k-Anonimidad para grafos

Klara Stokes
Dept. of Computer Engineering and Maths
Universitat Rovira i Virgili
Email: klara.stokes@urv.cat

***Resumen*—En este trabajo se estudian aspectos de privacidad sobre grafos, estructuras que permiten representar, por ejemplo, las redes sociales. En particular, se estudia la definición de k-anonimidad para grafos y la (k,l)-anonimidad, la cual es una relajación de la primera. Se ha observado que si un grafo es k-anónimo respecto a su vecindario, entonces es k-anónimo respecto a cualquier otra propiedad de los grafos. Esto hace que la definición de k-anonimidad presentada proporcione k-anonimidad en cualquier situación. En el presente trabajo estudiamos las propiedades de estos grafos desde el punto de vista combinatorio.**

I. Introducción

Las relaciones pueden representarse mediante grafos. Encontramos de forma frecuente bases de datos que contienen relaciones, como en el caso, por ejemplo, de la epidemiología y de la sociología. La privacidad de datos es un área de investigación que se focaliza en la protección de bases de datos de manera que los investigadores puedan acceder a ellas sin violar la privacidad de los individuos que han suministrado sus datos. Una aproximación popular para la protección de datos es la k-anonimización. Una tabla es k-anónima si existen k copias de cada registro cuando restringimos la tabla a cualquier colección de atributos (cuasi-identificadores) que hacen posible la reidentificación.

Se ha argumentado que conseguir la k-anonimidad para grafos es más complicado que para tablas tabulares puras (véase por ejemplo [9]), y se han sugerido varios cuasi-identificadores. En este trabajo argumentamos que la k-anonimidad para grafos es un caso particular de la k-anonimidad para datos tabulares, ya que un grafo puede ser representado en forma de tabla a través de la matriz de adyacencia, sin perdida de información. Si podemos k-anonimizar esta matriz, entonces el grafo será k-anónimo en relación a cualquier posible propiedad de grafos que podamos imaginar.

En este trabajo se presenta una caracterización de los grafos k-anónimos desde el punto de vista combinatorio.

Para una revisión del tema de la privacidad de grafos véase, por ejemplo, [10]. Un trabajo relacionado con este es [1], que presenta una visión de la k-anonimidad basada en *clustering*. En esta aproximación el resultado de la protección consiste en una estructura de supervértices y superaristas que resume los *clusters*, previamente construidos por un algoritmo de *clustering*. Se observa que este metodo tiene la desventaja de no proporcionar un grafo como resultado final.

La estructura de este trabajo la siguente. En la sección II se revisa el concepto de k-anonimidad para datos en forma de tabla. En la sección III se introduce el concepto de k-anonimidad para grafos. En la sección IV se presenta la relajación (k,l)-anonimidad y algunas de sus propiedades. En la sección V se introducen algunos resultados sobre la k-anonimidad para grafos dirigidos. El trabajo finaliza con unas conclusiones.

II. k-Anonimidad

Una base de datos (tabla) es una colección de registros que corresponden a individuos o entidades. Un registro se divide en atributos (nombre, número de registro personal, etc.). Como hemos mencionado en la introducción, la privacidad de datos investiga técnicas que permitan la publicación de bases de datos evitando la revelación de información privada ó sensible a terceras personas. El concepto de k-anonimidad ([5], [6], [7], [8]) engloba un conjunto de técnicas que proporcionan privacidad de datos. Después de proteger una base de datos, tenemos k-anonimidad si cada registro se encuentra camuflado entre otros $k-1$.

