Privacidad y anonimización de datos

PID_00247391

Jordi Casas Roma

Tiempo mínimo de dedicación recomendado: 4 horas

Universitat Oberta de Catalunya

Índice

Introducción

En este módulo didáctico trataremos el papel que tiene la preservación de la privacidad en el proceso de publicación de datos. En concreto, veremos los principales modelos de preservación de la privacidad, entre los cuales destacamos los métodos de aleatorización o perturbación de datos, el modelo de *k*-anonimidad y el modelo de la privacidad diferencial.

A continuación, profundizaremos en los métodos de enmascaramiento de datos y el modelo de *k*-anonimidad, que representa uno de los modelos más conocidos y empleados en los procesos de anonimización. Veremos la teoría subyacente a cada modelo, así como ejemplos de aplicación en datos estructurados en formato de tablas y, también, en datos semiestructurados en formato de red o grafo.

El objetivo de estos procesos es asegurar la privacidad de los datos de los individuos cuando se publica información que contiene datos personales. Por un lado, la publicación de estos datos es muy útil para la investigación que realizan instituciones, universidades y empresas; pero por otro lado, se debe evitar la violación de privacidad que pudieran sufrir los individuos que aparecen en estos conjuntos de datos.

Objetivos

En los materiales didácticos de este módulo encontraremos las herramientas indispensables para asimilar los siguientes objetivos:

1. Conocer las principales amenazas relacionadas con la privacidad de los usuarios que puede suponer la publicación de datos.
2. Conocer los principales modelos teóricos que permiten preservar la privacidad de los usuarios en procesos de publicación de la información.
3. Conocer qué mecanismos de enmascaramiento de datos pueden evitar o dificultar la tarea de identificación de usuarios por parte de un atacante.
4. Identificar cuál es el método más apropiado para cada entorno de publicación de datos, considerando el objetivo del posterior análisis a realizar y el tipo de datos a emplear.
5. Conocer los principales modelos de anonimización en datos estructurados en formato de tablas.
6. Conocer los principales modelos de anonimización en datos semiestructurados en formato de red o grafo.

1. Introducción

La minería de datos (*data mining*, en inglés) es el proceso de extraer información útil, interesante y desconocida hasta el momento de conjuntos de datos. El éxito de la minería de datos se basa en la disponibilidad de datos de calidad sobre los que ejecutar estos procesos. En este sentido, la recopilación de la información digital por parte de gobiernos, corporaciones e individuos debe facilitar el intercambio y la disponibilidad de datos a gran escala para su posterior análisis. Existe una demanda de intercambio de datos entre varios actores impulsada, por una parte, por los beneficios mutuos y, por otra parte, por las regulaciones que requieren que ciertos datos sean publicados. Por ejemplo, en Estados Unidos se requiere que los hospitales de California publiquen los datos demográficos específicos de todos los pacientes dados de alta en sus instalaciones. Generalmente, la publicación de datos en abierto (*open data*, en inglés) que incluyan datos personales puede inducir a un quebrantamiento de la privacidad, si no se tratan de forma adecuada.

Lecturas complementarias

D. M. Carlisle; M. L. Rodrian; C. L. Diamond. (2007, julio) *California inpatient data reporting manual, medical information reporting for California, 5th edition*. Informe técnico. Oficina de Desarrollo y Planificación Sanitaria Estatal.

1.1. Antecedentes y contextualización

En junio de 2004, el Comité Consultivo de Tecnologías de la Información (*Information Technology Advisory Committee*) de Estados Unidos publicó un informe titulado «Revolucionando la atención sanitaria a través de las tecnologías de la información». Un punto clave fue establecer un sistema nacional de registros médicos electrónicos que fomentara el intercambio de conocimientos médicos. Ejemplos similares se pueden encontrar en prácticamente todos los dominios. Por ejemplo, Netflix*, un popular servicio de alquiler de películas en línea publicó un conjunto de datos que contiene calificaciones de sus películas de quinientos mil suscriptores, en un intento por mejorar la precisión de las recomendaciones de las películas basadas en las preferencias personales. De forma similar, AOL**, un conocido buscador de páginas web que opera en Estados Unidos, publicó un conjunto de registros de consultas, pero rápidamente se vio obligado a retirar los datos debido a la identificación de un usuario en los datos.

*** http://www.netflix.com**
**** http://www.aol.com/**

Lecturas complementarias

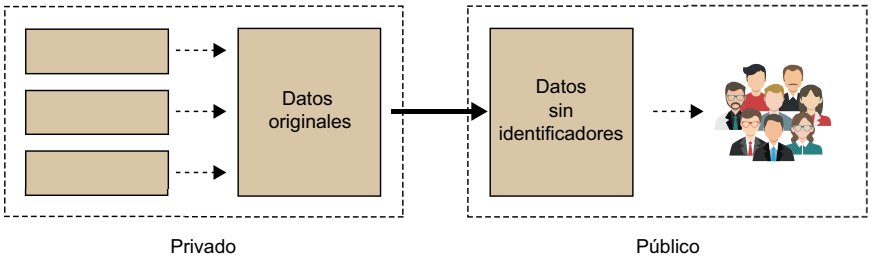
M. Barbaro y T. Zeller (2006, 9 de agosto). «A face is exposed for AOL searcher no. 4417749». *New York Times*.

Una tarea de gran importancia es el desarrollo de métodos y herramientas que permitan la publicación de datos, de manera que los datos publicados mantengan su utilidad al mismo tiempo que preservan la privacidad de los usuarios que aparecen en ellos. Este proceso se llama preservación de la privacidad en la publicación de datos (*privacy-preserving data publishing* o PPDP, en inglés), que puede ser visto como una respuesta de carácter técnico para complementar las políticas de privacidad que cada país o región implementa.

1.2. Publicación de datos

El escenario típico de la recopilación y publicación de datos se describe en la figura 1. En la fase de recogida de datos, el titular de los datos recopila información de los distintos usuarios. A continuación, el propietario de los datos recopilados, «datos originales» en la figura, debe proteger y asegurar la privacidad de los usuarios que aparecen antes de hacer público el conjunto de datos recopilados. Este proceso, llamado *anonimización* o *preservación de la privacidad*, será el encargado de asegurar que no es posible identificar a un usuario dentro del conjunto de datos protegidos o anónimos. En la fase de publicación de los datos, el propietario de los datos recopilados publica los datos protegidos para su posterior explotación. Esta publicación de datos protegidos puede hacerse de forma pública y accesible para cualquier persona o entidad, o bien de forma privada a un conjunto de empresas o centros autorizados.

Figura 1. Escenario básico para la publicación de datos



Publicación de datos médicos

Por ejemplo, un hospital recoge datos de los pacientes y comparte los registros de los pacientes con un centro médico externo. En este ejemplo, el hospital es el titular de los datos, los pacientes son propietarios de sus propios datos y el centro médico externo es el receptor de los datos. Las tareas de minería de datos que el centro médico externo puede realizar sobre los datos protegidos pueden ser de cualquier tipo, desde un simple recuento del número de hombres con diabetes hasta un sofisticado análisis de grupos de pacientes según sus características fisiológicas y demográficas.

1.3. Preservación de la privacidad en la publicación de datos

Desde el punto de vista de la privacidad o anonimización, los atributos de un conjunto de datos se dividen en cuatro clases, según el tipo de información que contienen:

- Los **identificadores** son un conjunto de atributos que permiten identificar de forma explícita a un individuo. El nombre, DNI o número de la seguridad social son ejemplos de atributos identificadores.

- Los **cuasi identificadores** son un conjunto de atributos que potencialmente podrían identificar a un individuo.

- Los **atributos sensibles** presentan información específica y sensible de un individuo en concreto, como por ejemplo las enfermedades que padece, su salario o sus preferencias sexuales o religiosas.

- Finalmente, los **atributos no sensibles** son todos los atributos que no caben en ninguna de las categorías anteriores.

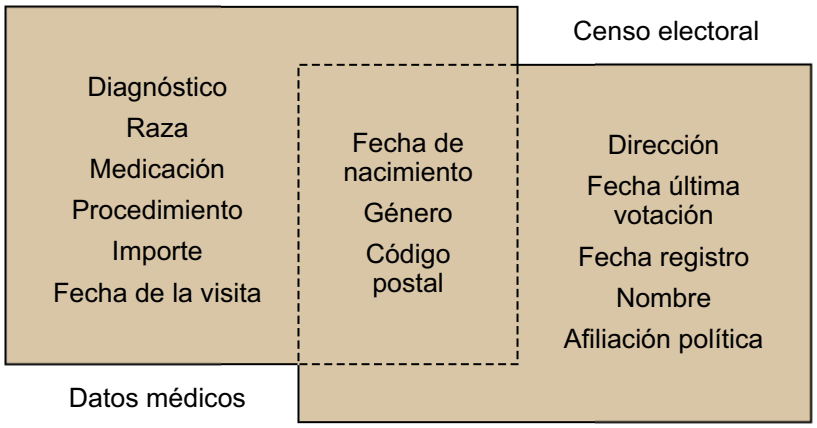
Obviamente, los atributos identificadores deben ser eliminados antes de publicar los datos. En caso contrario, la identificación de los usuarios es directa y trivial en los datos publicados.

Aun eliminando todos los identificadores, un estudio de Sweeney en el año 2002 consiguió romper la privacidad de un gobernador de Estados Unidos. En este trabajo, el nombre y otros datos públicos del censo electoral fueron combinados con una base de datos médicos utilizando el código postal, la fecha de nacimiento y el género. La figura 2 muestra la intersección de datos entre los dos conjuntos. Ninguno de estos atributos puede utilizarse para identificar a una persona de forma única, pero su combinación conduce frecuentemente a identificar a una única persona o a un grupo reducido de personas. Estos atributos son los llamados cuasi identificadores. De esta forma, los autores de este trabajo consiguieron identificar los datos médicos del gobernador después de combinarlos con los datos públicos del censo electoral. Estudios posteriores mostraron que el 87 % de la población de Estados Unidos podía ser identificada de forma similar utilizando información públicamente accesible.

Lectura complementaria

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

Figura 2. Combinación de atributos cuasi identificadores usado por Sweeney en 2002



En el ejemplo anterior, la identidad de un registro se ve comprometida por medio de la combinación de distintos cuasi identificadores. Para llevar a cabo este tipo de ataques a la privacidad de los usuarios es necesario que el atacante tenga cierto conocimiento externo del usuario objetivo. Este conocimiento puede ser obtenido mediante muchas y muy distintas fuentes. Por ejemplo, un atacante puede darse cuenta de que su jefe ha sido hospitalizado durante unos días concretos, por lo tanto puede saber que aparecerá en los registros que se

hagan públicos de los pacientes del hospital. Por otro lado, no le será muy difícil descubrir el código postal, fecha de nacimiento y género del individuo en cuestión, con lo cual puede efectuar este ataque para descubrir información confidencial de su jefe.

La tabla 1 presenta un ejemplo de conjunto de datos médicos. Como se puede ver, los identificadores han sido eliminados del conjunto, los cuasi identificadores contienen los valores originales (código postal, género y edad) y se mantiene inalterable el atributo sensible (enfermedad) que normalmente es la información más interesante para el análisis o minería de datos, pero también es la información que debemos evitar que pueda ser vinculada a un usuario concreto.

Tabla 1. Ejemplo de tabla de datos médicos

ID	Código postal	Género	Edad	Enfermedad
1	08500	Hombre	29	Cáncer
2	17600	Mujer	48	Hepatitis
3	08242	Hombre	21	Gripe
4	25128	Hombre	36	Cáncer
5	17488	Mujer	45	Diabetes
6	08401	Mujer	58	Gripe
7	08760	Mujer	72	Hepatitis
8	43840	Mujer	22	Cáncer
9	43500	Hombre	25	Diabetes
10	25310	Mujer	43	Gripe

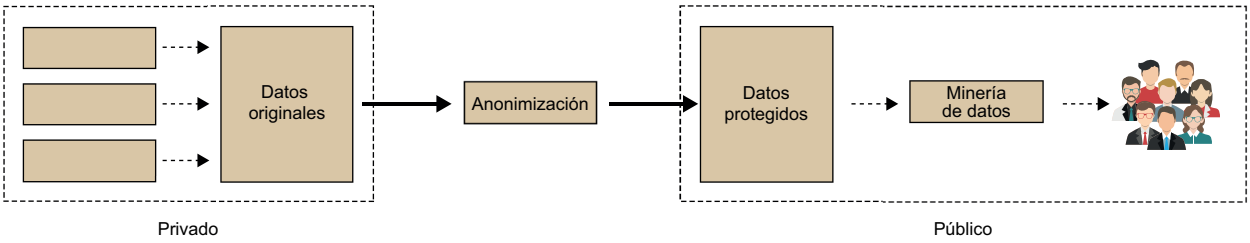
En estos datos de ejemplo, un atacante puede identificar a un usuario de forma similar al ataque realizado por Sweeney en el año 2002. Utilizando una combinación de los cuasi identificadores obtenemos registros individuales, que permiten identificar de forma única a un usuario dentro del conjunto. Por ejemplo, si una persona conocida por el atacante fue hospitalizada durante el periodo previo a la publicación de los datos y el atacante sabe que es un hombre, que reside en la población con código postal 08242 y que tiene veintiún años, entonces el atacante puede saber, sin lugar a dudas, que el registro médico del usuario corresponde al registro número 3 de los datos publicados, y que por lo tanto esta persona sufrió una gripe.

Para evitar este tipo de ataques, el propietario de los datos debe aplicar una serie de operaciones sobre los cuasi identificadores para impedir que puedan ser usados para identificar a cualquier usuario dentro de los datos protegidos.

