Estudi de la privacitat de dades (Privacy-preserving data publishing)

Enric Alibech Romero
Universitat Autònoma de Barcelona,
Barcelona, España
enricalibech@gmail.com

***Resum*—A causa de la imminent digitalització de la informació personal que s'acumula en els fitxers dels estats, empreses i xarxes socials, la preservació de la privacitat està sent un dels trencaclosques més difícils d'afrontar i mitigar en aquests últims anys, tant pels responsables de la seguretat de les dades així com pels mateixos usuaris d'Internet. En aquest paper es presenta l'estat de l'art actual d'alguns dels algoritmes que busquen protegir la privacitat de les dades, analitzant les possibles amenaces i riscos implicats. Així com una descripció de les dues línies de treball més estudiades actualment, com són el Graph-Modification i el Differential Privacy, on es realitzarà una explicació de cadascuna de les tècniques emprades, fent especial referencia a la modificació en les arestes i els vèrtexs, random perturbation i k-anonimitat. Finalment i emprant un dataset d'una xarxa d'interconnexions com un graf, es mostrarà la comparativa d'un conjunt d'indicadors que avaluen la pèrdua d'informació que es produeix a l'hora d'anonimitzar les dades segons l'algoritme d'anonimització escollit, a partir d'un conjunt de proves empíriques realitzades sobre el dataset original.**

***Index Terms*—** *(Privacitat, Graph-Modification, Differential Privacy, k-anonimitat, mineria de dades, anonimització)*

***Abstract*— Due to the impending digitalization of personal information stored in states and companies files and social networks, the preservation of privacy is being one of the most difficult puzzles to mitigate and cope with in recent years, by those who are responsible for the security of such data security as well as Internet users themselves. This paper presents the current state of art of some of the algorithms which target is to protect the privacy of the data, by analyzing the potential threats and risks involved. Also a description of the two lines of work studied nowadays, such as Graph-Modification and Differential Privacy, where there will be an explanation of each of the techniques used, with a particular reference to the modifications in the edges and vertices, random perturbation and k-anonymity. Finally, from a set of empirical tests performed on the dataset of a network of interconnections used as a graph, will be shown the comparison of a set of indicators that evaluate the information loss produced as a result of the anonymization process chosen.**

***Index Terms*—** *(Privacy, Graph-Modification, Differential Privacy, k-anonimity, data minig, anonimity)*

1. Introducció

La ràpida i constant evolució de la tecnologia web, les xarxes socials i l'aparició d'un nou paradigma en les relacions personals ha produït a la vegada un greu problema sobre com tractar i salvaguardar totes aquestes dades emmagatzemades a la xarxa, la privacitat dels usuaris i sobretot de fer front a l'aparició de noves amenaces relacionades amb l'extracció il·legal de dades personals, suplantació d'identitat i possibles fraus.

La recerca de diferents mètodes i tècniques d'anonimització o preservació de la privacitat, dissenyades amb l'objectiu de protegir l'anonimat d'aquestes dades personals dels usuaris, és un camp molt actiu en la recerca actual [1]. Aquest serà l'objectiu principal que s'estudiarà en aquest projecte, on es realitzarà un extens anàlisi de les principals propostes actuals en referència a la preservació de l'anonimat i l'ús públic de les dades.

2. Abast i Objectius del treball

Davant de la situació actual on es recull i s'emmagatzema grans quantitats de dades de xarxes socials i navegació a Internet, els investigadors durant el pas dels anys han exposat diferents treballs sobre com protegir l'anonimat de les dades en diferents escenaris [21] [22].

La solució que es proposa en aquest projecte és un detallat anàlisi sobre l'estat de l'art de les diferents tècniques que s'estan proposant actualment i que tenen com a objectiu prioritari preservar la privacitat de les dades a la xarxa, mantenint l'anonimat d'aquestes, i alhora poder fer ús de les mateixes amb eines com mineria de dades o *data analitycs*.

És per això, que es realitzarà una comparació entre les dues principals línies de treball que hi ha actualment i on s'analitzaran les bases de les dues propostes, així com l'efectivitat real dels seus algorismes avui en dia.

Les propostes en les quals farem èmfasi seran:

- *Graph-Modification* [16].
- *Differential Privacy* [17].

Els objectius són:

- Estudiar els principals mètodes d'anonimització de dades en format de graf, especialment els

mètodes basats en la modificació de l'estructura (*Graph-Modification*) i els basats en la privacitat diferencial (*Differential Privacy*).
- Analitzar les bases i solucions proposades pels dos mètodes prèviament citats.
- Comparar els algorismes proposats per cadascun dels mètodes estudiats per tal de donar una visió de contrast entre ells.
- Extreure conclusions sobre l'efectivitat i la viabilitat d'ambdós mètodes.

3. Metodologia

S'han establert una sèrie de sessions de seguiment entre alumne i professor que tenen com a objectiu una fluida comunicació sobre el progrés i els dubtes que aniran sorgint al llarg del projecte de final de grau, el qual té una durada d'un semestre.

S'ha dividit el projecte en tres etapes:

1. Durant la primera etapa del treball es realitza una recerca general sobre els estudis i treballs realitzats pels investigadors, els quals han estat publicats en congressos i revistes científiques sobre projectes relacionats amb la privacitat de les dades i del seu posterior tractament i ús. Després d'un primer contacte i entrevista amb el tutor, es van escollir quines serien les dues principals metodologies a analitzar.

D'altra banda, en aquest projecte també es recullen les principals preocupacions sobre les conseqüències d'un mal tractament a l'hora de protegir dades personals sensibles i dels possibles riscos així com les a actuals amenaces sobre la informació sensible d'usuaris d'Internet.

2. La segona etapa de treball s'aprofundirà a la recerca dels estudis realitzats sobre les tècniques de modificació per grafs i de la privacitat diferencial, on es realitzarà la descripció de l'estat de l'art dels dos algoritmes.

3. La darrera etapa es realitzarà la comprovació del comportament dels dos algoritmes escollits a partir d'una sèrie de proves empíriques sobre les diferents variants dels mètodes proposats, amb l'objectiu de poder realitzar una valoració més precisa a l'hora de comparar-los.