Formalmente, sea T una tabla con atributos A, y sea $B \subset A$ un conjunto de atributos de la tabla. Denotamos mediante $T[B]$ la proyección de la tabla en los atributos de B. Suponemos que cada registro contiene información sobre un individuo único. Un identificador I en la base de datos es un atributo que determina de forma unívoca a los individuos correspondientes a los registros. En particular, cada entrada en $T[I]$ es única. Un cuasi-identificador QI en la base de datos es una colección de atributos $\{A_1,\ldots,A_n\}$ que pertenecen al dominio público (esto es, que son conocidos por el adversario), tal que su combinación puede identificar de forma única, o casi única, a un registro [2]. Esto significa que la estructura de la tabla permite que una entrada en $T[QI]$ sea única, o casi única. En el primer caso tenemos que la entrada en $T[QI]$ identifica de forma única al individuo correspondiente al registro, y en el segundo caso existe la posibilidad de que un número pequeño de individuos se coaliguen para hacer la identificación posible.

Definición 1. *Una tabla T que representa una base de datos y que tiene como cuasi-identificador QI es k-anónima si todas las secuencias en $T[QI]$ aparecen como mínimo k veces en $T[QI]$.*

III. k-Anonimidad para grafos

Un grafo es un par (V,E) donde V es un conjunto de vértices y E es un conjunto de aristas que relacionan los vértices dos a dos. Por lo tanto, E es una familia de 2-subconjuntos de V. Un camino en un grafo es una secuencia de vértices en V tal que existen aristas en E que unen los pares de vértices consecutivos. Una componente conexa en un grafo es un subconjunto de vértices $U \subseteq V$ de manera que entre cada par de vértices en U existe un camino. Todo grafo es unión de sus componentes conexas. Una componente conexa de un solo vértice es un vértice que está desconectado del resto del grafo. Un vértice de estas características puede disfrutar de k-anonimidad sólo si existen $k-1$ otras componentes conexas de un sólo vértice. Para vértices que se encuentran en una componente conexa que consiste en más de un vértice, es fácil ver que su anonimidad sólo depende de su propia componente conexa. Por lo tanto, en este trabajo suponemos que los grafos tienen al menos dos vértices y que son de una sola componente conexa, es decir, que son conexos.

El concepto de k-anonimidad se introdujo para tablas. Para su aplicación a otras estructuras de datos, es necesario representar éstas también de forma tabular. Por ejemplo, si queremos aplicar la k-anonimidad a los grafos, la matriz de adyacencia es una representación del grafo en forma de tabla. Recordemos que la matriz de adyacencia de un grafo es una matriz en la que las filas y las columnas se indexan por los vértices del grafo y que las entradas en la tabla representan el número de aristas entre los vértices. En este caso, obtenemos la tabla tomando las filas de la matriz como los registros, y las columnas como los atributos de la tabla.

En algunas aplicaciones podría parecer necesario incluir en la tabla que representa el grafo otra información adicional como el grado de los vértices. Sin embargo, es importante subrayar que la información en la matriz de adyacencia es suficiente para deducir cualquier otra información sobre el grafo. A continuación presentamos la definición de k-anonimidad que consideramos apropiada para grafos.

Sea (V, E) un grafo y sea v un vértice de V. Definimos los vecinos de v como el conjunto de vértices a distancia uno de v, esto es, el conjunto de vértices

$$N(v) := \{u \in V : (u, v) \in E\}.$$

Entonces el grado de v es $d(v) = |N(v)|$. Utilizando $N(v)$ definimos un grafo k-anónimo como sigue.

Definición 2. *Sea $G(V, E)$ un grafo. Decimos que G es un grafo k-anónimo si para cada vértice v en V existen como mínimo $k-1$ otros vértices distintos $\{v_i\}_{i=1}^{k-1}$ de V tales que $N(v_i) = N(v)$ para todo $i \in \{1, \ldots, k-1\}$.*

Esta definición de k-anonimidad es apropiada para un propietario de datos que ha determinado que el cuasi-identificador del grafo son los conjuntos de vecinos de los vértices. Un grafo que es k-anónimo de acuerdo con esta definición tiene una matriz de adyacencia en la que cada fila aparece k veces. Como hemos dicho anteriormente, la matriz de adyacencia es una representación del grafo sin pérdida de información en el sentido de que determina de forma completa el grafo. Por tanto, se puede deducir que un grafo k-anónimo de acuerdo con la definición 2 es k-anónimo respecto a cualquier otra propiedad del grafo puesto que estas propiedades estan implícitas en la matriz. En este tipo de grafo existe una partición del conjunto de vértices tal que los vértices de una misma parte comparten exactamente los mismos vecinos. Es necesario observar que no únicamente comparten los vecinos sinó también los no vecinos.