La anonimización o PPDP persigue ocultar la identidad y la información sensible de los usuarios que aparecen en conjuntos de datos, asumiendo al mismo tiempo que la utilidad de los datos debe ser retenida en los datos protegidos. Es decir, el análisis ejecutado sobre los datos protegidos debe revelar información útil y verdadera, de forma similar a los resultados que se obtendrían en los mismos análisis utilizando los datos originales (no protegidos).

Existen multitud de estrategias de anonimización, pero en general este tipo de técnicas buscan las formas de ocultar los detalles que puedan hacer a un individuo único dentro del conjunto de datos. El objetivo es que un único individuo sea indistinguible respecto a un conjunto de individuos suficientemente grande para proteger su identidad, de tal forma que el atacante solo puede deducir cierta información con una cierta probabilidad. La figura 3 muestra que, a diferencia del escenario básico de publicación de datos visto anteriormente, en este caso se añade la tarea de anonimización o PPDP previamente a la publicación de los datos.

Figura 3. Escenario básico considerando la preservación de la privacidad en la publicación de datos



2. Modelos teóricos de privacidad

En este apartado veremos algunos de los principales modelos para la preservación de la privacidad en procesos de publicación de datos. En concreto, veremos tres modelos muy relevantes en tareas de anonimización, como son la aleatorización o perturbación de los datos, el modelo de *k*-anonimidad y la privacidad diferencial.

2.1. Tipología de modelos de protección

Esencialmente, existen dos enfoques principales para limitar el riesgo de divulgación en procesos de publicación de datos:

- **Protección no interactiva**, mediante la cual se genera y se libera una versión protegida del conjunto de datos original recopilado de los sujetos de datos.

- **Protección interactiva**, mediante la cual se realiza un análisis de datos consultado por el usuario en el conjunto de datos original y, a continuación, se devuelve una versión protegida de los resultados al usuario.

Cuando el tipo de análisis de datos es desconocido en el momento de la protección y publicación de datos, la protección no interactiva es la única solución viable. Sin embargo, para un análisis establecido de datos conocidos previamente, se puede optar por la protección interactiva, ya que permite ajustar el nivel de protección al análisis que se está realizando, lo que teóricamente permite maximizar la utilidad y precisión de los resultados.

2.2. Aleatorización

El modelo de *aleatorización* o *perturbación de datos* consiste, simplemente, en introducir ruido en los datos originales, de tal forma que un atacante no pueda saber, a ciencia cierta, si la información que está extrayendo es cierta o ha sido alterada durante este proceso de anonimización aleatoria.

Por otro lado, si se introduce demasiado ruido en los datos originales, la privacidad quedará preservada pero la utilidad de los datos puede llegar a ser nula. En efecto, cuando el grado de perturbación o introducción de ruido en los datos originales es muy elevado, estamos generando datos aleatorios, de

forma que el riesgo de romper la privacidad será nula, pero también será nula la información que se puede extraer de los datos.

Por lo tanto, un proceso de aleatorización debe introducir una cantidad de ruido que:

- sea suficiente para que un atacante no pueda estar seguro de la veracidad de un dato en concreto,
- y, al mismo tiempo, los datos y la información general del conjunto de datos se debe preservar para que los análisis realizados sobre el conjunto de datos anónimos sea próximo (tan próximo como sea posible) al resultado obtenido utilizando el conjunto de datos original.

Ejemplo de perturbación aleatoria

Un ejemplo típico de perturbación aleatoria en datos numéricos consiste en añadir ruido a los atributos numéricos. La tabla 2 muestra un ejemplo en donde vemos el salario de un conjunto de diez individuos. Este valor, que se considera sensible, debe ser anonimizado para preservar la identidad de los individuos, ya que si el atacante conoce el salario de uno de los usuarios que aparece en el conjunto de datos, podría ser capaz de reidentificar al usario en el conjunto de datos anónimo y, por lo tanto, inferir otros datos asociados al usuario (no mostrados en esta tabla de ejemplo).

La tercera columna muestra una posible configuración de valores después de añadir ruido aleatoriamente por medio de una distribución normal, con el objetivo de perturbar los datos originales y dificultar la posible reidentificación de un individuo.

Tabla 2. Ejemplo de salarios y valores perturbados mediante introducción de perturbación aleatoria

ID	Salario ($\times 10^3$)	Valor perturbado ($\times 10^3$)
1	45	45
2	30	30
3	99	106
4	92	93
5	73	67
6	27	27
7	84	80
8	94	94
9	62	64
10	14	16

Es interesante notar que, aunque los valores han sido modificados, el rango de valores perturbados se mantiene próximo al valor original. Así, el valor medio de los salarios en conjunto original es de 62,0, mientras que en los datos perturbados es de 62,2.

Una de las principales debilidades de este modelo es que no suele preservar los valores *outliers*, es decir, valores «extremos» o muy alejados del rango de los demás valores. Con lo cual, aun sabiendo que el valor ha sido perturbado, es posible reidentificar con una probabilidad muy alta al usuario 10, ya que su valor está muy por debajo de los demás.

Lectura complementaria

P. Samarati (2001). *Protecting Respondents' Identities in Microdata Release* (vol. 13(6), págs. 1010-1027). IEEE Transactions on Knowledge and Data Engineering.

Lectura complementaria

L. Sweeney (2002). «*k*-anonymity: a model for protecting privacy». *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*(vol. 10(5), págs. 557-570).

2.3. *k*-anonimidad

El modelo de *k*-anonimidad, introducido por Samarati y Sweeney, es uno de los modelos de protección no interactiva más ampliamente investigado y empleado en la publicación de datos.

La *k*-anonimidad es una propiedad de los datos que garantiza que un individuo no pueda ser distinguido de otros $k-1$ individuos también representados en esos datos.

Para conseguir este objetivo, podemos aplicar diferentes técnicas diferentes, como por ejemplo reemplazar valores concretos por otros valores de una categoría más general o eliminar ciertos valores.

Ejemplo de *k*-anonimidad

Inicialmente, partimos de una base de datos organizada en forma de tabla, como la que se muestra en la tabla 3. Esta tabla muestra la información referente a un grupo de personas que sufren alguna enfermedad. Los cuasi identificadores son su código postal, su edad y su nacionalidad, mientras que la enfermedad sería el atributo sensible que queremos proteger. El objetivo es que cualquier combinación de cuasi identificadores aparezca al menos *k* veces.

Tabla 3. Base de datos no anonimizada

Cuasi identificadores			**Atributos sensibles**
Cod. Postal	Edad	Nacionalidad	Enfermedad
13053	28	Rusa	Arritmia
13068	29	Española	Arritmia
13068	21	Japonesa	Infección
13053	23	Española	Infección
14853	50	India	Cáncer
14853	55	Rusa	Arritmia
14850	47	Española	Infección
14850	49	Española	Infección
13053	31	Española	Cáncer
13053	37	India	Cáncer
13068	36	Japonesa	Cáncer
13068	35	Española	Cáncer

La tabla 4 muestra la versión *k*-anonimizada de la tabla 3. En este ejemplo, se ha dado un valor $k = 4$, por lo que cada combinación de cuasi identificadores aparece cuatro veces. En el caso del código postal y de la edad, se ha utilizado la técnica de generalización. Al eliminar la última o dos últimas cifras del código postal, este sigue aportando información sobre el área geográfica del individuo, pero ahora se trata de un área más extensa. Para la edad se ha generalizado indicando únicamente límites (superiores o inferiores) de edad. La técnica de supresión ha sido aplicada al atributo de nacionalidad, ya que no existen suficientes combinaciones para garantizar *k*-anonimidad para $k = 4$.

Tabla 4. Base de datos anonimizada

Cuasi identificadores			**Atributos sensibles**
Cod. Postal	Edad	Nacionalidad	Enfermedad
130**	<30	*	Arritmia
130**	<30	*	Infección
130**	<30	*	Infección
1485*	>40	*	Cáncer
1485*	>40	*	Arritmia
1485*	>40	*	Infección
1485*	>40	*	Infección
130**	<40	*	Cáncer
130**	<40	*	Cáncer
130**	<40	*	Cáncer
130**	<40	*	Cáncer

La *k*-anonimidad tiene sus ventajas e inconvenientes. Su principal ventaja es que un atacante no puede identificar a su víctima con una probabilidad superior a $\frac{1}{k}$, lo que proporciona un límite teórico asociado al modelo de privacidad. Por otro lado, debemos considerar que el valor de *k* en este modelo es el encargado de establecer el nivel de privacidad, donde al aumentar su valor (y el nivel de privacidad), se reduce la utilidad de los datos.

2.4. Privacidad diferencial

Lectura complementaria

C. Dwork (2006). «Differential Privacy». En: *International Conference on Automata, Languages and Programming*(vol. 4052, págs. 1–12). Conferencia.

La privacidad diferencial fue introducida por Dwork como un modelo de privacidad para la protección interactiva en el contexto de las bases de datos estadísticas, es decir, para proteger los resultados de las consultas a una base de datos. En este contexto, un mecanismo de anonimización se sitúa entre el usuario, que envía las consultas, y el controlador de base de datos, que las responde. Para preservar la privacidad de los individuos, el mecanismo de anonimización debe garantizar que la contribución de los datos de un individuo al resultado global de la consulta es limitada.

Formalmente, un algoritmo o función A es ϵ-diferencialmente privado si, y solo si, para todos los conjuntos de datos D_1 y D_2 que difieren un solo individuo, se cumple:

$$\frac{Pr[A(D_1) \in S]}{Pr[A(D_2) \in S]} \leq e^{\epsilon} \tag{1}$$

donde ϵ es un número real positivo y $S \subset rango(A)$.

De acuerdo con esta definición, la privacidad diferencial es una condición en el mecanismo de publicación, no en el conjunto de datos en sí. Intuitivamente, esto significa que para cualquier par de conjuntos de datos similares, esto es que se diferencian en un solo individuo, un determinado algoritmo diferencialmente privado se comportará aproximadamente del mismo modo en ambos conjuntos de datos. La definición da garantías de que la presencia o ausencia de un individuo no afectará significativamente al resultado final del algoritmo.

La intuición detrás de la privacidad diferencial nos dice que «la presencia o ausencia de un individuo en un conjunto de datos no debe modificar significativamente los resultados del análisis». Este precepto es muy adecuado para la limitación del riesgo en la publicación de datos: si los datos de un individuo tienen un impacto significativo en los resultados de un análisis, probablemente la privacidad de este individuo esté en riesgo. Así, intuitivamente, la privacidad diferencial asegura que los datos estén adecuadamente protegidos.

Ejemplo de privacidad diferencial

Un ejemplo sencillo, pero que nos permite introducir el concepto básico de la privacidad diferencial es el siguiente: supongamos que pedimos a un grupo de personas que responda a la pregunta «¿Tienes la enfermedad X?»

La respuesta del individuo seguirá el siguiente procedimiento:

- Tirar una moneda.
- Si sale «cara», entonces el individuo responderá con honestidad a la pregunta formulada.
- Si sale «cruz», luego se tira la moneda de nuevo y se responde «sí» si sale cara, y «no» si sale cruz.

La confidencialidad proviene de la refutabilidad de las respuestas individuales.

Cuando tenemos una gran cantidad de respuestas, los resultados son significativos, ya que las respuestas positivas se dan en una cuarta parte por las personas que no tienen la enfermedad X y tres cuartas partes por las personas que realmente la tienen. Así, si p es la proporción verdadera de personas con la enfermedad X, entonces esperamos obtener respuestas positivas de:

$$\frac{1}{4}(1-p)+\frac{3}{4}p=\frac{1}{4}+\frac{p}{2}$$

Por lo tanto, es posible estimar p sin comprometer la privacidad de ninguno de los usuarios que responden a la pregunta que les formulamos.

2.5. Resumen

En este apartado hemos descrito la división principal que existe entre los métodos de protección de datos, esto es métodos interactivos y no interactivos. La propia naturaleza del problema y los datos indican cuál es la mejor alternativa en cada caso, aunque como se ha comentado anteriormente, la protección no interactiva es la única solución viable cuando se desea publicar los datos para cualquier tipo de análisis posterior, mientras que la protección interactiva es una buena opción para un análisis establecido de datos conocido previamente.

El modelo de aleatorización permite trabajar con grandes volúmenes de datos, e incluso con datos continuos (*streaming*), dado su bajo coste de cálculo y espacio. Sin embargo, no existen garantías del nivel de protección conseguido por los datos, y especialmente en el caso de los valores *outliers* puede ser posible la reidentificación de ciertos individuos.

El modelo de la k-anonimidad, en cambio, proporciona garantías específicas sobre la probabilidad de reidentificación, que se pueden ajustar mediante el parámetro k. Por contra, su principal debilidad reside en que es necesario formular el conocimiento del adversario sobre la cuasi identificadores para poder aplicar el modelo. Adicionalmente, el coste de cálculo suele ser mucho más alto que en el caso del modelo de aleatorización.

Finalmente, la privacidad diferencial se basa en un método interactivo, lo que permite responder a ciertas preguntas sobre la base de datos, pero no permite la publicación directa de los datos. Sin embargo, a pesar de su popularidad entre los investigadores y el paso adelante que ofrece en términos de garantías de privacidad, la privacidad diferencial solo se está desplegando en un número limitado de aplicaciones del mundo real, debido principalmente a la baja utilidad de los resultados ofrecidos.

Como en muchos de los problemas relacionados con los datos, como por ejemplo la propia minería de datos, no existe un algoritmo óptimo para todos los problemas y todos los entornos de datos. Se deberá analizar y escoger el mejor (o la combinación de varios) para cada caso concreto.

3. Anonimización de datos tabulares

Tradicionalmente, los datos se han presentado en forma de tablas, donde cada fila corresponde a un registro o elemento (un usuario en nuestro caso) y cada columna corresponde a un atributo, propiedad o característica. Por lo tanto, cada registro tiene un valor concreto para cada uno de los atributos de la tabla. La tabla 5 muestra el formato típico de un conjunto de datos en forma de tabla. En este caso, podemos ver un total de n filas o registros y m atributos.

