Acquisto G., Naldi M., *Big Data e Privacy by Design- Anonimizzazione Pesudonimizzazione Sicurezza*, G.Giappichelli Editore (2017)

[3] Gazzetta ufficiale dell'Unione europea REGOLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 27 aprile 2016, L 119/1

[4] infoGDPR, sito web: www.infogdpr.eu/pseudonimizzazione-gdpr-58.html (ultimo accesso: Febbraio 2019)

[5] Rodotà S., voce *"Privacy"*, in Enciclopedia Italiana Treccani, sito web: http://www.treccani.it/enciclopedia/privacy (Ultimo accesso: Febbraio 2019)

[6] L. Brandeis, S. Warren, *The Right to Privacy*, in Harvard Law Review, Volume 4, Articolo 5, 1890

[7] Rodotà S., voce *"Diritto alla riservatezza"*, in Enciclopedia Italiana Treccani, sito web: http://www.treccani.it/enciclopedia/diritto-alla-riservatezza/ (Ultimo accesso: Febbraio 2019)

[8] Garante per la protezione dei dati personali, sito web: https://www.garanteprivacy.it/web/guest/home/autorita (Ultimo accesso: Febbraio 2019)

[9] Roger Clarke's Web-Site, sito web: http://www.rogerclarke.com/DV/Intro.html (ultimo accesso: Febbraio 2019)

[10] Biczók G., Chia P.H. (2013) *Interdependent Privacy: Let Me Share Your Data*. In: Sadeghi AR. (eds) Financial Cryptography and Data Security. FC 2013. Lecture Notes in Computer Science, vol 7859. Springer, Berlin, Heidelberg

[11] Tharntip Chutikulrungsee, Oliver Burmeister, Yeslam Al-Saggaf, Maumita Bhattacharya. *Denial of Choice: Group Level Disclosure of Private Information*. David Kreps; Gordon Fletcher; Marie Griffiths. 12th IFIP International Conference on Human Choice and Computers (HCC), Sep 2016, Salford, United Kingdom. IFIP Advances in Information and Communication Technology, AICT-474, pp.229-240, 2016, Technology and Intimacy: Choice or Coercion.

[12] Alsarkal, Y., Zhang, N., & Xu, H. (2018). *Your Privacy Is Your Friend's Privacy: Examining Interdependent Information Disclosure on Online Social Networks*. In Proceedings of the 51st Hawaii Intl Conf on System Sciences.

[13] Mankiw, N.: *Principles of Economics*. Available Titles CourseMate Series, vol. 1. South-Western Cengage Learning (2008)

[14] Centro assistenza di Facebook, https://www.facebook.com/help/1727608884153160/ (ultimo accesso: Febbraio 2019)

[15] Riferimento per le autorizzazioni di Facebook Login, https://developers.facebook.com/docs/facebook-login/permissions#reference-user_friends (ultimo accesso: Febbraio 2019)

[16] Chia, P.H., Yamamoto, Y., Asokan, N.: *Is this app safe? A large scale study on application permissions and risk signals.* In: Proceedings of the 21st International Conference on World Wide Web, WWW 2012. ACM, New York (2012)

[17] H. Harkous , K. Aberer, *"If You Can't Beat them, Join them": A Usability Approach to Interdependent Privacy in Cloud Apps*, Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy, March 22-24, 2017, Scottsdale, Arizona, USA

[18] Henne, B., Szongott, C., and Smith, M. *SnapMe if you can: privacy threats of other peoples' geo-tagged media and what we can do about it.* In Proceedings of the sixth ACM conference on Security and privacy in wireless and mobile networks (New York, NY, USA, 2013), WiSec '13, ACM, pp.95–106

[19] A. M. Olteanu, K. Huguenin, R. Shokri, M. Humbert, J.-P. Hubaux, *Quantifying interdependent privacy risks with location data*, IEEE Transactions on Mobile Computing, 2016.

[20] Anind K. Dey. *Understanding and using context.* Personal Ubiquitous Comput., 5(1):4–7, January 2001

[21] L. Bedogni, M. Levorato, *Rising User Privacy Against Predictive Context Awareness through Adversarial Information Injection*

[22] H. Harkous, R. Rahman, B. Karlas, and K. Aberer. *The curious case of the PDF converter that likes Mozart: Dissecting and mitigating the privacy risk of personal cloud apps.* Proceedings on Privacy Enhancing Technologies, 2016(4):123–143, 2016.

[23] Drive API, https://developers.google.com/resources/api-libraries/documentation/drive/v3/php/latest/class-Google_Service_Drive.html (Ultimo accesso: Febbraio 2019)

[24] Alexandra-Mihaela Olteanu, Kévin Huguenin, Italo Dacosta, Jean-Pierre Hubaux. *Consensual and Privacy-Preserving Sharing of Multi-Subject and Interdependent Data.* 25th Network and Distributed System Security Symposium (NDSS), Feb 2018, San Diego, CA, United States. 2018, Proceedings of the 25th Network and Distributed System Security Symposium (NDSS)