$$\begin{pmatrix}
0&0&0&1&1&1&0&0&0&1&1&1&1&1&1\\
0&0&0&1&1&1&0&0&0&1&1&1&1&1&1\\
0&0&0&1&1&1&0&0&0&1&1&1&1&1&1\\
1&1&1&0&0&0&0&0&0&1&1&1&0&0&0\\
1&1&1&0&0&0&0&0&0&1&1&1&0&0&0\\
1&1&1&0&0&0&0&0&0&1&1&1&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1\\
1&1&1&1&1&1&0&0&0&0&0&0&1&1&1\\
1&1&1&1&1&1&0&0&0&0&0&0&1&1&1\\
1&1&1&1&1&1&0&0&0&0&0&0&1&1&1\\
1&1&1&0&0&0&1&1&1&1&1&1&0&0&0\\
1&1&1&0&0&0&1&1&1&1&1&1&0&0&0\\
1&1&1&0&0&0&1&1&1&1&1&1&0&0&0
\end{pmatrix}$$

Figura 1. Matriz de adyacencia del grafo 3-anónimo de la figura 2

A continuación se enumeran algunas de las propiedades de los grafos k-anónimos.

Proposición 3. *Sea $G = (V, E)$ un grafo k-anónimo de acuerdo con la definición 2. Entonces se cumplen las condiciones siguientes:*

- *El grado mínimo del grafo es mayor que k;*
- *Existe una partición $V = V_1 \cup \ldots \cup V_n$ tal que*
  - *para todo i, $|V_i| \geq k$,*
  - *para todo $u, v \in V_i$, tenemos que $N(v) = N(u)$;*
  - *para todo $i \neq j$, $v \in V_i$ y $w \in V_j$, tenemos que $N(v) \neq N(w)$;*
- *Sea $v \in V$ un vértice de V prefijado. Entonces existe una partición de $N(v)$ en una o más clases de equivalencia. Las clases de esta partición forman un subconjunto de las clases de la partición del conjunto de vértices total.*

Observemos que la proposición 3 no implica en general que $N(v) \cap N(w) \neq \emptyset$ para $v \in V_i$ y $w \in V_j$.

Cuando todos los vértices tienen el mismo grado decimos que el grafo es regular. Un grafo regular que sea k-anónimo tiene, a parte de las propiedades descritas en la Proposición 3, las propiedades adicionales que se describen en la proposición siguiente.

Proposición 4. *Sea G un grafo regular de grado d que sea k-anónimo. Entonces se cumple*

- *El grado de anonimidad k es menor o igual que d;*
- *Si todas las clases de equivalencia V_i tienen la misma cardinalidad $|V_i| = k$, entonces k divide a d y a $|V|$.*

En la figura 2 se representa un grafo 3-anónimo con 15 vértices y 54 aristas. La matriz de adyacencia de este grafo es la presentada en la figura 1.

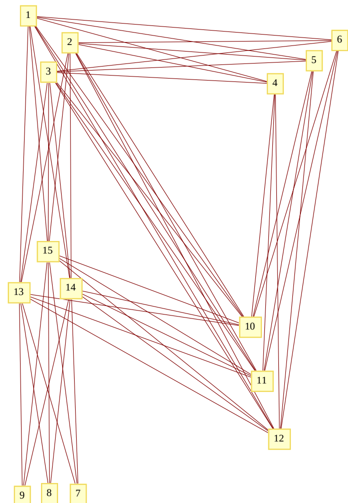

Figura 2. Grafo 3-anónimo

IV. (k, l)-ANONIMIDAD PARA GRAFOS

El problema de la definición presentada de k-anonimidad para grafos es que puede ser bastante restrictiva para grafos pequeños. Cuando k es grande en relación al orden $|V|$ del grafo habrá sólo un número pequeño de grafos no isomórficos que satisfagan el criterio de k-anonimidad. Por lo tanto, habrá muchos grafos que estén lejos de ser k-anónimos, y al quererlos proteger, representándolos por grafos k-anónimos, la pérdida de información será grande. Este hecho justifica la relajación de la k-anonimidad que se discute a continuación.