4. Planificació

La planificació del projecte es va organitzar en diferents fases, les quals es descriuran a continuació. La taula 1 presenta les dates compreses entre fases i els dies emprats per a cadascuna d'elles:

Fase	Inici	Fi	Dedicació
Fase 1	3/9/17	16/9/17	13 dies
Fase 2	16/9/17	2/10/17	16 dies
Fase 3	2/10/17	6/11/17	35 dies
Fase 4	6/11/17	18/12/17	42 dies
Fase 5	18/12/17	2/1/18	15 dies

Fase 1: Es realitza una entrevista inicial amb el tutor i primer contacte amb el projecte: Es realitza la primera recerca i estudi del material proposat pel tutor. A l'entrevista es dissenyen les bases inicials del projecte i quin serà el camí a seguir.

Fase 2: Recerca d'informació/estudis realitzats, per al posterior anàlisi detallat sobre l'estudi de *Graph-Modification*. Durant la recerca d'informació s'ha anat realitzant a la vegada un breu resum sobre els estudis llegits per tal de poder agilitzar l'avaluació de l'algoritme i comparativa amb l'altre mètode. Aquesta recerca va ser en gran part agilitzada degut a la col·laboració del tutor Jordi Casas i la seva tesi.

Fase 3: Recerca d'informació/estudis realitzats, per al posterior anàlisi detallat sobre l'estudi de *Differential Privacy*. Durant la recerca d'informació s'ha anat realitzant a la vegada un breu resum sobre els estudis llegits per tal de poder agilitzar l'avaluació i comparativa amb l'altre algoritme d'anonimització. A diferència del *Graph-Modification* hi ha a la xarxa molta més recerca sobre aquest estudi basant-me en gran part en els estudis realitzats per Cynthia Dwork.

Fase 4: Es realitza un estudi comparatiu i d'anàlisi de l'estat de l'art actual dels dos teoremes, durant aquest període es realitza una reunió de seguiment amb el tutor per poder fer un intercanvi d'opinions i aclarir els aspectes on havien sorgit alguns dubtes.

D'altra banda, a l'hora de comparar els dos algoritmes es va identificar que el tractament de les dades i els requeriments dels *datasets* que s'havien de fer, en cadascun dels algoritmes era diferent. Mentre que el *Graph-Modification* estava dirigit al treballar amb data set orientat a grafs i unes modificacions basades en teoria de grafs, el *Differential Privacy* actua d'una manera diferent i l'algoritme originàriament no treballa amb grafs, complicant la possibilitat de fer una correcta comparativa entre ells. És per això, que s'han buscat treballs d'investigadors que estudien com extrapolar els comportaments i aplicacions de l'algoritme de *Differential Privacy* a estudis amb Grafs amb la intenció de poder realitzar una comparació més acurada.

Fase 5: Preparació dels scripts sobre els algoritmes que estem treballant, per tal de poder realitzar les proves empíriques sobre les metodologies escollides i poder realitzar una comparativa de resultats.

5. Definició del Problema

Actualment hi ha una gran quantitat de dades personals que estan sent extretes, emmagatzemades i tractades d'Internet, a partir de les xarxes socials o qualsevol altre tipus de fonts de dades d'informació. Es planteja un escenari bàsic (Figura 1) de tractament de dades on es mostra un exemple d'un hospital que recull dades dels pacients i comparteix els informes de salut dels pacients amb un altre centre. L'hospital és el titular dels drets de les dades, mentre que els pacients són els propietaris de les seves pròpies dades i el centre extern és el receptor de les dades. Les tasques de mineria de dades que pot realitzar aquest centre extern sobre les dades personals poden ser de qualsevol tipus,

des de un recompte del nombre de pacients que tenen algun tipus de malaltia fins a un anàlisis amb més detall.

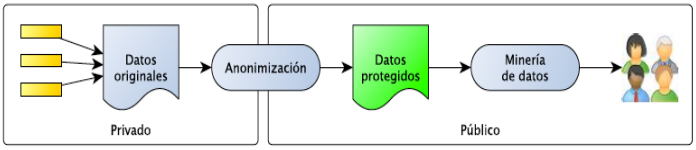

Figura 1. Escenari bàsic. Imatge extreta de [21].

Ens trobem en una situació on es treballa amb la creació de diferents tècniques d'anonimització com podria ser, l'eliminació i/o ocultació dels detalls sobre les dades que poden arribar a identificar un element de la base de dades com a únic. Però tot i així no es pot descuidar que, aquestes bases de dades podrien ser objecte d'un possible atacat dirigit per part d'un adversari, per tal de poder conèixer aquestes dades que estan sent protegides i re-identificar a un element.

L'estratègia habitual emprada per protegir la informació sensible d'aquests atacs és a partir de l'eliminació dels identificadors claus o l'intercanvi d'aquests identificadors amb uns de caràcter més genèrics, per així trencar la possible associació de les dades reals amb les dades introduïdes a la base de dades. Aquest procediment és conegut amb el nom de *Naïve anonymitzation*.

5.1. Mètodes d'emmascarament.

S'han classificat en tres, els diferents procediments d'emmascarament que es descriuran a continuació. Tots ells tenen com a objectiu impedir la identificació d'un individu en una base de dades protegida. Aquesta classificació depèn del tractament que es fa amb les dades [12, 21]:

1. Mètodes perturbatius: El conjunt de dades original és perturbat d'alguna mena, mitjançant soroll, i com a conseqüència el nou conjunt de dades anònim pot contenir informació errònia.
2. Mètodes no perturbatius: A través de la substitució del valor original per un altre valor no tan important més genèric, però que tot i així no és incorrecte. Aquest mètode utilitza les següents tècniques:

    2.1. Generalització i recodificació: S'aplica sobre els atributs categòrics i la protecció de les dades s'aconsegueix gràcies a la combinació de diferents categories amb una categoria més general.

    2.2. Codificació *top* i *bottom*: Substitució dels valors més importants i menys importants del *dataset* original per un valor categòric més general. Aquestes tècniques s'utilitzen en els casos on només poques dades tenen valors extrems, és a dir, tant d'importància molt alta com molt baixa.

    2.3. Eliminació local: Eliminació d'alguns dels valors del *dataset* original per una etiqueta especial que indica que aquest valor concret ha estat eliminat.

3. Generadors de dades sintètics: Es crea noves dades artificials, les quals substituiran les dades originals, l'objectiu d'aquest mètode és que en el cas d'un adversari aconseguint dades del *dataset* anonimitzat no comprometi la privacitat d'aquestes, ja que les dades extretes per l'atacant no seran dades que es puguin associar a cap element de la base de dades.