Tabla 5. Ejemplo de tabla con n registros y m atributos

Atributo 1	**Atributo 2**	...	**Atributo *m***
$valor_{1-1}$	$valor_{1-2}$	...	$valor_{1-m}$
...	...	...	...
$valor_{n-1}$	$valor_{n-2}$	...	$valor_{n-m}$

En los últimos años, las técnicas de preservación de la privacidad en tablas han sido ampliamente estudiadas por un grupo importante de investigadores de distintas instituciones y universidades de todo el mundo.

3.1. Métodos de enmascaramiento

Los métodos de enmascaramiento permiten modificar los datos originales con el objetivo de impedir o dificultar la identificación de un usuario en los datos protegidos. Estos métodos se pueden clasificar en tres categorías básicas en función de cómo se manipulan los datos originales para definir el conjunto de datos protegidos.

- **Métodos perturbativos**. El conjunto de datos original es perturbado de algún modo, y el nuevo conjunto de datos puede contener información errónea. Por ejemplo, se puede introducir ruido en algunos atributos, es decir, alterar su valor de forma más o menos aleatoria. De este modo, algunas combinaciones de valores desaparecen en el conjunto de datos protegidos. Al mismo tiempo, las combinaciones en los datos protegidos ya no corresponden a los del conjunto de datos original. Esta ofuscación dificulta la identificación de usuarios en el conjunto de datos protegidos por parte de los atacantes.

- **Métodos no perturbativos**. La protección se logra mediante la sustitución del valor original por otro valor que no es incorrecto pero es menos específico, es decir, más general. Por ejemplo, reemplazamos un número por un intervalo. En general, los métodos no perturbativos reducen el nivel de

detalle del conjunto de datos. Esta reducción del nivel de detalle provoca que diferentes registros tengan las mismas combinaciones de valores, lo que dificulta la identificación de usuarios por parte de un atacante.

- **Generadores de datos sintéticos**. En este caso, en lugar de distorsionar los datos originales, se crean nuevos datos artificiales para sustituir los valores originales. Formalmente, los generadores de datos sintéticos construyen un modelo de datos nuevos a partir del conjunto de datos original y, posteriormente, generan de forma aleatoria un nuevo (y protegido) conjunto de datos que si bien sigue las pautas de los datos originales, no contiene información privada de ningún usuario. Por ejemplo, podemos sustituir la edad de un conjunto de individuos por valores aleatorios generados a partir del valor medio y la varianza observada en los datos originales.

Por otro lado, los métodos de enmascaramiento deben considerar la peculiaridades de los distintos tipos de atributos para reducir las probabilidades de identificación mientras se mantiene la utilidad de los datos protegidos. Los atributos se pueden clasificar en dos categorías básicas:

Lectura complementaria

C. Benjamin; M. Fung; K. Wang; A. Wai-Chee Fu; S. Yu Philip (2011). *Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques*. Estados Unidos: CRC Press.

- Los **atributos numéricos** permiten realizar operaciones aritméticas entre ellos como, por ejemplo, la sustracción o adición. Los ingresos y la edad son ejemplos típicos de tales atributos. Con respecto al riesgo de identificación, los valores de los atributos numéricos son propensos a ser únicos en una base de datos y, por lo tanto, pueden provocar la identificación de usuarios si no se toman medidas oportunas para su anonimización.

- Los **atributos categóricos** pueden tomar valores en un conjunto finito y las operaciones numéricas estándar no tienen sentido en este tipo de atributos. Podemos distinguir tres grupos principales dentro de los atributos categóricos:

  – **Nominales**. El valor de estos atributos se representa mediante etiquetas que proporcionan información. Por ejemplo, el color del pelo o el estado civil son atributos categóricos nominales.

  – **Ordinales**. En este caso los atributos presentan un orden o escala relevante entre ellos. Por ejemplo, el nivel de estudios (primaria, secundaria, bachillerato, grado, etc.) es un atributo categórico ordinal. En estos atributos las operaciones de mínimo y máximo tienen sentido.

  – **Estructurados**. Estos atributos mantienen una relación de clase y subclase entre ellos. Por ejemplo, las profesiones pueden seguir una jerarquía dada, donde por ejemplo dentro de la clase *médico* podemos encontrar muchos especializaciones, tales como ginecólogo, pediatra, etc. En algunos casos la jerarquía puede ser explícita y en otros se puede inferir a partir de los valores dados y de sus relaciones.

A continuación veremos cada uno de los tres tipos de métodos de enmascaramiento y presentaremos algunos métodos concretos y ejemplos que faciliten la comprensión de su funcionamiento. Aun así, una revisión exhaustiva de los métodos de enmascaramiento queda fuera del alcance de este texto.

3.1.1. Métodos perturbativos

Los métodos perturbativos alteran los datos e introducen ruido para dificultar el proceso de identificación de un usuario por parte de un atacante. Existen multitud de métodos en la literatura, aunque una revisión completa de todos ellos escapa a los objetivos de este texto y solo veremos aquí algunos de los métodos más comunes y utilizados en la actualidad.

El primer método que veremos es conocido como *ruido aditivo* (*additive noise*) y consiste en añadir distorsión o ruido en los datos originales, ya sea siguiendo o no la misma distribución de los datos originales. Un ejemplo simple de este tipo de método es introducir el ruido siguiendo una distribución normal $N(0,p\sigma)$, donde σ representa la desviación estándar de los datos originales y p es el parámetro que controla la cantidad de ruido introducido en los datos. Este método fue desarrollado originalmente para tratar con atributos numéricos, aunque posteriormente se han añadido extensiones para poder trabajar con atributos categóricos.

Distribución normal

La distribución normal, también conocida como distribución gaussiana, es una distribución de probabilidad de variable continua. La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado punto.

La tabla 6 muestra una nueva versión del ejemplo que hemos presentado en la tabla 1. En este caso, hemos aplicado el método *additive noise* al atributo «edad», introduciendo ruido en este atributo y disminuyendo, por lo tanto, la probabilidad de que un atacante puede identificar a un usuario dentro de los datos protegidos utilizando información externa.

Tabla 6. Datos médicos después de aplicar *additive noise* en el atributo «edad»

ID	Código postal	Género	Edad	Enfermedad
1	08500	Hombre	32	Cáncer
2	17600	Mujer	46	Hepatitis
3	08242	Hombre	28	Gripe
4	25128	Hombre	38	Cáncer
5	17488	Mujer	48	Diabetes
6	08401	Mujer	61	Gripe
7	08760	Mujer	59	Hepatitis
8	43840	Mujer	20	Cáncer
9	43500	Hombre	24	Diabetes
10	25310	Mujer	55	Gripe

El segundo método que veremos, y que es ampliamente conocido y utilizado por empresas y administraciones, es conocido como *microagregación* (*microaggregation*). Este método se basa en crear grupos de datos según su similitud, y posteriormente reemplazar el valor original de cada dato por el valor promedio de todos los valores del grupo al que pertenece. Por lo tanto, para cada

valor específico de uno o más atributos existirán siempre un conjunto de registros; nunca un registro único que permita que sea identificado un usuario. Este método se puede aplicar sobre un único atributo, y entonces se conoce como microagregación univariante, o sobre dos o más atributos, y en este caso se conoce como microagregación multivariante.

El método permite decidir cuántos registros se juntan en un mismo grupo. Es importante notar que cuantos más registros se juntan en cada grupo, mayor es el nivel de privacidad que se consigue, pero también es mayor el nivel de ruido. Al igual que el método anterior, este también fue desarrollado para atributos numéricos, aunque posteriormente también se han desarrollado extensiones que permiten tratar con atributos categóricos.

La nueva versión del ejemplo, después de aplicar microagregación univariante sobre el atributo «edad» se puede ver en la tabla 7. En este caso hemos parametrizado el método para que cree grupos de dos registros. Como podemos ver en la tabla, para cada posible valor del atributo «edad», siempre existen dos o más registros con el mismo valor. Por ejemplo, en el conjunto de datos original mostrado en la tabla 1, la edad de los usuarios de la fila 1 y 9 es de veintinueve y veinticinco años, respectivamente. Después del proceso aplicado, se ha asignado la edad de veintisiete años a ambos usuarios. De esta forma, no es posible que un atacante pueda identificar de forma única a un usuario utilizando la información de los atributos que hemos utilizado para crear los grupos, en este caso concreto, la edad de los individuos.

Tabla 7. Datos médicos después de aplicar *microagregación univariante* en el atributo «edad»

ID	Código postal	Género	Edad	Enfermedad
1	08500	Hombre	27	Cáncer
2	17600	Mujer	46	Hepatitis
3	08242	Hombre	21	Gripe
4	25128	Hombre	39	Cáncer
5	17488	Mujer	46	Diabetes
6	08401	Mujer	65	Gripe
7	08760	Mujer	65	Hepatitis
8	43840	Mujer	21	Cáncer
9	43500	Hombre	27	Diabetes
10	25310	Mujer	39	Gripe

Para finalizar con los métodos perturbativos, veremos un tercer método que también es muy conocido y utilizado en entornos empresariales y gubernamentales. El método, conocido como *intercambio de rango* (*rank swapping*, en inglés), se basa en intercambiar aleatoriamente los valores de un mismo atributo entre distintos registros. Para conseguir que los datos no sean excesivamente perturbados, este método ordena todos los valores del atributo presentes en la tabla, y a continuación el intercambio se realiza entre valores que se encuentran dentro de un rango acotado y definido como parámetro del método. De esta forma se intenta minimizar el ruido y mantener la utilidad de

los datos protegidos. Este método puede ser aplicado a atributos numéricos y categóricos ordinales.

La tabla 8 muestra los resultados de aplicar el método *rank swapping* sobre el atributo «edad». Como se puede ver, los valores del atributo han sido intercambiados entre los registros con valores próximos. Por ejemplo, si ordenamos los valores del atributo «edad» de los datos originales, mostrados en la tabla 1, obtenemos el vector $\{21,22,25,29,\ldots\}$. Si aplicamos este método en un rango de solo dos posiciones obtenemos el vector $\{22,21,29,25,\ldots\}$, que corresponde a los datos presentados en la tabla 8. Los dos primeros valores corresponden a los registros 3 y 8 que, como se puede apreciar en la tabla, han intercambiado sus valores. De forma similar, los registros 1 y 9 también han intercambiado sus valores durante el proceso de anonimización. Y así, sucesivamente todos los registros de la tabla. Cuanto mayor es el rango que se utiliza en el intercambio, mayor es el grado de privacidad y mayor es, también, la distorsión de los datos originales.

Tabla 8. Datos médicos protegidos aplicando *rank swapping* en el atributo «edad»

ID	Código postal	Género	Edad	Enfermedad
1	08500	Hombre	25	Cáncer
2	17600	Mujer	45	Hepatitis
3	08242	Hombre	22	Gripe
4	25128	Hombre	43	Cáncer
5	17488	Mujer	48	Diabetes
6	08401	Mujer	72	Gripe
7	08760	Mujer	58	Hepatitis
8	43840	Mujer	21	Cáncer
9	43500	Hombre	29	Diabetes
10	25310	Mujer	36	Gripe

3.1.2. Métodos no perturbativos

A diferencia de los métodos vistos en el subapartado anterior, los métodos que veremos en este no introducen ruido o distorsión en los datos originales. La información protegida que se publica continúa siendo totalmente verdadera, aunque se generalizan o suprimen algunas partes de la información que podrían ayudar a un atacante a identificar de forma única a un usuario dentro de los datos protegidos.

Distinguiremos dos métodos básicos, por un lado la generalización de atributos, y por otro la supresión de atributos.

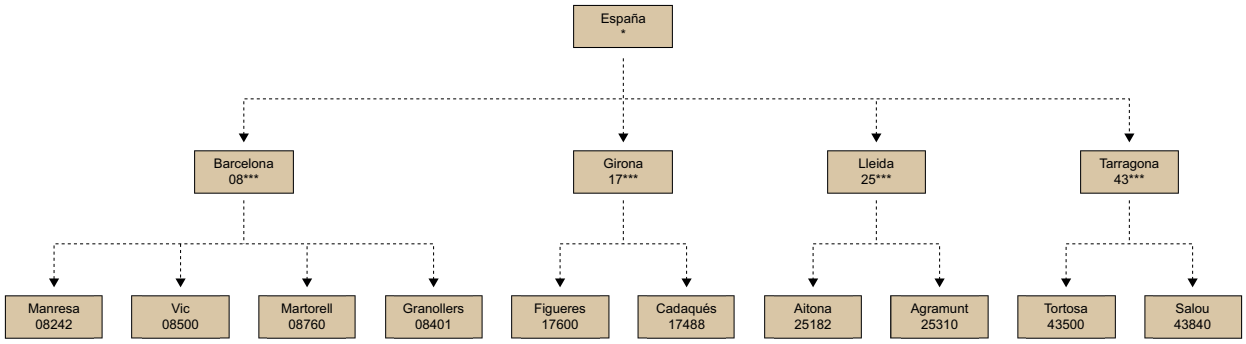
El método de **generalización** se aplica normalmente a atributos categóricos, aunque se puede aplicar a atributos numéricos sin problema alguno. En este caso los datos se protegen reemplazando un conjunto de atributos por un valor más general que los incluya a todos. Por lo tanto, este método no introduce ruido o falsea los datos, simplemente los hace más generales o menos

específicos, de forma que las individualidades de cada usuario del conjunto de datos original quedan difuminadas entre los demás usuarios en el conjunto de datos protegido.

En el caso de los atributos numéricos, la agregación se puede implementar mediante la construcción de rangos de valores. Por ejemplo, si tenemos registros con los valores 3,11,7,19 podemos crear dos grupos y asociar un rango concreto a cada grupo. En este caso, podríamos crear los rangos [0,10) y [10,20), y en este caso el primer y el tercer registro tendrían el mismo valor [0,10) en los datos protegidos.