Siguiendo una idea presentada en [3], introducimos un segundo parámetro l y consideramos un grafo (k, l)-anónimo si es k-anónimo en relación a cualquier subconjunto de cardinalidad como máximo l de los vecinos de los vértices del grafo. La frase *un subconjunto de cardinalidad como máximo l del vecindario de un vértice* tiene dos interpretaciones diferentes. Cada interpretación da lugar a una definición de (k, l)-anonimidad y lo mejor es dejar que el contexto de anonimización determine cual es la interpretación adecuada.

Si el grafo no tiene aristas múltiples (un par de vértices pueden conectarse mediante una arista como máximo), entonces el vector fila a_v de la matriz de adyacencia que representa el conjunto de vecinos $N(v)$ de un vértice v es un vector en el espacio $\{0, 1\}^n$, donde $|V| = n$ es el número de vértices del grafo. Si el grafo tiene aristas múltiples, entonces el espacio es $(\mathbb{N} \cup \{0\})^n$. Sea $I \subseteq [1, |V|]$ un subconjunto de índices de cardinalidad $|I| \leq l$ y sea $a_v[I]$ los componentes sobre I del vector a_v que representa $N(v)$. Entonces podemos interpretar $a_v[I]$ como *un subconjunto de cardinalidad como máximo l del vecindario del vértice v*. Esta interpretación implica que, al caracterizar el vértice v, tengamos en cuenta no solo a sus vecinos, sino también a los vertices del grafo que no lo son, y conduce a la siguiente definición de (k, l)-anonimidad.

Definición 5. *Sea $G = (V, E)$ un grafo. Sea $\{a_v : v \in V\}$ el conjunto de filas de la matriz de adyacencia de G. Decimos que G es (k, l)-anónimo si para todo vértice $v \in V$ y para todo subconjunto de índices $I \subseteq \{1, \ldots, |V|\}$ de cardinalidad $|I| \leq l$ hay como mínimo $k - 1$ vértices $\{v_i\}_{i=1}^{k-1}$ distintos de manera que los componentes de a_{v_i} y a_v sobre I coinciden, para $i \in \{1, \ldots, k - 1\}$.*

Suponemos que la información accesible a un adversario está restringida al subgrafo inducido por un subconjunto de cardinalidad menor o igual que l de los vértices del grafo que este quiere atacar. Entonces, si el grafo atacado satisface la definición 5, el adversario no puede reidentificar ningún vértice en el grafo con probabilidad más grande que $1/k$.

Si en cambio consideramos que el vecindario de un vértice es el conjunto de vértices adyacentes, entonces obtenemos una interpretación diferente de la frase *un subconjunto de cardinalidad como máximo l del vecindario del vértice v*. Esto conduce a una definición distinta de (k, l)-anonimidad, en que no tenemos en cuenta los vértices que no son del vecindario.

Definición 6. *Sea $G = (V, E)$ un grafo. Decimos que G es (k, l)-anónimo si para todo vértice $v \in V$ y para todo subconjunto $S \subseteq N(v)$ de cardinalidad $|S| \leq l$ hay al menos $k - 1$ otros vértices distintos $\{v_i\}_{i=1}^{k-1}$ tal que $S \subseteq N(v_i)$ para $i \in \{1, \ldots, k\}$.*

En un grafo que satisface la definición 6 tenemos que un adversario que tiene acceso a un subgrafo inducido por un subconjunto de los vértices de G conteniendo menos de l de los vecinos del vértice v, sólo puede reidentificar a v con una probabilidad que no supera $1/k$.