La solució proposada i estudiada davant d'aquest tipus d'atacs, o almenys per minimitzar la probabilitat de què l'atacant sigui capaç de fer-ho en aquest projecte, és mitjançant la utilització de tècniques que introdueixin soroll a la mostra inicial de la base de dades, per tal de fer improbable la re-identificació en el cas que es produeixi un atac, tal com fan els mètodes perturbatius. El detall de les tècniques utilitzades en aquests mètodes es troba a les Seccions 6 i 7.

El resultat en tots els procediments d'emmascarament prèviament comentats és que bàsicament quan es construeix un nou graf a partir d'aquestes modificacions es produeix una disminució de la utilitat respecte del graf original. Com a conseqüència, la falta de coherència de les dades extretes entre el *dataset* inicial respecte al anonimitzat produeix que els anàlisis posteriors sobre les dades obtingudes del graf siguin més ineficients i s'haurà d'avaluar si la informació resultant tot i estar protegida segueix sent valida.

D'altra banda, a l'hora de definir quin és el principal problema de com s'hauria de protegir la privacitat de les dades dels usuaris, la literatura especialitzada coincideix en establir una sèrie de passos [7][8]:

- Primer, identificar quin és el valor a protegir, en el cas concret de la privacitat, quina informació sensible dels usuaris han de ser protegits.
- Segon, determinar quin coneixement podria utilitzar l'adversari, per tal de realitzar un atac.
- Tercer, especificar l'ús de la informació publicada, per així poder escollir correctament quin mètode o tractament d'anonimització és l'adient per a cada cas concret.

5.2. Riscos associats.

A partir de la recerca realitzada en els articles [24][25] s'han identificat dues categories de riscos que afecten la informació addicional o complementària que implica els individus que es troben dintre de les bases de dades i on un atacant podria treure profit per tal d'extreure informació sensible:

- Que es conegui la identitat d'un individu associat a un vèrtex.
- Que es conegui l'atribut, és a dir, es rebel·la informació sensible sobre els atributs del vèrtex. O que es coneguin els veïns, com per exemple, es

produiria en el cas que es rebel·li informació sobre les connexions entre dos individus [6].

Basant-se en els articles [26][11] on es tracten els atributs categoritzats en una base de dades s'han de considerar:

- Identificadors: conjunt d'atributs que permeten identificar de forma inequívoca a un individu.
- Quasi-Identificadors: conjunt d'atributs que combinats amb informació addicional o complementaria, un atacant podria conèixer i re-identificar a un usuari de la base de dades.
- Atributs Confidencials: contenen informació sensible sobre un usuari, però que no entren en cap dels apartats anteriors.

Finalment, es classifiquen dos possibles tipus d'atacs sobre les xarxes socials: els atacs actius es produeixen quan l'atacant tracta de comprometre la privacitat de la víctima mitjançant la creació de nous usuaris i relacions entre ells abans de que es produeixi algun mètode d'anonimització. Per tal que un cop es realitzi l'anonimització de la xarxa social, aquesta també contingui els usuaris creats, l'atacant tindrà informació privilegiada sobre la nova versió. D'altra banda, els atacs passius, són aquells que es produeixen després que la base de dades original hagi passat per algun mètode de anonimització i serà llavors quan l'atacant tracti d'obtenir informació sensible.

5.3. Desemmascarant la identitat

Existeixen dos riscos relacionats amb el desemmascarament de la identitat d'un element de la base de dades:

1. Re-identificació: Risc definit com a una estimació del nombre de re-identificadors que podria extreure un adversari en cas d'un atac a la base de dades.
2. Unicitat dels elements: El risc de desemmascarar la unicitat d'un element es mesurat com la probabilitat de l'aparició d'un conjunt de combinacions úniques entre atributs del *dataset* anonimitzat i que aparegui exactament igual en el *dataset* original.

5.4. Pèrdua de informació.

La pèrdua de informació [13] és una mesura que serveix per poder quantificar quan ha estat distorsionat el graf final respecte al graf original. A l'apartat 8, on s'exposa l'estudi empíric realitzat es detallarà cada un dels indicadors que es tenen en compte per aquest projecte, els quals pertanyen al grup anomenat *Generic information loss mesaures*, extret del article[18], aquest conjunt d'indicadors et permeten poder quantificar la perduda d'informació respecte al *dataset* originat un cop s'han realitzat les tècniques d'anonimització que s'ha escollit a l'hora de realitzar aquest estudi.

5.5. Notació

Tenint $G = (V, E)$ com a graf simple i no dirigit, on V és un conjunt de nodes i E un conjunt d'arestes dins del graf G. Es defineix $\{i, j\}$ per fer referencia a la aresta que va des del node i cap el node j. És defineix $n = |V|$ per a fer referencia al nombre de vèrtexs i $m = |E|$ pel nombre d'arestes. Finalment, es designa $G = (V, E)$ per a referir-se al graf original i $\tilde{G} = (\tilde{V}, \tilde{E})$ per a fer referencia al graf anonimitzat.

6. GRAPH-MODIFICATION

Les tècniques de Graph-Modification[16] apliquen un emmascarament de tipus pertorbatiu, el qual, tal com s'ha explicat prèviament consisteix a aplicar soroll sobre el conjunt de dades que conformen el graf inicial de manera que el graf resultant no pugui ser identificat per part d'un adversari.

Els mètodes que es presentaran a continuació, primer realitzen una sèrie de modificacions o transformacions i un cop modificat exposen les dades:

1. Modificacions en les arestes i/o dels vèrtexs, aquest procediment consisteix en la modificació a través d'afegir o treure dades identificatives abans de fer-ho públic.
2. *Uncertain Graphs*, consisteix a afegir o eliminar arestes de forma parcial sobre el graf inicial, a partir de l'assignació d'una probabilitat sobre cada eix en una xarxa anònima.
3. Generalització, en aquest cas s'agrupen les arestes i els vèrtexs en uns grups anomenats super-vèrtexs o super-arestes, els quals tindran la informació sensible de cada vèrtex o aresta amagada, però el graf haurà canviat tant després d'anonimitzar-lo que les dades finals deixarien de tenir un valor real.

Tot i l'existència dels diferents mètodes prèviament mencionats, aquest projecte s'ha centrat en l'estudi de la primera tècnica, Modificacions en les arestes i/o dels vèrtexs, donat que serà el que s'utilitzarà més endavant a la comparació de tècniques d'emmascarament de la Secció 8. Estudi Empíric d'aquest mateix paper.