Por otro lado, en el caso de los atributos categóricos, es necesario disponer de la jerarquía de los atributos, ya sea de forma implícita o explícita. La figura 4 muestra la jerarquía del atributo «código postal». Podemos ver que los distintos municipios se agrupan en provincias, y estas, a su vez, se agrupan formando la raíz de toda la estructura, que en este caso sería el nivel del país. Vemos que podríamos haber incluido un nivel extra que hiciera referencia a las comunidades autónomas, pero se han obviado para mantener la simplicidad del ejemplo. De este modo, si queremos generalizar la información sin perturbarla, podemos generalizar el código postal a nivel de provincia, de tal forma que los datos son absolutamente correctos, pero más generales que los datos originales.

Figura 4. Jerarquía del atributo «código postal»



La tabla 9 muestra el resultado de aplicar la generalización sobre los atributos «código postal» y «edad». En el primer caso, como se puede observar, se han generalizado todos los valores a nivel de provincia, de tal forma que no es posible identificar un municipio de forma única. En el segundo caso, el atributo «edad» ha sido generalizado usando rangos de valores, de tal forma que en lugar de indicar que la edad del individuo que aparece en el registro 1 es veintinueve años, en los datos protegidos indicamos que la edad del individuo está en el rango comprendido entre veinte y treinta años.

El método de **supresión** de atributos consiste, simplemente, en eliminar los atributos. Cuando no es posible utilizar la generalización u otro método y el

Tabla 9. Datos médicos protegidos aplicando *generalización* en los atributos «código postal» y «edad»

ID	Código postal	Género	Edad	Enfermedad
1	08***	Hombre	[20,30)	Cáncer
2	17***	Mujer	[30,50)	Hepatitis
3	08***	Hombre	[20,30)	Gripe
4	25***	Hombre	[30,50)	Cáncer
5	17***	Mujer	[30,50)	Diabetes
6	08***	Mujer	[50,80)	Gripe
7	08***	Mujer	[50,80)	Hepatitis
8	43***	Mujer	[20,30)	Cáncer
9	43***	Hombre	[20,30)	Diabetes
10	25***	Mujer	[30,50)	Gripe

atributo puede presentar una brecha de privacidad, se utiliza este recurso para eliminar el valor del atributo e indicar que el valor ha sido suprimido.

3.1.3. Generación sintética de datos

En los últimos años, ha surgido una nueva tendencia para la preservación de la privacidad de los datos. Consiste en la publicación de datos sintéticos en lugar de los datos originales. La idea principal es que los datos sintéticos no pueden poner en peligro la privacidad de los individuos, dado que los datos no son «reales». Los métodos de generación de datos sintéticos consisten en construir un modelo de los datos y a continuación generar datos artificiales utilizando el modelo generado.

Este enfoque presenta dos dificultades: por un lado, aunque los datos son sintéticos, la identificación de usuarios es aún posible y, por lo tanto, el riesgo de divulgación debe ser analizado para este tipo de datos. Por otra parte, como los datos sintéticos se generan a partir de un modelo de datos particular, construido a partir de los datos originales, todos aquellos aspectos que no están incluidos explícitamente en el modelo no son incluidos en los datos. Debido a esto, el análisis de los datos protegidos podrían llevar a resultados sensiblemente diferentes a los que llevarían los datos originales.

Existen multitud de métodos para la generación de datos sintéticos. La distorsión de datos por medio de la probabilidad de distribución (*data distortion by probability distribution*) fue uno de los primeros procedimientos de protección que pueden ser clasificados como generación sintética de datos. Este procedimiento se define por los tres pasos siguientes:

Función de densidad de probabilidad

En la teoría de la probabilidad, la función de densidad de probabilidad de una variable continua describe la probabilidad relativa según la cual dicha variable aleatoria tomará determinado valor.

- En primer lugar, identificar una función de densidad de probabilidad subyacente a los datos y determinar sus parámetros.

- A continuación, generar series distorsionadas mediante la función de densidad de probabilidad estimada.

- Para finalizar, se sustituye la serie original por la serie distorsionada en los datos protegidos.

El procedimiento se definió originalmente para funciones de densidad univariante (aplicados sobre una variable, en nuestro caso un atributo), aunque también puede ser aplicado a funciones de densidad de multivariantes (más de un atributo).

3.2. *k*-anonimidad en tablas

El modelo de *k*-anonimidad no es un método de enmascaramiento o protección, sino un modelo o condición que debe ser satisfecho por el conjunto de datos protegido. Aun así, generalmente conseguimos cumplir la *k*-anonimidad por medio de los métodos de protección o enmascaramiento que hemos visto en los subapartados anteriores. Es decir, sobre el conjunto de datos originales se les aplica uno o más de los métodos de enmascaramiento vistos anteriormente, con la finalidad de conseguir un conjunto de datos protegido que cumpla con las restricciones o condiciones necesarias para el modelo de la *k*-anonimidad.

Un conjunto de datos cumple el modelo de la ***k*-anonimidad** si, y solo si, para cualquier combinación de atributos cuasi identificadores existen *k* o más registros que comparten los mismos valores.

En otras palabras, cada registro en un conjunto de datos *k*-anónimo es indistinguible de, como mínimo, otros $k-1$ registros con respecto al conjunto de cuasi identificadores. Por lo tanto, la probabilidad de identificación de un usuario en un conjunto de datos *k*-anónimo con respecto a los cuasi identificadores es de como máximo $\frac{1}{k}$.

La tabla 10 muestra el conjunto de datos de ejemplo 2-anónimo, es decir, anonimizado para cumplir con la *k*-anonimidad con un valor de $k = 2$. Según indica el modelo, para cada posible conjunto de atributos cuasi identificadores (código postal, género y edad) siempre existen, al menos, dos registros con los mismos valores. Por lo tanto, la probabilidad de identificación de un usuario es de $\frac{1}{2}$ como máximo. Como se puede apreciar en la tabla, para el enmascaramiento de los datos hemos aplicado generalización en el «código postal», supresión en el '«género» y microagregación univariante en la «edad».

Después del proceso de anonimización, la tabla puede ser publicada con la certeza de que un atacante solo podrá identificar a un usuario con una probabilidad de, como máximo, $\frac{1}{2}$. Es importante subrayar que en los casos reales, donde los conjuntos de datos son muy grandes (con miles o millones de regis-

Tabla 10. Datos médicos protegidos que cumple con el modelo de k-anonimidad con valor $k = 2$

ID	Código postal	Género	Edad	Enfermedad
1	08***	Hombre	25	Cáncer
2	17***	Mujer	46,5	Hepatitis
3	08***	Hombre	25	Gripe
4	25***	*	39,5	Cáncer
5	17***	Mujer	46,5	Diabetes
6	08***	Mujer	65	Gripe
7	08***	Mujer	65	Hepatitis
8	43***	*	23,5	Cáncer
9	43***	*	23,5	Diabetes
10	25***	*	39,5	Gripe

tros) se debe escoger un valor de k mucho más grande que en el ejemplo que acabamos de presentar, donde la probabilidad de identificar a un individuo será mucho más baja, es decir, $\frac{1}{k} << 1$.

3.3. Privacidad diferencial en tablas

La privacidad diferencial asume la presencia de un actor de confianza que contiene el conjunto de datos, recibe consultas enviadas por los usuarios y devuelve resultados para estas consultas.

Supongamos que existe una base de datos con información sobre el salario medio de una población o zona geográfica específica. Además, sabemos que Pedro se traslada a esta zona en breve, con lo cual un atacante puede realizar la consulta antes y después de que sea incorporado el registro con la información de Pedro. La privacidad diferencial previene la obtención de información relacionada con el salario de este individuo, asegurando que nadie pueda probar que el individuo en cuestión esté dentro o fuera de la base de datos.

Consideremos dos bases de datos o tablas: la inicial D_1, sin la información de Pedro, que se puede ver en la tabla 11; y D_2, con la información de Pedro, que se puede ver en la tabla 12.

Tabla 11. Ejemplo de la base de datos inicial (D_1)

ID	Salario (USD)
1	50.000
2	58.000
3	72.000
4	59.000
5	68.000

Como podemos ver, la única diferencia entre D_1 y D_2 es la incorporación de los nuevos datos, es decir, solo difieren en un registro. En este caso, hablaremos de *bases de datos adyacentes*.

Tabla 12. Ejemplo de la base de datos con la información de Pedro (D_2)

ID	Salario (USD)
1	50.000
2	58.000
3	72.000
4	59.000
5	68.000
6	110.000

Para que la base de datos sea diferencialmente privada, necesitamos seleccionar una función aleatoria, un mecanismo M, que agregue ruido a los conjuntos de datos que producirán un resultado R.

Como D_1 y D_2 son adyacentes, la probabilidad de que $M(D_1) = R$ es cercana a la probabilidad de que $M(D_2) = R$. Más formalmente podemos escribir:

$$\frac{P[M(D_1) = R]}{P[M(D_2) = R]} < e^{\epsilon} \tag{2}$$

Para ϵ pequeño, observamos que $e^{\epsilon} \approx 1 + \epsilon$ y, si las probabilidades son idénticas, obtenemos:

$$1 - \epsilon < \frac{P[M(D_1) = R]}{P[M(D_2) = R]} < 1 + \epsilon \tag{3}$$

La cantidad y el tipo de ruido que M agrega está limitado por la sensibilidad global de la función de consulta, f, que se aplicará a los datos. La sensibilidad global se puede escribir:

$$\Delta f = max[f(D_1) - f(D_2)] \tag{4}$$

para todos los posibles conjuntos de datos adyacentes.

Si consideramos una consulta de recuento, entonces $\Delta f = 1$, dado que dos conjuntos de datos adyacentes pueden ser diferentes en un registro como máximo. Dwork demostró que el ruido con una distribución laplaciana mantiene la privacidad diferencial si el valor del ruido laplaciano es ajustado con el parámetro $b = \frac{\Delta f}{\epsilon}$. Por lo tanto, una base de datos en la que se aplica la consulta de recuento será diferencialmente privada si utiliza un mecanismo aleatorio que añade ruido laplaciano con $b = \frac{1}{\epsilon}$.

La siguiente pregunta es qué valor de ϵ debemos seleccionar. Cuanto mayor sea b, más ruido habrá que añadir a la respuesta para lograr la privacidad diferencial. Por lo tanto, un ϵ más pequeño implica más ruido.

A medida que aumentamos Δf (mayor sensibilidad global), necesitamos valores de ϵ más pequeños para proporcionar suficiente ruido. Consideremos una función de consulta que calcula el salario medio. En este caso, la sensibilidad global es igual al salario más alto posible en los conjuntos de datos (es decir, el peor escenario).

Sin embargo, a pesar de su popularidad entre los investigadores y el paso adelante que ofrece en términos de garantías de privacidad, la privacidad diferencial solo se está empleando en un número limitado de aplicaciones del mundo real. La principal razón es la pobre utilidad de los resultados obtenidos mediante modelos diferencialmente privados. A excepción de una serie de aplicaciones de buen comportamiento, la privacidad diferencial tiene un impacto considerable en la utilidad de los datos para ser ampliamente utilizado en el análisis o minería de datos.

A diferencia de los métodos o algoritmos basados en aleatorización o *k*-anonimidad, que se pueden aplicar sobre (casi) cualquier conjunto de datos e independientemente de la tarea a realizar, en el caso de la privacidad diferencial se requiere un método o algoritmo específico para cada tipo de consulta que deseamos realizar a la base de datos. Ejemplos de consultas pueden ser: valores medios de ciertos campos (por ejemplo, salarios) o histogramas de valores (por ejemplo, edades de la población), entre muchos otros.

3.4. Resumen

En este apartado hemos revisado los principales métodos de enmascaramiento de datos tabulares, que permiten modificar los datos originales con el objetivo de impedir o dificultar la identificación de un usuario en los datos protegidos.

- Los métodos perturbativos se basan en la idea de introducir ruido en los datos originales para dificultar el proceso de reindentificación, siendo el ruido aditivo y la microagregación dos de los principales métodos de esta categoría.

- En segundo lugar, hemos presentado los métodos no perturbativos que, en lugar de introducir ruido en los datos, introducen cierto grado de «incertidumbre», ya sea por medio del proceso de generalización o bien de supresión de la información. Es decir, su objetivo es presentar datos menos precisos que impidan la reidentificación de los usuarios, sin introducir, en ningún caso, datos perturbados o falsos en el conjunto de datos anónimos.

- La generación de datos sintéticos permite generar un nuevo conjunto de datos con propiedades similares al conjunto original. Este proceso no es trivial, y en muchas ocasiones requiere cierta información sobre los objetivos de las tareas de minería de datos que se van a aplicar a continuación, con el

fin de ajustar los datos sintéticos para maximizar la utilidad del conjunto de datos anónimo.

A continuación hemos visto cómo algunas de las técnicas perturbativas y no perturbativas permiten aplicar el modelo de *k*-anonimidad en datos tabulares. El modelo de *k*-anonimidad es uno de los modelos más importantes, más estudiados y más utilizados en la preservación de datos tabulares.

Finalmente, hemos introducido el modelo de la privacidad diferencial para datos tabulares. Cabe destacar la diferencia de enfoque con los métodos anteriores, ya que este modelo permite conseguir un alto grado de privacidad en entornos interactivos. Sin embargo, como hemos discutido, ha recibido ciertas críticas sobre la utilidad de los datos anónimos que hace que, hasta el momento, solo sea aplicable a un conjunto reducido de problemas reales.

4. Anonimización de redes y grafos

En los últimos años la representación de datos en formato de red ha experimentado un importante auge en todos los niveles. Este formato permite representar estructuras y realidades más complejas que los tradicionales datos relacionales, que utilizan el formato de tuplas. En un formato semiestructurado cada entidad puede presentar, al igual que los datos relacionales, una serie de atributos en formato numérico, nominal o categórico. Pero además, el formato de red permite representar de un modo más rico las relaciones que puedan existir entre las distintas entidades que forman el conjunto de datos. Un claro ejemplo de esta situación lo presentan las redes sociales.