Las diferencias entre las dos definiciones de (k, l)-anonimidad pueden ser ilustradas usando los ejemplos del grafo cubo, que vemos en la Figura 3 y el grafo hipercubo de la Figura 4. Es fácil ver que en el cubo, fijado un vértice v, y un subconjunto de vecinos de v de cardinalidad dos, hay otro vértice que también tiene este subconjunto de vecinos. En cambio, si consideramos tambien los vértices que no son vecinos de v, entonces podemos encontrar subconjuntos de cardinalidad dos que identifican unívocamente a v; todos los subconjuntos que contienen un vecino y el vértice opuesto a v. Por ejemplo, en el caso del vértice 1, tenemos que su vecindario es $\{2, 3, 4\}$ y su vértice opuesto es el 8. Entonces cada uno de los subconjuntos $\{2, 8\}$,$\{3, 8\}$,$\{4, 8\}$ identifica al vértice 1, ya que no hay otro vértice que tenga esas combinaciones de vecindario y complemento de vecindario.

En cambio, en el hipercubo, fijando cualquier subconjunto S de cardinalidad 2 de los 16 vértices del grafo, para todo vértice v hay otro vertice v' de manera que $N(v)$ coincide con $N(v')$ sobre S. Deducimos que el cubo es $(2, 2)$-anónimo según la definición 6, pero no según la definición 5, mientras que el hipercubo es $(2, 2)$-anónimo según las dos definiciones.

IV-A. *Propiedades de los grafos (k,l)-anónimos*

El hecho de que la definición 5 y la definición 6 son relajaciones de la definición 2, implica que un grafo (k, l)-anónimo no tiene porque ser necesariamente k-anónimo. De hecho se pueden encontrar ejemplos de grafos (k, l)-anónimos

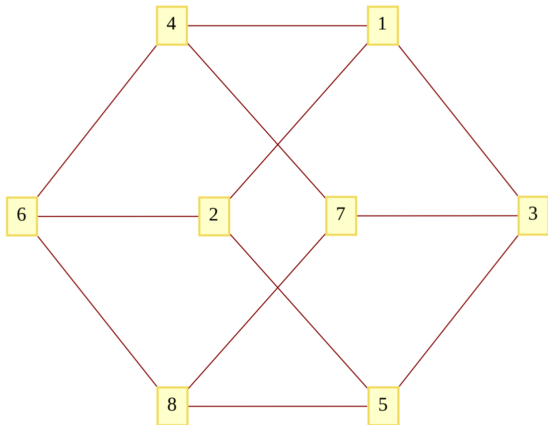

Figura 3. El grafo cubo

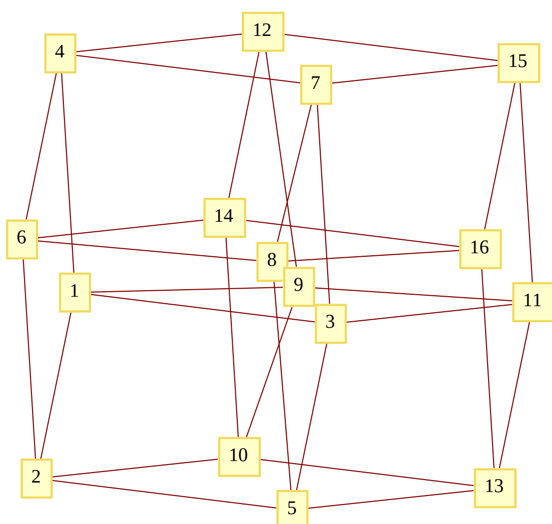

Figura 4. Un hipercubo

en que todo vértice está unívocamente identificado por alguna propiedad, por ejemplo, su grado.