6.1. Modificació en les arestes i els vèrtexs.

Els mètodes que es descriuran a continuació poden ser combinats entre si per tal de crear combinacions més complexes i fer més difícil la seva re-identificació[8][9]:

- Afegir/Eliminar aresta: Consisteix a eliminar un eix existent *{vi, vj} pertanyen E* i afegir un nou *{vk, vp} no pertanyen E*. Amb la Figura 2(a) podem veure la operació en qüestió.
- Rotació d'arestes: Es produeix quan tenim tres nodes vi, *vj, vk pertanyen V*, els quals *{vi, vj}* $\in$ *E* i *{vi, vp}* $!\in$ *E*.

S'elimina l'eix {vi, vj} i es crea un nou eix *{vj, vp}* tal i com es pot observar a la Figura 2(b).

- Canvi d'arestes: Quan es disposa de 4 nodes *vj, vi, vk, vp que pertanyen a V*, on *{vi, vj}, {vk,vp}* ϵ *E* i *{vi, vp}, {vk,vj}* !ϵ *E*. Es defineix que es borra *{vi, vj}* i *{vk, vp}* i s'afegeixen dues noves arestes: *{vi, vp}, {vk,vj}* com podem veure a la figura 2(c).

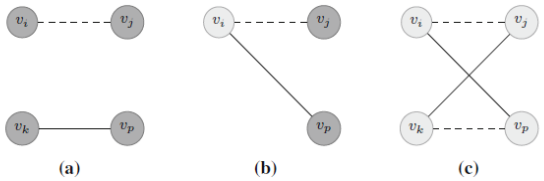

Figura 2. Imatge extreta de [16]. Operacions del algoritme *Graph-Modification*. a)afegir/eliminar aresta, b)Rotació d'arestes, c)Canvi d'arestes.

Per aquests tres tipus de modificacions, el nombre d'arestes i nodes es manté igual, però el grau de distribució d'arestes canvia en els casos de: Rotació d'arestes i Afegir/Eliminar arestes, mentre que en el cas de Intercanvi d'arestes quedaria igual.

Els mètodes prèviament descrits d'afegir o treure nodes del graf original poden ser realitzats amb la intenció de colpir un objectiu concret, en aquest tipus de cas la tècnica es coneix com a *constrained pertubation*. D'altra banda, existeix el mètode *random pertubation*, el qual consisteix en la modificació d'arestes mitjançant tècniques de modificacions aleatòries.

6.2. Random perturbation

El mètode que es descriurà a continuació consisteix a afegir soroll a les dades del graf inicial de forma aleatòria, per tal de minimitzar de manera probabilística l'èxit sobre els atacs que tenen com a objectiu conèixer la identitat del node inicial o re-identificació.

Les dues estratègies presentades[1][5] en l'algoritme de *random perturbation* són les següents:

- Afegir/Eliminar arestes de forma aleatòria: Com hem vist a la secció 6.1, les tècniques d'anonimització consisteixen a afegir o eliminar arestes del graf original amb un algoritme aleatori. S'ha de tenir en compte que aquesta estratègia preserva el mateix nombre d'arestes que el graf inicial.
- Canvi d'arestes de forma aleatòria: Canviar de forma aleatòria les parelles d'arestes sobre els nodes inicials. Aquesta estratègia preserva el nombre d'arestes i el grau dels vèrtexs.

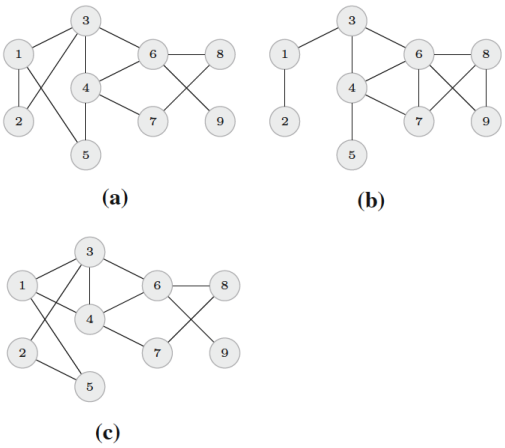

Figura 3. Imatge extreta de [16]. *Random perturbation*. a) Graf original (b) i (c) representen possibles outputs del graf original un cop s'han aplicat les dues estratègies de Afegir/eliminar arestes i canviar les arestes de forma aleatòria.

Com a exemple d'una estratègia de *random perturbation* tenim la figura 3. extreta de [16] En el cas (a) representa el graf original *G* abans de ser modificat, d'altra banda tenim el cas (b) aplicant el Afegir/eliminar arestes de forma aleatòria on s'han eliminat les arestes{1,5} i {2,3}, mentre que s'han creat les arestes {6,7} i {8,9}. Finalment en el cas c) s'aplica la tècnica de canvi d'arestes de forma aleatòria i obtenim que les arestes {1,2} i {4,5} s'han intercanviat amb {1,4} i {2,5} respectivament. Destacar que les dues tècniques conserven el nombre d'arestes i vèrtexs inicials.

6.3. k-Anonymity.

El model de k-anonymity[10][13] no és un procediment com els que hem vist anteriorment, sinó més aviat una condició que haurà de ser satisfeta pels *datasets*.

Té com a objectiu la preservació de la privacitat en conjunts de dades relacionals o estructurades. Aquest tipus de model utilitza els mètodes de protecció que s'han anat nombrant en els apartats 6.1 i 6.2 amb la finalitat d'aconseguir les restriccions o condicions necessàries per complir amb el model de k-anonimitat, el qual especifica que:[15] "Un conjunt de dades compleix el model de la k-anonimitat si, i només si, per a qualsevol combinació d'atributs casi-identificadors existeixen k o més registres que comparteixen els mateixos valors". És a dir, cada registre d'un conjunt de dades k-anonim és indistingible de com a mínim altre k-1 registre respecte al conjunt de casi identificadors. Això implica que un atacant no podrà re-identificar a un node individual amb una probabilitat més alta que: $\frac{1}{k}$.

Aquesta màxima en el que es basa k-anonimitat no presenta possibilitat de descobriment dels atributs o identificadors dels valors dels *datasets* que compleixen el model, tot i així el k-anonimitat es presenta com a un problema NP-difícil.