La literatura utiliza los términos *red* y *grafo* de manera indistinta. Generalmente podemos encontrar referencias a redes o a grafos sin apenas matices, sin diferencias importantes en su significado. En este texto se utilizan los términos red y grafo indistintamente. Aun así, algunos autores apuntan a una sutil diferencia entre ambas terminologías. Por ejemplo, A. Barabási señala que la terminología de red (*network*) se utiliza a menudo para referirse a sistemas reales, mientras que la terminología de grafo (*graph*) se utiliza generalmente para referirnos a las representaciones matemáticas de las redes.

Lectura complementaria

A. Barabási; M. Pósfai (2016). *Network science*. Cambridge: Cambridge University Press.

4.1. Introducción a la teoría de redes y grafos

En este subapartado introduciremos la definición y notación básica de la teoría de grafos. Los grafos son la forma más natural de representación de las redes reales, y es en este sentido en el que necesitamos introducir los conceptos básicos para poder representar las redes reales.

Un grafo es una pareja de conjuntos $G = (V,E)$, donde $V = \{v_1,v_2...,v_n\}$ es el conjunto de nodos o vértices y $E = \{e_1,e_2,\ldots,e_m\}$ es un conjunto de aristas que unen dos nodos $e_i = \{v_i,v_j\}$ de forma bidireccional, es decir, el nodo v_i está conectado al nodo v_j y viceversa. En este caso, hablamos de **grafos no dirigidos**, **bidireccionales** o **simétricos**.

Cuando las relaciones no son bidireccionales, hablaremos de **grafos dirigidos**, **unidireccionales** o **asimétricos**. En este caso, se representa el grafo como pareja de conjuntos $G = (V,A)$, donde es el conjunto de nodos o vértices, igual que en el caso anterior, y $A = \{a_1,a_2,\ldots,a_m\}$ es un conjunto de arcos que unen dos nodos $a_i = \{v_i,v_j\}$ de forma unidireccional, esto es, el nodo v_i está conectado al nodo v_j.

Se llama **orden** de G a su número de nodos, $|V|$, que por convenio es referenciado por la letra n. Asimismo, el número de aristas, $|E|$, es referenciado por la letra m y se le llama **tamaño** del grafo.

Los **nodos adyacentes** o vecinos, denotados como $\Gamma(v_i)$, se definen como el conjunto de nodos unidos a v_i por medio de una arista. En este caso, el **grado** de un nodo se define como el número de nodos adyacentes, es decir, $|\Gamma(v_i)|$, aunque generalmente el grado del vértice v_i se denota como $deg(v_i)$. La **secuencia de grados** (*degree sequence*) es una secuencia numérica de n posiciones en que cada posición i indica el grado del nodo v_i.

En un grafo dirigido G se define a los sucesores de un nodo v_i, $\Gamma(v_i)$, como el conjunto de nodos a los cuales se puede llegar usando un arco desde v_i. Se define el grado exterior de un nodo como el número de sucesores $|\Gamma(v_i)|$. De forma similar, se puede definir a los antecesores de un nodo v_i, $\Gamma^{-1}(v_i)$, como el conjunto de nodos desde los cuales es posible llegar a v_i usando un arco. Se define el grado interior de un nodo como el número de antecesores, es decir, $|\Gamma^{-1}(v_i)|$.

En el caso de las redes sociales, los datos se suelen representar utilizando los grafos, dado que permiten una representación natural de las relaciones existentes entre un conjunto de usuarios (representados mediante nodos o vértices en el contexto de los grafos) de la red. Existen varios formatos de grafo que permiten representar cada una de las redes existentes en la realidad y que se adaptan a las particularidades de cada una de ellas. Por ejemplo, podemos encontrar redes con relaciones simétricas, como por ejemplo Facebook, donde si el usuario A es amigo del usuario B, necesariamente B es amigo de A. Por otro lado, podemos encontrar ejemplos de redes asimétricas, como por ejemplo Twitter, que se representan mediante grafos dirigidos o asimétricos, y donde la relación de «seguir» de un usuario A hacia un usuario B no tiene por qué ser recíproca.

Representación mediante grafos

Vamos a suponer que deseamos modelar las relaciones entre los usuarios de Twitter, mostrando la relación «seguir» que se establece entre dos usuarios de la red. Para representar la información que nos interesa podemos utilizar un grafo dirigido o asimétrico, en donde creamos un arco entre los nodos A y B si el usuario A «sigue» al usuario B. La figura 5 muestra un posible escenario donde podemos ver que los usuarios A, B y C «siguen» al usuario D. Por su parte, el usuario D «sigue» a los usuarios A, E y F.

A continuación veremos cómo modelar las relaciones en una red simétrica o no dirigida, como puede ser, por ejemplo, Facebook. En esta red los usuarios establecen «relaciones de amistad» bidireccionales, es decir, si un usuario A es «amigo» de un usuario B, entonces implícitamente el usuario B también es «amigo» del usuario A. La figura 6 muestra un posible ejemplo en el que vemos las relaciones de amistad entre siete usuarios. Podemos ver que el usuario A es amigo de D, el cual es también amigo de A y de C, E y F.

Figura 5. Grafo dirigido o asimétrico

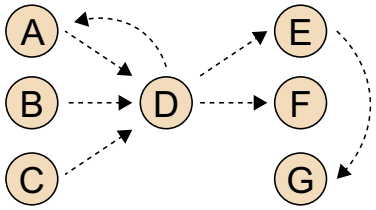
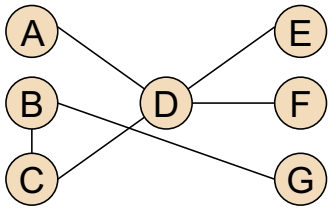


Figura 6. Grafo no dirigido o simétrico



En estos casos, la propia estructura de la red contiene información de gran utilidad para el análisis y el estudio de las redes, pero también puede ser utilizado por un atacante para obtener información e identificar a un usuario dentro de los datos protegidos. Los métodos de enmascaramiento específicos para redes sociales se basan en modificar la estructura de los grafos, ya sea añadiendo o eliminando aristas o vértices del grafo, para introducir ruido en los datos y dificultar el proceso de identificación de usuarios en los datos protegidos.

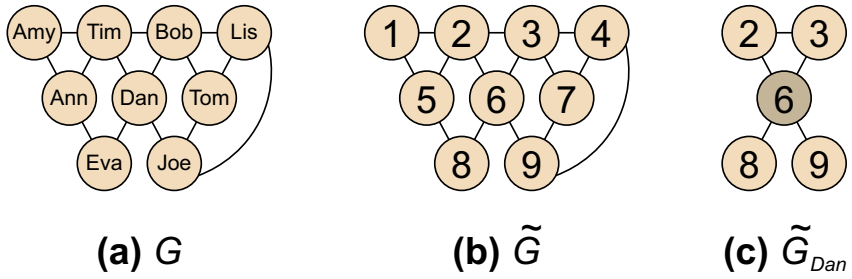
4.2. Definición del problema

Actualmente, se están recopilando grandes cantidades de datos sobre redes sociales, que a menudo contienen información personal y privada de usuarios e individuos. Aunque se realizan procesos básicos de anonimización de datos, como la eliminación de nombres u otros identificadores de claves, la información restante puede ser sensible y útil para que un atacante vuelva a identificar usuarios e individuos dentro del conjunto de datos anónimos.

Para resolver este problema, se han desarrollado métodos que introducen ruido en los datos originales con el fin de obstaculizar los procesos posteriores de reidentificación. Una estrategia natural para proteger la información sensible es reemplazar los atributos de identificación con identificadores sintéticos. Nos referimos a este procedimiento como *anonimización simple* (*naïve anonymization*). Esta práctica trata de proteger la información sensible rompiendo la asociación entre la identidad del mundo real y los datos sensibles.

La figura 7a muestra un ejemplo de una red social, donde cada vértice representa un individuo y cada arista indica la relación de amistad entre dos individuos. La figura 7b presenta el mismo grafo después de un proceso de anonimización simple, donde los identificadores de vértices han sido eliminados y la estructura se mantiene inalterada. En este escenario, un atacante puede romper la privacidad y volver a identificar a un usuario en el grafo anónimo. Por ejemplo, si un atacante sabe que el usuario *Dan* tiene cuatro amigos y que dos de ellos son amigos entre sí, entonces puede construir un subgrafo a distancia uno de *Dan*, representado en la figura 7c. A partir de este subgrafo, el atacante puede reidentificar al usuario *Dan* en el grafo anónimo, dado que es el único nodo con este patrón de subgrafo a distancia uno. Y por consiguiente, puede romper la privacidad del usuario.

Figura 7. Ejemplo de anonimización simple, donde G es el grafo original, $\widetilde{G}$ es la versión anónima de la red y $\widetilde{G}_{Dan}$ representa el subgrafo a distancia uno del usuario Dan



Zhou y Pei mostraron que, para definir el problema de la preservación de la privacidad en la publicación de datos de redes sociales, necesitamos formular las siguientes cuestiones: en primer lugar, necesitamos identificar la información que se debe preservar. En segundo lugar, es necesario modelar el conocimiento que un adversario puede usar para atacar la privacidad de la red. Y, en tercer lugar, necesitamos especificar el uso de los datos publicados para que un método de anonimización pueda intentar retener la utilidad, tanto como sea posible, mientras preserva la privacidad de la información contenida en la red.

Lectura complementaria

B. Zhou; J. Pei (2008). «Preserving Privacy in Social Networks Against Neighborhood Attacks». En: *IEEE International Conference on Data Engineering* (págs. 506-515). Conferencia. Washington D. C.: IEEE Computer Society.

Respecto a la información que se debe preservar en las redes sociales, se han identificado tres categorías principales de amenazas a la privacidad:

1) La **divulgación de la identidad** (*identity disclosure*) ocurre cuando se revela la identidad de un individuo asociado con un vértice del grafo anónimo.

2) La **divulgación de los atributos** (*attribute disclosure*) no busca identificar necesariamente un vértice, sino revelar atributos o datos sensibles del vértice. Los datos sensibles asociados a cada vértice se ven comprometidos.

3) La **divulgación de las relaciones** (*link disclosure*) ocurre cuando se revela la relación sensible entre dos individuos.

La divulgación de identidad y la divulgación de las relaciones se aplican a todo tipo de grafos. Sin embargo, la divulgación de los atributos solo se aplica a las redes que contienen información asociada a los vértices. La revelación de la identidad a menudo conduce a la divulgación de los atributos, debido a que la revelación de identidad ocurre cuando un individuo se reidentifica dentro de un conjunto de datos, mientras que la divulgación de los atributos ocurre cuando se identifica información sensible que un individuo desea mantener privada.

Determinar el conocimiento del adversario es el problema clave. Se ha propuesto un amplio abanico de conocimientos de los adversarios, en conjunción con su correspondiente ataque y método de protección. En el criptoanálisis, los autores distinguen entre dos tipos básicos de ataques, que puede ser también una clasificación válida para los ataques a redes sociales:

- **Ataques activos**, donde un el adversario intenta comprometer la privacidad mediante la creación estratégica de nuevas cuentas de usuario y vínculos con otros usuarios antes de que se publique la versión anónima en la red, de modo que estos nuevos vértices y aristas se encuentren presentes en la versión anónima.

- **Ataques pasivos**, que son llevados a cabo por individuos que tratan de descubrir las identidades de los vértices solo después de que la red anónima haya sido publicada.

4.3. Métodos de anonimización

Existen tres grandes familias de técnicas de anonimización en grafos que se basan en la modificación de la estructura para preservar la privacidad de los datos de una red. Estos son:

- **Modificación de aristas y vértices**: estas técnicas transforman el grafo mediante modificaciones de aristas o vértices (añadiendo o eliminando) y luego publican los datos perturbados. Los datos se ponen así a disposición para cualquier tipo de análisis, sin restricciones.

- **Grafos inciertos** (*uncertain graphs*): este enfoque está basado en la adición o eliminación de aristas de forma «parcial», asignando una probabilidad de existir a cada arista de la red anónima. En lugar de crear o eliminar aristas, se considera el conjunto de todas las aristas posibles y se asigna una probabilidad de existir a cada una de ellas.

- **Métodos de generalización** (*generalization*): estos métodos buscan vértices similares y los agrupan en particiones, de forma que los detalles sobre los individuos quedan ocultos. El principal problema es que la utilidad

del grafo puede reducirse considerablemente después del proceso de anonimización, especialmente cuando el objetivo es preservar las estructuras locales de la red.

4.3.1. Métodos basados en la modificación de aristas o vértices

Los enfoques de modificación de aristas o vértices anonimizan un grafo por medio de modificaciones (es decir, añadiendo o eliminando) de aristas o vértices en el grafo. Estas modificaciones pueden hacerse de forma aleatoria, y este caso nos referiremos a ellos como *métodos aleatorios* (*randomization*), *perturbativos* (*random perturbation*) o *técnicas de ofuscación aleatoria* (*obfuscation*). Sin embargo, la modificación se puede realizar con el fin de cumplir con algunas restricciones deseadas, como en el caso de los métodos basados en la *k*-anonimidad.