En consecuencia, resulta obvio que si se sospecha que un adversario puede tener más información auxiliar que la de un subgrafo inducido como los descritos antes, entonces la (k, l)-anonimidad por si sola es una medida de proteccion insuficiente. No obstante, si se puede suponer que la información auxiliar extra es de una cierta naturaleza, entonces la (k, l)-anonimidad puede ser combinada con otras medidas para conseguir un nivel de privacidad acceptable. Por ejemplo, si la información auxiliar adicional es el grado de los vértices, entonces se puede combinar la (k, l)-anonimidad con la k-anonimidad de grado (véase [4]).

Sin embargo, cuando la información auxiliar sobre el grafo accesible para el adversario está restringida a

- el subgrafo inducido por un subconjunto de l vértices del grafo original en el caso de la definición 5, ó
- el subgrafo inducido por un subconjunto de vértices del grafo original que contiene a los sumo l de los vecinos de cada vértice, en el caso de la definición 6,

entonces la información que el adversario tiene sobre el grado de los vértices está igualmente restringida, de manera que la protección que proporciona la (k, l)-anonimidad es la que se espera.

De la definición se deduce directamente que k tiene que ser más pequeño que el número de vértices del grafo, $k \leq |V|$. Si suponemos que el grafo no contiene bucles, entonces el conjunto de k vértices que comparten vecinos y el conjunto de los l vecinos que comparten, tienen que ser disjuntos, de manera que $k + l \leq |V|$. Estas cotas no pueden ser mejores, ya que el grafo completo (V, E) sobre $n := |V|$ vértices es $(k, n - k)$-anónimo para todo $k \leq n$.

Consideremos un grafo (k, l)-anónimo según la definición 6. Sea d su grado mínimo. Si d es más pequeño que l, de manera que hay un vértice con un número de vecinos más pequeño que l, entonces v puede compartir a los sumo d vecinos con otros vértices, por lo que el grafo solo puede ser (k, d)-anónimo. Por lo tanto, en la definición 6, tenemos siempre que $l \leq d$.

Observemos que si un vértice comparte un conjunto de l vecinos con k otros vértices, entonces también comparte cualquier subconjunto de esos l vecinos con los mismos k vértices. Así, si un grafo es (k, l)-anónimo según la definición 6, entonces también es (k, l')-anónimo según la definición 6 para todo $l' \leq l$. Además, si un vértice comparte un conjunto de l vecinos con k vértices, entonces comparte el mismo conjunto de l vecinos con cualquier subconjunto de los k vértices. Por tanto, si un grafo es (k, l)-anónimo según la definición 6 entonces también lo es para todo $k' \leq k$.

Resumimos los resultados anteriores en la siguiente proposición.

Proposición 7. *Sea $G = (V, E)$ un grafo (k, l)-anónimo según la definición 6. Entonces las siguientes condiciones son ciertas:*

- *Si G carece de bucles, entonces $k + l \leq |V|$;*
- *l es más pequeño o igual que el grado mínimo de G;*
- *G es (k, l')-anónimo para todo $l' \leq l$;*
- *G es (k', l)-anónimo para todo $k' \leq k$.*

V. k-Anonimidad para grafos dirigidos

En un grafo dirigido, todo vértice v tiene dos vecindarios: el vecindario entrante $N_I(v)$ y el vecindario saliente $N_O(v)$. El grado entrante del vértice v es $d_I(v) = |N_I(v)|$ y el grado saliente del vértice v es $d_O(v) = |N_O(v)|$.

La matriz de adyacencia de un grafo dirigido se distingue de una matriz de adyacencia de un grafo no dirigido en que en general no es simétrica. Un vértice u puede tener un vértice v como vecino saliente, sin que v tenga a u como vecino saliente. No obstante, en este caso v tiene a u como vecino entrante (aunque u no tiene a v como vecino entrante). Concluimos que la matriz de adyacencia de un grafo dirigido (posiblemente con bucles y aristas múltiples) es cualquier matriz en la que los componentes son o bien cero, o bien, números naturales.

A continuación proporcionamos una caracterización de los grafos dirigidos k-anónimos, en términos de su matriz de adyacencia.