7. DIFFERENTIAL PRIVACY

El mètode *Differential Privacy* elaborat per Cyntia Dwork, va ser motivat per la màxima dita feta per Dalenius[6] on afirmava que: "Cap informació sensible sobre un individu pot ser coneguda sense haver abans entrat a una base de dades on estigui emmagatzemada aquella informació". A partir d'aquí, el treball d'aquesta autora s'ha enfocat en la recerca sobre l'existencia de la possibilitat d'arribar a publicar les dades sensibles d'un nombre persones, que sigui alhora representatiu de la societat actual i finalment, sigui preservada la privacitat dels individus que es troben dintre de la base de dades[23], és a dir, tenim una base de dades formada per un conjunt de nodes interconnectats entre ells i volem publicar propietats d'aquesta xarxa de nodes sense que el que es publiqui pugui produir un risc sobre la informació individual d'aquests nodes o les arestes d'aquests.

Per tal de donar resposta a aquest paradigma, Cyntia dona un exemple conegut com a "Terry Gross Example[17], on se suposa que: l'alçada d'una persona és considerada una dada molt sensible dintre d'una base de dades i que com a conseqüència que es reveli informació al respecte es considera una violació sobre la privacitat de les dades de l'usuari. D'altra banda, tenim que a la base de dades s'emmagatzema l'alçada mitjana de dones de diferents nacionalitats. Un atacant, el qual té accés a la base de dades i que a la vegada coneix informació addicional, que especifica que Terry Gross és dues polzades més baix que la mitjana de les dones lituanes, pot arribar a conèixer l'alçada d'en Terry Gross, però això és degut a que disposa d'accés a informació privilegiada, com és la mitjana de l'alçada de les dones, mentre que en canvi, si un atacant no té accés a la base de dades només podrà conèixer que en Terry Gross és dues polzades més baix que la mitja de dones lituanes."

Amb aquest exemple el que tracta d'explicar és que simplement amb la informació resultant de la base de dades, no s'aconsegueix aprendre nova informació que relacioni a algun individu en concret, en aquest cas Terry Gross, és a partir de la informació complementaria obtinguda des de fora de la base de dades que podem arribar a conèixer l'alçada d'en Terry Gross, ja que la seva presencia o no a la base de dades no modifica el resultat final de conèixer aquesta informació, és a dir, estigues a la base de dades com si no hi fos, haguéssim conegut l'alçada d'en Terry Gross amb la mateixa probabilitat, ja que hem conegut la seva alçada a partir d'informació aliena a la base de dades.

Definició de *Diferential Privacy*: *Per a tots els parells de Datasets X i Y que només difereixen en la informació d'un node/persona, la probabilitat que el resultat sigui S quan el Dataset és X és gairebé la mateixa que si fos amb el Dataset Y. En el cas d'intercanviar la X, per la Y el resultat seria el mateix.*

Formula 1:

$$\Pr[M(X) \in S] \leq e^{\varepsilon} \times \Pr[M(Y) \in S]$$

En el cas de la Formula 1. La funció M és qui aplica l'algoritme aleatori sobre els *datasets* X i Y i es considera que el *privacy loss/budget* (ε) és 0 (determina com de precís serà el resultat, és a dir, el grau de perturbació aplicat al *dataset* original).

8. ESTUDI EMPÍRIC

Per tal de poder mesurar, avaluar i comprar els algoritmes que tenen com a objectiu protegir les dades sensibles d'un *dataset*, s'han establert els mètodes d'anonimització proposats anteriorment a les seccions 6 i 7 com els encarregats de protegir els nodes del nostre *dataset* original.

S'ha treballat utilitzant l'indicador de pèrdua de informació, per tal de poder quantificar el soroll introduït i els resultats obtinguts dels diferents algoritmes. La pèrdua de informació definida com: "La diferència dels resultats entre el graf original i els grafs resultant un cop s'han aplicat les tècniques d'anonimització del graf inicial."

Com a indicadors per mesurar aquest el fenomen de pèrdua de informació en grafs s'han emprat:

- Distància Mitja (DM) [3]: Valor mitja de les distàncies que existeixen entre cada parell de nodes del graf. Es pot observar a l'equació 1:

$$DM\ (G) = \frac{\sum_{i,j} d_{i,j}}{\frac{n}{2}} \qquad (1)$$

- Diàmetre (D) [3]: la distancia entre dos vèrtexs és el menor nombre d'arestes d'un recorregut entre ells. Es pot observar a l'equació 2:

$$D\ (G) = \max\left(d_{i,j}\right), \forall\ i\ \neq j \qquad (2)$$

- Densitat del Graf (DG) [3]: El nombre d'arestes és pròxim al nombre d'arestes màxim que pot tindre el graf.

$$DG = \frac{2|E|}{|E|(|E|-1)} \qquad (3)$$

- Longitud Mitjana(LM) [3]: Càlcul de la longitud del camí per a tots els parells possibles i dona informació sobre la proximitat dels nodes.

$$LM\ (G) = \frac{n}{\sum d_{i,j}} \qquad (4)$$

- Coeficient de Clustering (C) [4]: Quantifica l'agrupació que es troba un vèrtex respecte dels seus veïns.

$$C\ (G) = \frac{\sum_{i,j} d_{i,j}}{\frac{n}{2}} \qquad (5)$$

8.1. Dataset

Com a mostra utilitzada per l'experiment, s'ha seleccionat un conjunt de dades públiques, format per un conjunt de dades reals.

Per aquest projecte s'ha emprat la *Dolphin Social Network*[28]. Es tracta d'un graf no dirigit extret d'una xarxa de les associacions freqüents entre una comunitat de 62 dofins a Nova Zelanda els indicadors del *dataset* inicial es troben descrits a la taula 2.

Algoritme	Nodes	Arestes	DM	D	LM	DG
Dolphin Graf	62	159	5,129	6	2,805	0,084

Taula 2. Propietats del graf inicial: nombre de nodes, nombre d'arestes, distancia mitja, diàmetre, longitud mitjana i densitat del graf.

8.2. Procediment

Tal com s'ha especificat prèviament el nostre *framework* experimental consta de tres mètodes: Dintre del *Graph-Modification* utilitzarem dues de les tècniques que s'han explicat prèviament, com són: *random perturbation* (RP), amb d'aquest algoritme s'ha provat les dues estratègies possibles (afegir/eliminar arestes i canviar arestes de forma aleatòria amb un percentatge d'arestes modificades del 2%, 5% i 10%), s'ha aplicat el model de k-anonimitat amb una k=10 sobre el *dataset* original i finalment s'ha aplicat l'algoritme de *Differential Privacy*, els quals són els encarregats d'injectar soroll a les dades del *dataset* original.