Definimos tres procesos básicos de modificación de aristas para cambiar la estructura de la red, agregando y/o eliminando aristas. Estos métodos son los más básicos, y se pueden mezclar para crear combinaciones complejas. Nos permiten modelar, de forma general y conceptual, la mayoría de los métodos de preservación de la privacidad. A continuación introduciremos estos métodos básicos, que se ilustran en la figura 8. Las líneas de trazos representan aristas existentes que se eliminarán, mientras que las líneas continuas constituirán las nuevas aristas. El color del nodo indica si cambia su grado (gris oscuro) o no (gris claro) después del proceso de modificación. Estos procesos son:

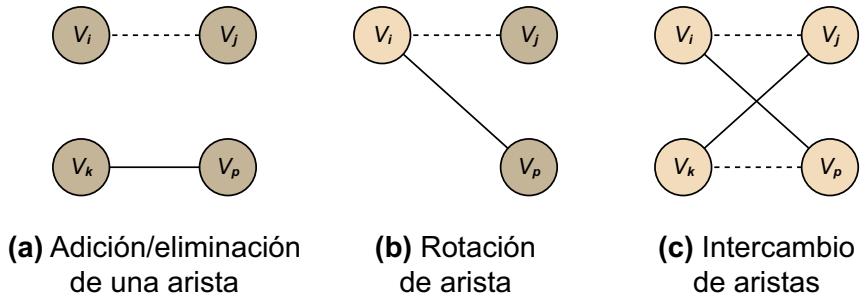
- **Adición/eliminación de una arista** (*edge add/del*). Es la modificación de arista más básica. Simplemente consiste en eliminar una arista existente $\{v_i,v_j\} \in E$ y añadir otra nueva $\{v_k,v_p\} \notin E$. La figura 8a ilustra este proceso.

- **Rotación de arista** (*edge rotation*). Esta ocurre entre tres nodos $v_i,v_j,v_p \in V$ tal que $\{v_i,v_j\} \in E$ y $\{v_i,v_p\} \notin E$. Se define como la supresión de la arista $\{v_i,v_j\}$ y la creación de la nueva arista $\{v_i,v_p\}$, tal y como se ilustra en la figura 8b.

- **Intercambio de aristas** (*edge switch*). En este caso, la operación se produce entre cuatro nodos $v_i,v_j,v_k,v_p \in V$ donde $\{v_i,v_j\},\{v_k,v_p\} \in E$ y $\{v_i,V_p\},\{v_k,v_j\} \notin E$. Se define como la eliminación de las aristas $\{v_i,v_j\}$ y $\{v_k,v_p\}$ y la creación de las aristas $\{v_i,v_p\}$ y $\{v_k,v_j\}$, como se puede ver en la figura 8c.

Para las tres operaciones presentadas, el número de nodos y aristas se mantiene inalterable, pero la distribución de los grados cambia en el caso de la rotación de aristas. Claramente, la adición/eliminación de aristas es el concepto más general y todas las otras perturbaciones pueden ser modeladas como un caso particular de la misma.

Figura 8. Operaciones básicas de modificación de aristas



Métodos aleatorios

Estos métodos se basan en la introducción de ruido aleatorio en los datos originales y han sido ampliamente investigados para los datos estructurados o relacionales.

Los enfoques basados en la introducción de ruido aleatorio, en general, protegen contra la reidentificación de una manera probabilística. Específicamente, los métodos basados en la adición/eliminación o rotación de aristas preservan la identidad de los usuarios (*identity disclosure*), cuando se supone que el conocimiento del adversario se basa en la información del grado o de los vértices vecinos del objetivo, y también contra la divulgación de las relaciones (*link disclosure*). Los métodos basados en el intercambio de aristas, sin embargo, no protegen contra la revelación de la identidad cuando un adversario tiene conocimiento sobre el grado de los vértices, dado que la secuencia de grados del grafo anónimo sigue siendo la misma.

Posiblemente, las dos estrategias más básicas de perturbación aleatoria en grafos son:

- Adición/eliminación aleatoria (*rand add/del*). Esta técnica aplica, reiteradamente, la operación de adición/eliminación de aristas al azar, considerando todo el conjunto de aristas, sin restricciones. Esta estrategia conserva el número de aristas en el grafo anónimo.

- Intercambio aleatorio (*rand switch*). Esta estrategia cambia aleatoria y reiteradamente pares de aristas existentes siguiendo la descripción del intercambio de aristas. En este caso, se preserva el número de aristas del grafo, pero también el grado de todos los vértices.

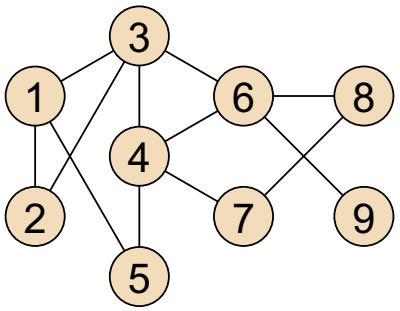
Ejemplo de un proceso de anonimización aleatorio

La figura 9 muestra un ejemplo de anonimización aleatoria, donde la red original se presenta en la figura 9a. A continuación, la figura 9b muestra una versión perturbada de la misma red utilizando el método de adición/eliminación aleatoria (*rand add/del*). Durante el proceso de anonimización, se han eliminado dos aristas ({1,5} y {2,3}) y

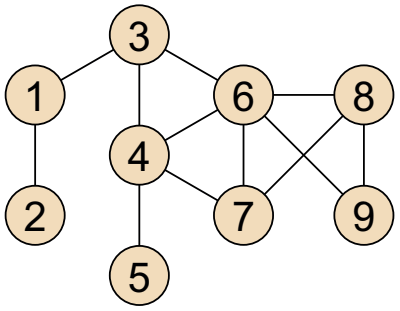
se han creado dos nuevas aristas ({6,7} y {8,9}). Una versión alternativa, en este caso empleando el método de intercambio aleatorio (*rand switch*), se presenta en la figura 9c, donde las aristas {1,2} y {4,5} han sido intercambiados, dando como resultado las aristas {1,4} y {2,5}.

Ambos métodos preservan el número de vértices y aristas. Además, el método de intercambio aleatorio también preserva la secuencia de grados, es decir, $d(G) = d(\widetilde{G}_{sw}) = \{3,2,4,4,2,4,2,2,1\}$, mientras que el método de adición/eliminación aleatorio no tiene esta propiedad, esto es $d(\widetilde{G}_{ra}) = \{2,1,3,4,1,5,3,3,2\}$.

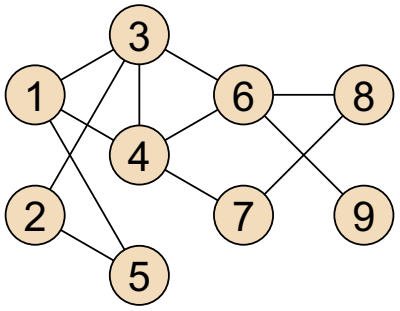
Figura 9. Ejemplo de perturbación aleatoria, donde G es el grafo original, $\widetilde{G}_{ra}$ y $\widetilde{G}_{sw}$ son las versiones anonimizadas empleando el método de adición/eliminación aleatoria (*rand add/del*) y el intercambio aleatorio (*rand switch*), respectivamente



(a) G



(b) $\widetilde{G}_{ad}$



(c) $\widetilde{G}_{sw}$

Hay y otros proponen un método llamado *random perturbation* para anonimizar los grafos simétricos usando la estrategia de adición/eliminación de aristas, eliminando aleatoriamente p aristas y luego añadiendo aleatoriamente p aristas falsas no existentes en el grafo. Las principales ventajas de este método son su simplicidad y su baja complejidad. Por el contrario, los usuarios con un número de vecinos muy superior a la media (*hubs*) no están bien protegidos y pueden ser reidentificados dentro del grafo anónimo.

Lectura complementaria

M. Hay; G. Miklau; D. Jensen; P. Weis; S. Srivastava (2007). *Anonymizing Social Networks*. Amherst: University of Massachusetts.

Otros trabajos posteriores intentan aplicar los métodos anteriormente comentados, pero en lugar de aplicarlos de forma aleatoria entre todas las aristas del grafo original, tratan de preservar las aristas que contribuyen a preservar la utilidad del grafo anónimo. En este sentido, los autores Ying y Wu desarrollaron dos métodos diseñados específicamente para preservar las características espectrales del grafo original, llamados *spctr add/del* y *spctr switch*. En la misma línea, Casas-Roma propone una estrategia que tiene como objetivo preservar las aristas más importantes de la red, tratando de maximizar la utilidad de los datos al tiempo que se logra el nivel de privacidad deseado.

Lectura complementaria

X. Ying; X. Wu (2008). *Randomizing Social Networks: a Spectrum Preserving Approach*. En: *SIAM Conference on Data Mining* (págs. 739-750). Conferencia. Atlanta: SIAM.

Lectura complementaria

J. Casas-Roma (2014). *Privacy-Preserving on Graphs Using Randomization and Edge-Relevance*. En: V. Torra (ed.). *International Conference on Modeling Decisions for Artificial Intelligence* (págs. 204-216). Conferencia. Tokio: Springer International Publishing Switzerland.

k-Anonimidad y derivados

Otra estrategia ampliamente adoptada de los enfoques de modificación de aristas y vértices tiene como objetivo cumplir con determinadas restricciones de privacidad. Es decir, las modificaciones no se realizan de forma aleatoria, sino que, al contrario, se busca realizar las mínimas modificaciones que permitan cumplir con las restricciones de privacidad deseadas. Probablemente, el modelo de *k*-anonimidad es el más conocido y utilizado en este grupo, aunque recientemente se han desarrollado otros modelos y extensiones.

El modelo de *k*-anonimidad, como se ha descrito en el subapartado 2.3, indica que un atacante no podrá distinguir a un usuario entre un grupo de k usuarios y, en consecuencia, no puede reidentificar a ningún individuo con una probabilidad mayor que $\frac{1}{k}$.

Se pueden emplear distintos conceptos como cuasi identificadores para aplicar el modelo de *k*-anonimidad en redes o grafos. Una opción ampliamente utilizada es usar el grado de los vértices como un cuasi identificador, entendiendo que es relativamente fácil conocer el número de «amigos» o «seguidores» que un usuario en concreto tiene en una red social. En consecuencia, suponemos que el atacante conoce el grado de algunos vértices objetivo. Si el atacante identifica un vértice único con el mismo grado en el gráfico anónimo, entonces ha reidentificado este vértice. Es decir, $deg(v_i) \neq deg_(v_j)\ \forall j \neq i$.

Lectura complementaria

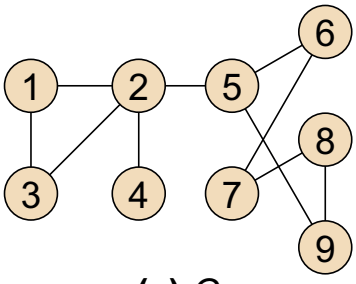
K. Liu; E. Terzi (2008). *Towards identity anonymization on graphs*. En: *ACM SIGMOD International Conference on Management of Data* (págs. 93-106). Conferencia. Nueva York: ACM Press.

Este modelo se llama *k*-anonimidad basada en el grado (*k-degree anonymity*) y fue introducida por Liu y Terzi. Por lo tanto, estos métodos se basan en la modificación de la estructura del grado (mediante modificaciones de aris-

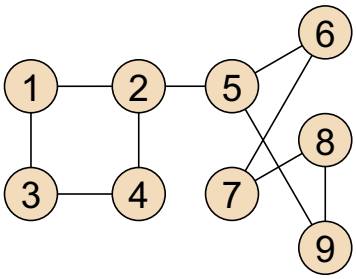
tas o vértices) para asegurar que todos los vértices satisfacen el modelo de *k*-anonimidad cuando se considera el grado como un cuasi identificador. En otras palabras, el objetivo principal es que todos los vértices tengan, al menos, $k-1$ otros vértices compartiendo el mismo grado.

Una red $G = (V,E)$ es *k*-anónima en el grado si, y solo si, todos sus vértices cumplen la propiedad de la *k*-anonimidad basada en el grado.

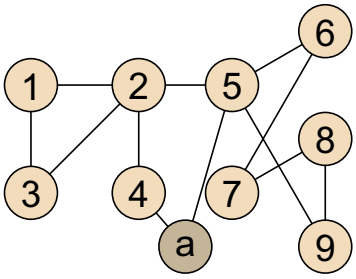
Figura 10. Ejemplo de *k*-anonimidad, donde G representa el grafo original, $\widetilde{G}_{em}$ y $\widetilde{G}_{va}$ son versiones 2-anónimas basadas en el grado a partir de modificaciones en las aristas y los vértices, respectivamente



(a) G



(b) $\widetilde{G}_{em}$



(c) $\widetilde{G}_{va}$

Ejemplo de *k*-anonimidad basada en el grado

Un ejemplo de *k*-anonimización basada en el grado se ilustra en la figura 10. La red original G, representada en la figura 10a, es $k = 1$ anónima respecto al grado, dado que su secuencia de grados es $d(G) = \{2,4,2,1,3,2,2,2,2\}$.

Un ejemplo de una red anónima con $k = 2$ se presenta en la figura 10b. En este caso, se ha empleado la modificación de aristas para cumplir con el modelo de *k*-anonimidad basada en el grado. Por lo tanto, el número de vértices es el mismo, es decir $\widetilde{n} = n$, y la perturbación se logra añadiendo y eliminando aristas. Su secuencia de grados es $d(\widetilde{G}_{em}) = \{2,3,2,2,3,2,2,2,2\}$. Por consiguiente, es una secuencia $k = 2$ anónima debido al hecho de que cada valor de grado aparece, al menos, dos veces en la secuencia de grados.

Alternativamente, podemos ver otra red $k = 2$ anónima basada en el grado en la figura 10c, pero en este caso se ha empleado la adición de vértices. Como se muestra, la estructura original permanece igual, pero se ha agregado un nuevo vértice (gris oscuro) y se han creado dos aristas $\{a,4\}$ y $\{a,5\}$ para cumplir con la propiedad de 2-anonimidad basada en el grado. Su secuencia de grados es $d(\widetilde{G}_{va}) = \{2,4,2,2,4,2,2,2,2,2\}$. Usando este modelo, el número de vértices se incrementa en uno ($\widetilde{n} = n+1$) y el número de aristas en dos ($\widetilde{m} = m+2$).