Per la realització de les proves s'han emprat els següents algoritmes:

1. EAGA [20][15] Algoritme que utilitza el mètode k-anonim per a xarxes de nodes, va variant la seqüència de grau dels nodes emprant algoritmes evolutius, i a continuació utilitza les tècniques d'intercanvi d'arestes o la d'afegir/eliminar arestes a l'estructura de la xarxa per tal d'aplicar l'anonimització del graf resultant.
2. RANDOM[19]: Algoritme aleatori que utilitza el mètode *random perturbation*, que ens permet aplicar sobre el nostre graf original les estratègies de: Afegir/Eliminar arestes de forma aleatòria i canviar d'arestes de forma aleatòria.
3. Diff-Privacy[14][18]: Algoritme adaptat a treballar amb grafs, on s'utilitza el graf original com a input i mitjançant l'algoritme de *Markov chain Monte Carlo*[27] crea un graf resultant que satisfà el *Differential Privacy*.

A l'hora de treballar amb els algoritmes, l'usuari ha de introduir el *dataset* escollit i els paràmetres adients per a cada algoritme com es descriuen a la taula que s'exposa a continuació:

Algoritme	Paràmetre	Valor
RP_1 - Afegir/Elim. Eix	% arèstes mod.	2
RP_2 - Afegir/Elim. Eix	% arèstes mod.	5
RP_2 - Afegir/Elim. Eix	% arèstes mod.	10
RP_1 - Canvi Arestes	% arèstes mod.	2
RP_2- Canvi Arestes	% arèstes mod.	5
RP_3- Canvi Arestes	% arèstes mod.	10
k-anonimitat	k	10
Differential Privacy	Privacy Budget	10

Taula 3. Paràmetres de configuració: Definint com a paràmetres el *percentatge d'arestes modificades*, el valor de *k* en referència a la k-anonimitat o el *privacy budget* establert per a cada un dels algoritmes d'anonimització.

D'altra banda, s'ha utilitzat l'eina de visualització de grafs Gephy. Es tracta d'una eina de software lliure que ens permet visualitzar les xarxes en forma de graf i avaluar els resultats, mitjançant les pròpies extensions de la pròpia eina, dels càlculs dels indicadors del graf inicial de forma ràpida i automàtica.

Així mateix, un cop aplicats els algoritmes d'anonimització i exportat els grafs perturbats amb el format adient, s'han pogut calcular també els indicadors dels grafs anonimitzats.

8.3. Resultats

En aquest apartat es volen mostrar els resultats obtinguts dels experiments realitzats. Al *dataset* inicial s'ha mesurat, mitjançant els indicador prèviament descrits, quin ha estat la variació final respecte la inicial en cada cas per tal de veure la diferència un cop s'han aplicat els tres mètodes d'anonimització amb els que hem treballat, tal com es detalla a la secció 8.2. Procediment.

Les dades genèriques del indicador de perduda d'informació es poden veure a la taula 4.

Algoritme	Nodes	Arestes	DM	D	LM	DG	C
Dataset Original	62	159	5,129	6	2,805	0,084	0,209
RP_1 - Afegir/Elim. Eix	62	159	5,129	8	3,188	0,084	0,28
RP_2 - Afegir/Elim. Eix	62	159	5,129	8	3,081	0,084	0,267
RP_3 - Afegir/Elim. Eix	62	159	5,129	6	2,986	0,084	0,251
RP_1 - Canvi Arestes	62	159	5,129	9	3,51	0,084	0,325
RP_2 - Canvi Arestes	62	159	5,129	8	3,175	0,084	0,257
RP_3 - Canvi Arestes	62	159	5,129	7	2,948	0,084	0,258
k-Anonimitat	62	159	5,129	8	3,357	0,084	0,303
Differential Privacy	62	322	5,279	8	3,377	0,088	0,273

Taula 4. Resultats obtinguts després de passar cadascun dels mètodes d'anonimització en comparació amb els resultats del graf original.

Si ens fixem en el detall de les dades s'observa que en tots els casos es preserva el nombre de nodes (62) en totes les metodologies tal i com s'havia explicat en les seccions 6 i 7 respectivament. És amb l'indicador d'arestes que es pot observar a la Taula 4. la primera diferència entre els algoritmes *Differential Privacy* respecte als altres mètodes, ja que com s'ha presentat en la secció 6, el *Graph-Modification* i les seves tècniques que s'han provat en aquests experiments, únicament fan variacions sobre arestes i no sobre nodes, es per això que el valor obtingut de *differential privacy* (322) es diferent a tots els altres mètodes i el *dataset* inicial(159). Així doncs, com existeix la mateixa situació que en el cas anterior amb el indicador de Distància Mitja, on per a tots els tipus d'algoritme relacionat amb el mètode *Graph-Modification* es manté la mateixa distància mitja per a tots (5,129) a diferència de l'obtinguda per el *Differential Privacy* (5,279). També ocorre el mateix amb l'indicador de la densitat del grau, seguint la lògica de la fórmula (3) al tenir en tots els casos del algoritme *Graph-Modification* el mateix nombre de nodes, és normal que tots tinguin el mateix valor (0.084) a diferència del altre algoritme (0,088). Tot i així, cal destacar que tant en els casos de densitat del graf com el de distancia mitja la diferencia entre un algoritme i l'altre no es tant destacada com l'anterior.

En el cas del indicador del diàmetre, s'observa una fluctuació de valors (Figura 4) entre els diferents algoritmes segons quin paràmetre se li ajustes. Partint de que el diàmetre inicial és 6, només l'algoritme de *Random Perturbation*, que utilitza la tècnica de afegir/eliminar nodes amb un percentatge d'arestes modificades de 10%, és el que aconsegueix tenir el mateix valor que el graf inicial, mentre que tots els altres algoritmes o la pròpia variació dels algoritmes amb paràmetres diferents, difereixen del *dataset* inicial.

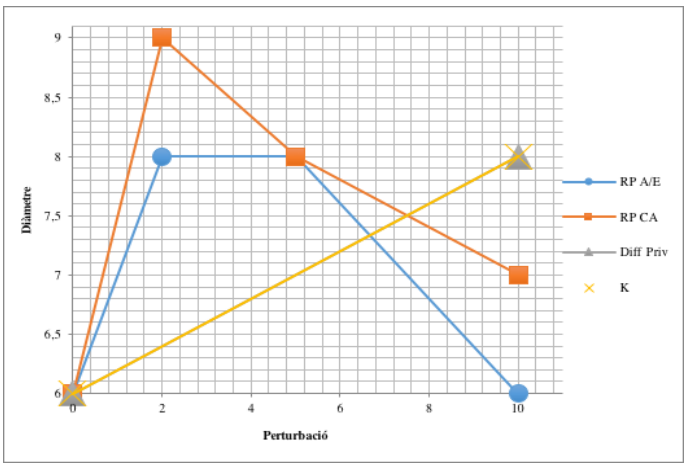

Figura 4. Valors obtinguts de l'indicador del Diàmetre per observar la difèrencia del *dataset* original respecte els diferents *datasets* anonimitzats.

En els valors sobre la longitud mitjana, la fluctuació de dades es mou en un rang màxim de 5 dècimes respecte al valor del *dataset* original, tal i com es pot observar a la Figura 5. És una altre cop, l'algoritme de *Random Perturbation*, amb un percentatge d'arestes modificades de 10%, el que aconsegueix tenir el mateix valor que el *dataset* inicial.

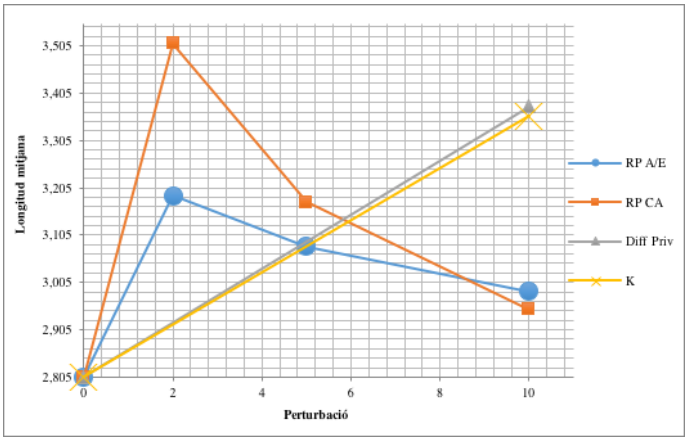

Figura 5. Valors obtinguts del indicador Longitud Mitjana per observar la diferencia del *dataset* original respecte als diferents *datasets* anonimitzats.

Cal destacar l'indicador del Coeficient de Clustering, el qual calcula el grau d'interconnexió amb els seus veïns, ja que els valors obtinguts només difereixen en un màxim de 1,2 dècimes en el cas de l'algoritme Random Perturbatión amb la tècnica de canvi d'arestes amb un percentatge del 2%, mentre que un cop més, és l'algoritme de Random Perturbation que utilitza la tècnica de afegir/eliminar nodes, amb un percentatge d'arestes modificades de 10%, el que aconsegueix tenir el valor més proper al *dataset* inicial, tal i com es pot observar a la Figura 6.

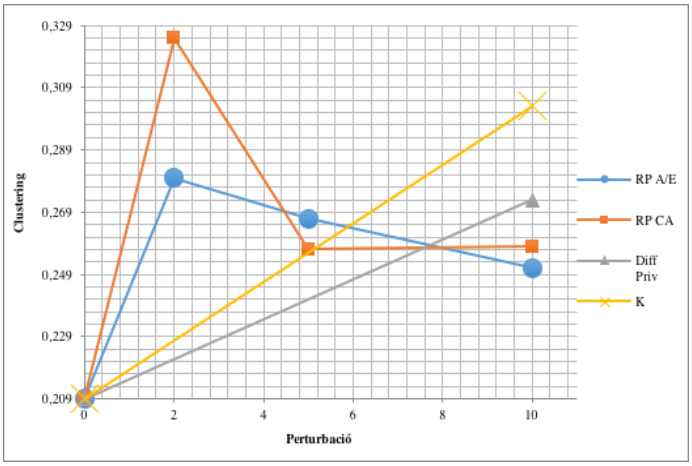

Figura 6. Valors obtinguts del indicador Coeficient de Clustering per observar la diferencia del *dataset* original respecte els diferents *datasets* anonimitzats.

Finalment, s'han escollit els valors dels indicadors de Distància mitjana, Diàmetre, Longitud Mitjana i el Coeficient de Clustering dels *datasets* resultants dels algoritmes de *Differential Privacy*, k-anonimitat i dintre l'algoritme de *Graph-Modification*, la tècnica de *Random Perturbation* amb un percentatge de 10% d'arestes modificades per a la realització d'un gràfic comparatiu, tal i com es pot observar a la Figura 7, amb el qual es podrà comparar de forma visual els indicadors dels *datasets* originats respecte al *dataset* incial.

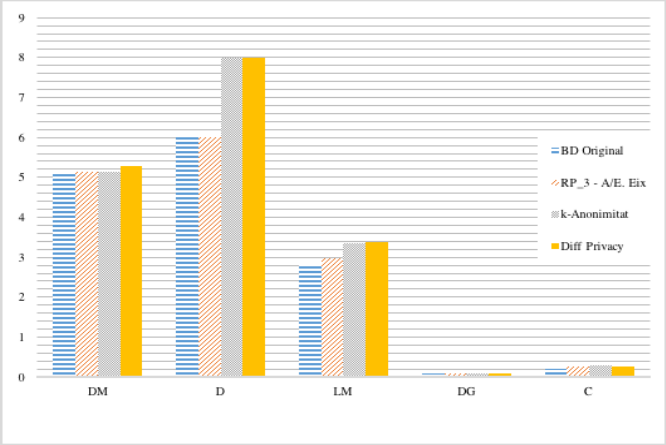

Figura 7. Gràfic comparatiu dels indicadors Distància Mitja, Diàmetre, Longitud Mitjana i Coeficient de Clustering, entre el *dataset* original i els algoritmes de *random pertubation* (amb la tècnica el percentatge de modificació d'arestes és el), k-anonimitat

S'ha escollit comparar la pèrdua d'informació dels algoritmes d'anonimització davant de la problemàtica existent a l'hora de comparar el grau de privacitat que ofereixen les dues metodologies a causa del diferent tractament de dades i conseqüent format de dades, ja que és l'únic indicador i mètrica que es pot obtenir en comú de les metodologies estudiades.

Així com, cal destacar que l'algoritme de *Differential Privacy* no pot ser pensat i utilitzat per a bases de dades de mida petita, ja que llavors perd la seva essència sobre la màxima en que es basa aquest algoritme, ja que l'afirmació que "un sol individu no ha de poder afectar el resultat de la base de dades" no tindria gaire sentit perquè qualsevol canvi afectaria proporcionalment a l'estructura de nodes de la base de dades.