Los propios autores que plantearon el concepto de *k*-anonimidad basada en el grado, Liu y Terzi, propusieron un método basado en la programación lineal entera y en el intercambio de aristas para construir un nuevo grafo *k*-anónimo basado en el grado, manteniendo un número de vértices y tratando de minimizar el número de aristas modificadas durante el proceso, esto es $V = \widetilde{V}$ y $E \cap \widetilde{E} \approx E$. El trabajo de Liu y Terzi inspiró a muchos otros autores que mejoraron este trabajo seminal, tanto en términos de velocidad como de escalabilidad (permitiendo abordar conjuntos de datos más grandes), utilizando diferentes tipos de heurísticas para la selección de las aristas a añadir o eliminar en el proceso de construcción del grafo anónimo.

Lectura complementaria

S. Chester; B. M. Kapron; G. Ramesh; G. Srivastava; A. Thomo; S. Venkatesh (2013). «Why Waldo befriended the dummy? *k*-Anonymization of social networks with pseudo-nodes». *Social Network Analysis and Mining* (vol. 3(3), págs. 381-399).

Los autores Chester y otros plantearon el mismo problema, pero permitiendo modificaciones al conjunto de vértices, en lugar de solo al conjunto de aristas, y esto ofrece algunas diferencias con respecto a la utilidad del grafo anónimo. En este trabajo crearon nuevas aristas entre vértices falsos y reales o únicamente entre vértices falsos. Sin embargo, los resultados mostraron que la pérdida de información aumenta cuando se utiliza simultáneamente la adición de nuevos vértices falsos y la creación de aristas.

El modelo de la *k*-anonimidad basada en el grado ha sido ampliamente estudiado, pero también se han considerado otros modelos derivados que incrementan el nivel de privacidad, normalmente incrementando también el nivel de complejidad de cálculo del grafo anónimo, y en consecuencia dificultan la aplicación de estos métodos en grafos de gran tamaño (cientos de miles o millones de vértices y aristas).

En lugar de usar el grado de vértice como cuasi identificador, otros trabajos consideran el subgrafo de vecindad a distancia uno de los vértices objetivos como cuasi identificador. Para un vértice $v_i \in V$, v_i es *k*-anónimo en G si hay al menos otros $k-1$ vértices $v_1,\ldots,v_{k-1} \in V$ tal que $\Gamma(v_i),\Gamma(v_1),\ldots,\Gamma(v_{k-1})$ son isomorfos. Entonces, G se considera *k*-anónimo basado en la vecindad a dis-

tancia uno si todos los vértices de G son k-anónimos considerando los vecinos a distancia uno del vértice objetivo.

Otros autores han modelado un conocimiento del adversario más complejo y, en consecuencia, han creado modelos de k-anonimidad basados en cuasi identificadores que implican un conocimiento mayor por parte del adversario. Por ejemplo, Hay y otros propusieron un método llamado k-anonimidad basado en un conjunto de candidatos. En este método, un vértice v_i es k-anónimo respecto a una pregunta Q si existen, al menos, otros $k-1$ vértices en el grafo con la misma respuesta a la pregunta Q. Formalmente, $|cand_Q(v_i)| \geq k$ donde $cand_Q(v_i) = \{v_j \in V : Q(v_i) = Q(v_j)\}$. Un grafo es k-anónimo respecto a una pregunta Q, si todos sus vértices son k-anónimos respecto a la pregunta Q. Este modelo es más general que los vistos anteriormente, y permite modelar el conocimiento del adversario con una consulta Q, que puede referirse a cualquier propiedad estructural de la red o grafo.

Lectura complementaria

M. Hay; G. Miklau; D. Jensen; D. Towsley; P. Weis (2008). «Resisting structural re-identification in anonymized social networks». *Proceedings of the VLDB Endowment* (vol. 1(1), págs. 102-114).

4.3.2. Grafos inciertos

En lugar de anonimizar los grafos añadiendo o eliminando aristas y vértices para satisfacer determinados parámetros de privacidad, métodos recientes han aprovechado la semántica de los grafos inciertos (*uncertain graphs*) para conseguir la protección de la privacidad deseada. Consideremos el grafo $G = (V,E)$ y denotemos V_2 como el conjunto de todos los pares de vértices $\binom{n}{2}$ desordenados de V, es decir $V_2 = \{(V_i,v_j) : 1 \leq i < j \leq n\}$. Un grafo incierto es un par $\widetilde{G} = (V,p)$, donde $p : V_2 \rightarrow [0,1]$ es una función que asigna las probabilidades existentes a todos las aristas posibles. Estas técnicas anonimizan un gráfico determinístico, convirtiéndolo en una forma incierta donde todas las posibles aristas existen con una cierta probabilidad en el rango [0,1].

Ejemplo de anonimización mediante un grafo incierto

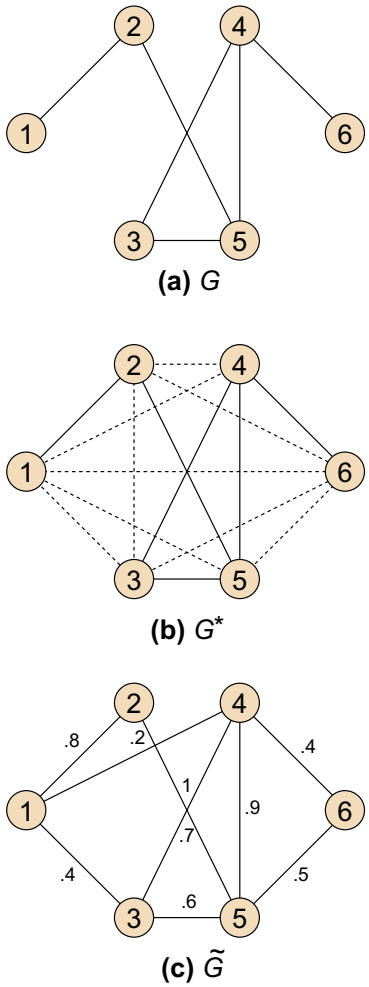
La figura 11 muestra el proceso de anonimización bajo el modelo de grafo incierto. El grafo original G está representado en la figura 11a, y la versión incierta del mismo grafo se muestra en la figura 11b. Como puede verse, hay todas las aristas posibles, es decir, $\binom{6}{2}$, y cada una está asignada a una probabilidad igual a uno (líneas negras) o cero (líneas punteadas). Así, G^* es la representación de G bajo el modelo de grafo incierto, pero no es perturbado ni anonimizado. La versión anónima se presenta en la figura 11c, donde la probabilidad de cada arista se perturba y se establece en el rango [0,1]. Las aristas con probabilidad igual a cero no se representan en $\widetilde{G}$ para preservar una visualización clara del grafo incierto perturbado.

El primer enfoque fue propuesto por Boldi y otros y se basa en inyectar incertidumbre en las redes originales y publicar los grafos inciertos resultantes. Desde una perspectiva probabilística, agregar una arista no existente $\{v_i,v_j\}$ corresponde a cambiar su probabilidad $p(\{v_i,v_j\})$ de cero a uno, mientras que eliminar una arista corresponde a cambiar su probabilidad de uno a cero.

Lectura complementaria

P. Boldi; F. Bonchi; A. Gionis; T. Tassa (2012). «Injecting Uncertainty in Graphs for Identity Obfuscation». *Proceedings of the VLDB Endowment* (vol. 5(11), págs. 1376-1387).

Figura 11. Ejemplo de anonimización de un grafo incierto, donde G representa la red original. El grafo incierto es G^*, donde las aristas existentes tienen asociada una probabilidad igual a uno (líneas negras) y las inexistentes una probabilidad igual a cero (líneas punteadas). $\widetilde{G}$ representa un posible grafo incierto después del proceso de anonimización



(a) G

(b) G^*

(c) $\widetilde{G}$

En este método, en lugar de considerar solo las probabilidades binarias de las aristas, permiten que las probabilidades tomen cualquier valor en el rango [0,1]. Por lo tanto, cada arista está asociada a una probabilidad específica en el gráfico incierto. Para mantener la utilidad del grafo anónimo, el método inyecta incertidumbre (o ruido) solo en un pequeño subconjunto de pares de vértices E_c, y supone que no existen otros pares de vértices, es decir $p(v_i,v_j) = 0\ \forall (v_i,v_j) \notin E_c$.

Un grafo incierto es (k,ϵ)-ofuscado respecto a la propiedad P si la entropía de la distribución $Y_{P(v)}$ sobre al menos $(1-\epsilon)n$ vértices de $\widetilde{G}$ es mayor o igual que $log_2(k)$, es decir, $H(Y_{P(v)}) \geq log_2(k)$.

Es importante remarcar que las estadísticas y las métricas deben ser definidas (o redefinidas) para ser aplicadas en este tipo de grafos, ya que casi todas fueron diseñadas para trabajar con grafos binarios y no pueden aplicarse directamente en este tipo de grafos inciertos. En esta dirección, el cálculo de las

estadísticas basadas en el grado, como por ejemplo el número de aristas, el grado medio, el grado máximo y la varianza de grado fueron propuestas en el trabajo de los mismos autores.

4.3.3. Métodos de generalización

Los métodos de generalización, también conocidos como *enfoques basados en clústeres* (*generalization* o *clustering-based approaches*), se basan, esencialmente, en agrupar vértices y aristas en particiones llamadas *supervértices* y *superaristas*. Los detalles sobre los individuos pueden ocultarse adecuadamente, pero el grafo queda reducido considerablemente después del proceso de anonimización, lo cual puede no ser deseable para analizar las estructuras locales. El grafo generalizado, que contiene las estructuras de enlace entre las particiones, así como la descripción agregada de cada partición, todavía se puede utilizar para estudiar las macropropiedades de la red original. Aun si contiene las propiedades del grafo original, no tiene la misma granularidad y escala. Los métodos de generalización ofrecen un buen nivel de anonimato, pero al precio de disminuir considerablemente la utilidad del grafo anónimo.

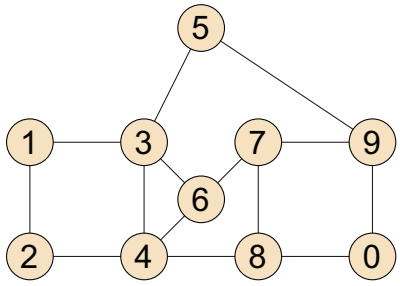
Los procesos de generalización reducen el tamaño del grafo, tanto en el número de vértices como en el número de las aristas. En esencia estos métodos producen un resumen de la red original, que también puede ser útil para reducir el tiempo de cálculo en los procesos de minería de grafos posteriores. Sin embargo, todos los métodos desarrollados hasta ahora necesitan disponer del grafo completo para poder analizarlo y decidir las particiones de vértices que va a realizar en el proceso de generalización. En consecuencia, no son capaces de tratar con grandes volúmenes de datos o con datos en tiempo real (*streaming data*).

Como todos los métodos antes mencionados, los enfoques de generalización también protegen contra la revelación de identidad. Por otra parte, es interesante subrayar que los enfoques de generalización también preservan contra la divulgación de atributo y de relación, ya que dos vértices de cualquier grupo son indistinguibles en función de sus relaciones o sus atributos.

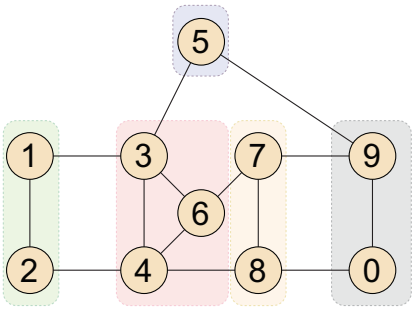
Ejemplo de generalización

La figura 12 presenta un ejemplo de proceso de anonimización basado en el método de generalización, donde G es la red original. En primer lugar, estos métodos calculan una partición de todo el conjunto de vértices, como por ejemplo la partición que se muestra en la figura 12b, donde todos los vértices has sido asignados a uno de los cinco clústeres. Este es el proceso más importante, ya que agrupa vértices con características similares, conduce el proceso de generalización a mejores resultados en términos de utilidad de datos y de pérdida de información. En segundo lugar, una vez que se crean las particiones, estos métodos agrupan todos los vértices de la misma partición en un supervértice y crean superaristas entre ellos. Una versión generalizada de G se puede ver en la figura 12c. Como se muestra, cada supervértice contiene información sobre el número de vértices y aristas internas entre ellos. Generalmente, cada superarista se etiqueta de acuerdo con el número total de aristas entre todos los vértices de cada supervértice.

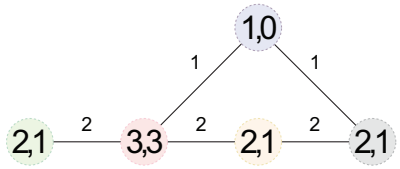
Figura 12. Ejemplo de generalización, donde G representa la red original. Se presenta una partición de muestra del conjunto de vértices y se utiliza para crear un gráfico generalizado $\widetilde{G}$



(a) G



(b) Partición del conjunto de vértices



(c) $\widetilde{G}$

Los autores Hay y otros aplicaron los métodos de generalización utilizando el tamaño de una partición para asegurar el anonimato de los vértices. Su método obtiene un supergrafo *k*-anónimo agrupando nodos en supervértices y aristas en superaristas. Cada supervértice contiene un conjunto mayor o igual a *k* vértices y cada superarista representa todas las aristas entre vértices en dos supervértices. Solo se publica la densidad de las aristas para cada partición, por lo que será difícil distinguir entre individuos dentro de una partición. Los autores evaluaron la efectividad de las consultas estructurales en redes reales de varios dominios y grafos aleatorios. Sus resultados mostraron que las redes son diversas en su resistencia a los ataques: las redes sociales y de comunicación tienden a ser más resistentes de lo que algunos modelos aleatorios podrían sugerir.