Mentre que amb els diversos mètodes de *Graph-Modification*, es presenta la problemàtica que són problemes NP-dificil i que per tant, a mesura que va creixen la mida del graf anonimitzat, es va convertint en un problema intractable, ja que per definició els problemes NP-dificil només són tractables amb conjunts de dades petites.

9. Conclusions

En aquest paper s'ha realitzat un estudi sobre l'estat de l'art actual enfront de la problemàtica que existeix a la nostra societat en matèria de tractament i privacitat de dades.

Primer de tot, s'ha realitzat una planificació d'aquest projecte on s'han anat marcant una sèrie de fites a assolir. A l'hora de plantejar el problema s'ha fet recerca sobre els estudis dels diferents experts en privacitat per tal de poder veure els diferents principals riscos i les conseqüències associades a un mal tractament sobre dades sensibles.

Un cop realitzada la definició del problema s'han exposat les principals tècniques d'anonimització de dades, on s'ha destacat l'ús dels mètodes pertorbatius, els quals introdueixen soroll a les bases de dades originals. Aquest serveix per si es produeix un atac per part d'un adversari, no sigui capaç de re-identificar cap dada o node.

El projecte s'ha centrat bàsicament en l'estudi i posterior exposició dels dos principals mètodes: *Graph-Modification* i *Differential Privacy*, on en el cas del primer mètode, s'han estudiat també les diferents estratègies o tècniques. Considerant aquests principals mètodes d'anonimització, s'ha procedit a la realització d'una sèrie de proves empíriques sobre un conjunt de dades reals utilitzant un *dataset* d'una xarxa universitària, on cada node representava a una persona dintre de la xarxa.

Finalment, s'han avaluat els diferents resultats obtinguts després d'anonimitzar els nodes a partir de les tècniques de: *Graph modification* amb random pertubation, k-anonimitat i *Differential Privacy*. Això ha servit per poder analitzar en quina mesura quedava afectada la pèrdua d'informació respecte al *dataset* del graf original.

10. Referències

[1] Aggarwal CC and Wang H (eds) (2010) Managing and Mining Graph Data. Springer, New York

[2] Guimerà R, Danon L, Díaz-Guilera A, Giralt F and Arenas A (2003) Self-similar community structure in a network of human interactions. Phys Rev E 68(6):065103

[3] Hay M, Miklau G, Jensen D,Weis P and Srivastava S (2007) Anonymizing Social Networks. Report, University of Massachusetts Amherst

[4] Hay M, Miklau G, Jensen D, Towsley D and Weis P (2008) Resisting structural reidentification in anonymized social networks. Proc VLDB Endow 1(1):102-114

[5] Ying X, Pan K, Wu X and Guo L (2009) Comparisons of randomization and k-degree anonymization schemes for privacy preserving social network publishing. In: Proceedings of the 3rd Workshop on Social Network Mining and Analysis (SNA-KDD). ACM Press, New York, pp 10:1-10:10

[6] T. Dalenius, Towards a methodology for statistical disclosure control. Statistik Tidskrift 15, pp. 429–222, 1977.

[7] Liu K, Terzi E (2008) Towards identity anonymization on graphs. In: ACMSIGMOD international conference on management of data, SIGMOD '08, ACM Press, New York, NY, USA, pp 93–106.

[8] Zhou B, Pei J (2008) Preserving privacy in social networks against neighborhood attacks. In: IEEE International

conference on data engineering (ICDE), IEEE Computer Society, Washington, DC, USA, pp 506–515

[9] Hay M, Miklau G, Jensen D,Weis P, Srivastava S (2007) Anonymizing social networks. Technical report No. 07-19, Computer Science Department, University of Massachusetts Amherst, UMass Amherst (Random Perturbation)

[10] L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. International Journal of Uncertainty, Fuzziness, and Knowledge-based Systems, vol. 10(5), pp. 571–588, 2002.

[11] Benjamin C. M. Fung,Ke Wang, Ada Wai-Chee Fu, and Philip S. Yu (2011). Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques. United States of America: CRC Press.

[12] Sweeney, L. (2002). k-anonymity: a model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 10(5), pp. 557–570.

[13] Vicenç Torra (2010). Privacy in Data Mining. Data Mining and Knowledge Discovery Handbook, pp. 687-716. Springer.

[14] Qian Xiao, Rui Chen, Kian-Lee Tan, Differentially Private Network Data Release via Structural Inference

[15] Casas-Roma J, Herrera-Joancomartí J, TorraV (2013) An algorithm for k-degree anonymity on large networks. In: IEEE international conference on advances on social networks analysis and mining (ASONAM), Niagara Falls, CA, IEEE Computer Society, pp 671–675

[16] Casas-Roma J, Herrera-Joancomartí J, TorraV (2013) A survey of graph-modification techniques for privacy-preserving on networks. Springer Science Business Media Dordrecht 2016.

[17] Dwork, C. (2006). Diferential privacy. In Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP).

[18] XIAO Qian, Differential Private Network Structural Inference, Disponble en: https://github.com/kaseyxiao/privHRG

[19] J. Casas-Roma, RGO (Random Graph Obfuscation) Algorithm, Disponible en: https://jcasasr.wordpress.com/software/random-based-algorithm/

[20] J. Casas-Roma, EAGA (Evolutionary Algorithm for Graph Anonymization) Algorithm, Disponible en: https://jcasasr.wordpress.com/software/eaga-algorithm/

[21] J. Casas-Roma, "Seguridad y privacidad en las Smart Cities". 2015.

[22] Cynthia Dwork. "A Firm Foundation for Private Data Analysis." In communications of the ACM, VOL 54, 2011, pp. 86-95.

[23] Frank McSherry, Kunal Talwar, "Mechanism Design via Differential Privacy".

[24] Paass, G. (1985) Disclosure risk and disclosure avoidance for microdata, Journal of Business and Economic Statistics 6 487-500.

[25] Lambert, D. (1993) Measures of Disclosure Risk and Harm, Journal of Official Statistics 9 313-331.

[26] Samarati, P. (2001) Protecting Respondents' Identities in Microdata Release, IEEE Trans. on Knowledge and Data Engineering, 13:6 1010-1027.

[27] Casas-Roma J, Herrera-Joancomartí J, Torra V. Anonymizing Graphs: Measuring Quality for Clustering

[28] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, Behavioral Ecology and Sociobiology 54, 396-405 (2003).