Lectura complementaria

M. Hay; G. Miklau; D. Jensen; D. Towsley; P. Weis (2008). «Resisting structural re-identification in anonymized social networks». *Proceedings of the VLDB Endowment* (vol. 1(1), págs. 102-114).

Uno de los métodos más conocidos de generalización fue introducido por Campan y Truta y es aplicable a redes no dirigidas con vértices etiquetados y aristas no etiquetados. Los atributos de los vértices pueden ser de tres tipos principales: identificadores, cuasi identificadores y atributos sensibles. Aplicamos el modelo de *k*-anonimidad a los cuasi identificadores para conseguir vértices indistinguibles en sus atributos o relaciones entre ellos. Los autores desarrollaron un nuevo método, llamado SaNGreeA, diseñado para anonimizar la información estructural. Agrupa vértices en varios grupos y, a continuación, se asigna una etiqueta para cada partición con información resumida (como el número de nodos en la partición).

Lectura complementaria

A. Campan; T. M. Truta (2009). «Data and Structural *k*-Anonymity in Social Networks». En: F. Bonchi; E. Ferrari; W. Jiang; B. Malin (eds.). *Privacy, Security, and Trust in KDD* (págs. 33-54). Berlín: Springer-Verlag.

4.4. Resumen

En este apartado hemos revisado la problemática de la publicación de datos en formato semiestructurado, o formato de red o grafo. En este tipo de datos, más allá de los atributos que puede contener cada individuo (representado por un nodo o vértice) y que pueden ser anonimizados mediante las técnicas vistas en el apartado 3, un atacante puede emplear información estructural para reidentificar a los usuarios dentro de la red anónima. Es decir, el número de relaciones y el subgrafo a distancia $d \geq 1$ pueden aportar información muy relevante que el atacante puede emplear para la reidentificación de los usuarios objetivo.

En los últimos años, este ha sido un campo muy activo en la investigación sobre privacidad y utilidad de los datos. Se han desarrollado gran cantidad de modelos basados en la modificación de aristas y vértices, especialmente modelos basados en la *k*-anonimidad, donde se ha buscado reducir la pérdida de información y mantener un nivel de privacidad adecuado. Aun así, algunos de estos modelos presentan un coste computacional muy elevado y no son aplicables a redes de cientos de miles o millones de nodos y aristas.

Los grafos inciertos han sido introducidos recientemente en el contexto de la privacidad. Aunque el modelo presenta interesantes propiedades, también presentan la problemática de que la mayoría de los modelos de minería de datos (concretamente minería de grafos, *graph mining*) no permiten trabajar con este tipo de grafos.

Los modelos de generalización, por otra parte, presentan interesantes propiedades de privacidad, al unir a todos los individuos similares en un único supernodo con la información agregada, pero no permiten análisis relacionados con las estructuras locales de las redes, ya que estas se pierden en el proceso de agrupamiento.

5. Conclusiones

Datos estructurados

Los datos estructurados son aquellos que siguen un patrón igual para todos los elementos y que además es conocido *a priori*. Por ejemplo, los datos de una hoja de cálculo presentan los mismos atributos para cada fila.

Tradicionalmente los datos se han almacenado en formato de tablas. Durante mucho tiempo este ha sido el formato «estándar» de almacenamiento y transferencia de la información. Debido a esto, la preservación de la privacidad se ha centrado durante muchos años en los datos en formato de tablas, también llamados *datos estructurados*, como por ejemplo los datos contenidos en bases de datos relacionales o en hojas de cálculo.

Todos los métodos vistos hasta ahora han sido desarrollados para trabajar con datos estructurados y semiestructurados (de tipo red o grafo), aunque muchos de ellos han sido extendidos para poder trabajar con otros tipos de datos.

En el contexto social actual, se generan datos en multitud de formatos distintos, que van mucho más allá de las tradicionales tablas. Por un lado, es innegable que los datos que generan los mismos ciudadanos en las distintas redes sociales son una fuente de información muy importante para la administración e instituciones de toda índole. De estas redes podemos obtener datos sobre las opiniones o preferencias de los ciudadanos, sus localizaciones en distintos momentos del día y un largo etcétera que crece día a día con la aparición de nuevas redes sociales. Por otro lado, es habitual que todo tipo de administraciones públicas de la ciudad, especialmente en el contexto de las ciudades inteligentes (*smart cities*), generen y publiquen información en distintos formatos: desde documentos censales de los comercios de la ciudad hasta todo tipo de servicios que se ofrecen a los ciudadanos. Por lo tanto, en este contexto es necesario, pero no suficiente, tratar la privacidad de los datos en formato estructurado y semiestructurado (de tipo red o grafo). La diversidad de formatos de los datos irá en aumento en los próximos años, y se deberá tratar de forma específica cada uno de los nuevos formatos, considerando sus características y especificaciones.

A continuación enumeraremos, muy brevemente, otros tipos de datos que por sus características o auge reciente han recibido mucha atención desde el punto de vista de la privacidad.

5.1. Localización y tiempo

En algunos casos los datos llevan asociada información referente a localización y el momento exacto en que se han producido. En estos casos, generalmente, la información relativa a la localización y el momento en que se pro-

duce es muy importante y puede resultar de gran interés para el estudio del comportamiento humano. Aun así, estos datos también pueden suponer una brecha en la privacidad de un usuario y pueden ser utilizados por un atacante para conseguir su identificación. Por ejemplo, supongamos que se publican datos relacionados con la movilidad y la localización de los ciudadanos. La red social Foursquare* es un ejemplo de servicio de localización y movilidad, donde los usuarios marcan (*check-in*) lugares específicos donde se encuentran y los comparten con los demás usuarios. En el conjunto de datos protegidos, los distintos registros de un mismo usuario indican el conjunto de sitios done un usuario ha estado en un determinado intervalo de tiempo. Si un usuario es identificado dentro de un conjunto protegido, el atacante puede obtener información de los lugares en que este ha estado y el momento exacto en que ha estado en cada lugar.

*** http://foursquare.com**

Foursquare

Foursquare es una red social creada en el año 2009 y basada en la localización de los usuarios.

Para preservar la privacidad en este contexto se puede entender la localización como un atributo numérico (por ejemplo, por medio de la coordenadas GPS del punto) y aplicar sobre este dato los métodos de enmascaramiento para datos numéricos, o también entender la localización como un punto dentro de un municipio, comarca y provincia y, por lo tanto, convertir estos datos en atributos categóricos y aplicar el concepto de generalización para enmascarar la localización real del usuario. Una aproximación similar puede ser empleada para tratar con los datos referentes al tiempo.

GPS

El sistema de posicionamiento global o GPS (*Global Positioning System*, en inglés) es un sistema que permite determinar la posición de un objeto facilitando su latitud y longitud.

5.2. Registros de búsqueda y acceso

Cada vez que un usuario realiza una búsqueda mediante un motor de búsqueda de internet, este almacena información sobre el usuario, el momento y el dispositivo desde donde se produce la búsqueda y los términos introducidos. Esta información es muy útil para el estudio del comportamiento humano e incluso para la predicción de epidemias, como el proyecto *Google Flu Trends**.

*** https://www.google.org /flutrends/about/**

Google Flu Trends

Mediante la agregación de registros de búsqueda en el motor de Google, el proyecto intentó intenta hacer predicciones precisas sobre la actividad de la gripe durante los años 2008-2014.

El ejemplo de AOL del que hemos hablado anteriormente demostró que una insuficiente anonimización de los datos puede conducir a un atacante a identificar usuarios dentro de los datos protegidos mediante sus búsquedas. La particularidad de este tipo de datos es que debemos tratar con la semántica de las consultas. Por ejemplo, si se quiere poder aplicar el concepto de generalización, es necesario «entender» el significado de una consulta para poder generalizarla y así evitar las particularidades que pueden conducir a la identificación de un usuario. Pero la semántica de las consultas puede resultar muy compleja y es necesario disponer de ontologías u otras herramientas semánticas para poder establecer equivalencias y jerarquías entre las distintas consultas.

5.3. Documentos

La publicación de documentos también lleva implícitos algunos problemas de privacidad. A diferencia de los casos hasta ahora comentados, cuando se publican documentos estamos trabajando con datos no estructurados, a diferencia de las tablas o registros, que son datos estructurados. Ciertamente, los documentos no presentan patrones de atributos conocidos *a priori*, al contrario, los documentos no presentan ningún tipo de estructura y para su análisis es necesario analizar todo el texto considerando la gramática y la semántica para poder extraer conocimiento del texto. En este sentido, es habitual crear índices que permitan asociar documentos o partes de los mismos a determinadas palabras o conjuntos de palabras.

La preservación de la privacidad en el caso de los documentos debe ser estudiada y analizada desde este doble punto de vista: por un lado el propio documento y, por el otro, los índices creados a partir del conjunto total de documentos que pueden guiar al atacante a la identificación de datos personales privados dentro de los mismos.

Estos son solo algunos ejemplos de las problemáticas que se deberán resolver en los próximos años relacionadas con la privacidad y anonimización de los datos. Los datos son una fuente innegable de información, que puede ser utilizada por las administraciones públicas para mejorar el rendimiento de las ciudades y la vida de sus ciudadanos, así como por parte de cualquier tipo de empresa, que puede optimizar su proceso de marketing y venta. Pero en ningún caso se debe permitir que estas «mejoras» sean al precio de sacrificar la privacidad de los usuarios, que pueden verse discriminados ante ciertas situaciones por los resultados de análisis y procesos de minería de datos.

Resumen

En este módulo didáctico hemos presentado la problemática básica que implica la publicación de datos en relación a la preservación de la privacidad. Hemos visto los modelos básicos de anonimización, esto es, aleatorización, *k*-anonimidad y privacidad diferencial, y también los métodos básicos de enmascaramiento.

Hemos introducido el modelo de *k*-anonimidad, que es probablemente el modelo más utilizado en la actualidad, tanto en el sector industrial como en el sector académico, para anonimizar datos previamente a su publicación. También hemos presentado el modelo de privacidad diferencial, que en los últimos años está atrayendo una gran parte de investigación en el campo de la anonimización y preservación de la privacidad en procesos de publicación de datos.

Durante el contenido de este módulo didáctico hemos profundizado en la problemática y soluciones de los datos estructurados en formato de tabla y los datos semiestructurados en formato de red o grafo. Pero como ya hemos comentado, el abanico de tipologías de datos y casuísticas de problemas no para de aumentar, del mismo modo que aumentan los datos accesibles para cualquier entidad o institución pública o privada. En los próximos años, posiblemente, asistiremos a un importante auge en la investigación de métodos de preservación de la privacidad, que permitan emplear la gran cantidad de datos que se generan actualmente en un entorno que respete la privacidad de los usuarios implicados en los análisis.

Glosario

Atributos identificadores. Conjunto de atributos que permiten identificar de forma explícita a un individuo, como por ejemplo mediante el nombre, DNI o número de la seguridad social.

Atributos cuasi identificadores. Conjunto de atributos que potencialmente podrían identificar a un individuo.

Atributos sensibles. Conjunto de atributos que presentan información específica y sensible de un individuo en concreto, y como tales deben poder ser asociados a un único individuo.

Datos estructurados. Aquellos que siguen un patrón igual para todos los elementos y que además es conocido *a priori*. Por ejemplo, los datos de una hoja de cálculo presentan los mismos atributos para cada fila.

Datos semiestructurados. Forma de datos que no contiene una estructura fija predefinida *a priori*, pero contiene etiquetas u otros marcadores para separar los elementos semánticos y hacer cumplir jerarquías de registros y campos de los datos. Por ejemplo, las documentos JSON o HTML.

Datos no estructurados. Aquellos que no siguen ningún tipo de patrón conocido *a priori*. Por ejemplo, dos documentos de texto o imágenes.

Grafo. Pareja de conjuntos $G = (V,E)$, donde $V = \{v_1,v_2...,v_n\}$ es el conjunto de nodos o vértices y $E = \{e_1,e_2,\ldots,e_m\}$ es un conjunto de aristas o arcos que unen dos nodos $e_i = \{v_i,v_j\}$.

GPS. El sistema de posicionamiento global o GPS (*Global Positioning System*, en inglés) es un sistema que permite determinar la posición de un objeto facilitando su latitud y longitud.

Método perturbativo. Método de enmascaramiento que altera los datos introduciendo cierto ruido o distorsión para dificultar el proceso de identificación de un usuario por parte de un atacante.

Método no perturbativo. Método de enmascaramiento basado en la generalización o el borrado de algunas partes de la información que podrían conducir a un atacante a identificar de forma única a un usuario dentro de los datos protegidos.

PPDP. Proceso de preservación de la privacidad en la publicación de datos (*privacy-preserving data publishing*), que estudia cómo publicar datos, de manera que una vez publicados mantengan su utilidad al mismo tiempo que preserven la privacidad de los usuarios que aparecen en ellos.

Bibliografía

Benjamin, C.; Fung, M.; Wang, K.; Wai-Chee Fu, A.; Yu, P. S. (2011). *Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques*. Estados Unidos de América: CRC Press.

Casas-Roma, J.; Herrera-Joancomartí J.; Torra, V. (2016). «A survey of graph-modication techniques for privacy-preserving on networks». En: *Artificial Intelligence Review* (vol. 47(3), págs. 341-366). DOI: 10.1007/s10462-016-9484-8.

Charu, C.; Yu, P. S. (2008). *Privacy-Preserving Data Mining: Models and Algorithms*. Nueva York: Springer.

Navarro-Arribas, G.; Torra, V. (2015). *Advanced research in data privacy*. Nueva York: Springer.

Torra, V. (2010). *Privacy in Data Mining. Data Mining and Knowledge Discovery Handbook* (págs. 687-716). Nueva York: Springer.

Torra, V.; Navarro-Arribas, G. (2014). «Data privacy». En: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (vol. 4(4), págs. 269-280).