En un grafo k-anónimo, los vecindarios salientes y entrantes tienen que ser repetidos al menos k veces. Dado cualquier vértice v, el hecho de que haya k vértices con el mismo vecindario saliente $N_O(v)$ que v, no asegura que haya al menos k vértices indistinguibles de v, si es que estos k vértices tienen vecindarios entrantes diferentes. Por lo tanto, en un grafo k-anónimo, para todo vértice v hay al menos k vértices diferentes que tienen exactamente los mismos vecindarios entrantes y salientes que v. Observemos que esta definición es la extensión natural de la definición de k-anonimidad para grafos a grafos dirigidos ya que coincide con la definición 2 en el caso de grafo no dirigidos, es decir, cuando $N_O(v) = N_I(v)$ para todo vértice $v \in V$.

En consecuencia, una propiedad que tiene la matriz de adyacencia de un grafo k-anónimo dirigido es que cada fila y cada columna se repita al menos k veces, creando bloques en la matriz en las que la información es redundante. Deducimos que la información codificada en esa matriz se puede representar en una matriz reducida, utilizando solamente una fila y una columna de cada una de esos bloques de filas y columnas redundantes. Como los bloques no tienen porqué ser de tamaño exactamente k, la matriz reducida tampoco no tiene porque ser cuadrada. La matriz de adyacencia del grafo k-anónimo se obtiene de la matriz reducida, añadiendo como minimo $k-1$ copias de cada fila y cada columna. Se puede ver fácilmente que las filas y las columnas de la matriz de adyacencia que resulta se pueden reordenar de manera que la matriz contenga bloques de tamaño $a \times b$ con $a, b \geq k$. Cada bloque contendrá el mismo número natural repetido $a*b$ veces.

Proposición 8. *Sea G un grafo k-anónimo dirigido. Entonces las siguentes afirmaciones son ciertas:*

- $k \leq \min\{d_I(v) : v \in V\}$ *y* $k \leq \min\{d_O(v) : v \in V\}$*;*
- $k \leq |V| - \min\{d_I(v) : v \in V\}$ *y* $k \leq |V| - \min\{d_O(v) : v \in V\}$*;*
- *Hay una particion del conjunto de vértices $V = V_1 \cup \cdots \cup V_n$, de manera que todo vértice v pertenece a una y sólo una de las partes V_i, que llamamos el conjunto de anonimidad de v. Esta partición satisface*
  - $|V_i| \geq k$ *para todo* i*;*
  - $N_O(v) = N_O(u)$ *y* $N_I(v) = N_I(u)$ *para todo* $u, v \in V_i$ *para todo* i*;*
  - $N_O(v) \neq N_O(w)$ *ó* $N_I(v) \neq N_I(w)$ *para todo* $v \in V_i$ *y* $w \in V_j$ *para todo* $i \neq j$*. No obstante, en general podemos tener* $N_O(v) \cap N_O(w) \neq \emptyset$ *y* $N_I(v) \cap N_I(w) \neq \emptyset$*.*

VI. Conclusiones

En este trabajo se han estudiado definiciones de k-anonimidad para grafos. Se ha observado que si un grafo es k-anónimo respecto a su vecindario, entonces es k-anónimo respecto a cualquier otra propiedad de los grafos. Esto hace que la definición de k-anonimidad presentada en este trabajo sea suficiente en cualquier situación. Luego hemos presentado resultados sobre las propiedades de los grafos k-anónimos, para grafos dirigidos y no-dirigidos. Esto es útil, por ejemplo, al disenyar algorismos de k-anonimización, ya que permite conocer los objectos que queremos construir. Para el caso en que la k-anonimización de un cierto grafo provoque demasiada pérdida de información, presentamos también una relajación de la k-anonimidad que hemos llamado (k,l)-anonimidad. Es importante notar que siendo una relajación, la anonimidad proporcionada por un grafo (k,l)-anónimo existe solo bajo la suposición de que la información que tenga un adversario está limitado.

Agradecimientos

La autora quiere dar las gracias a Vicenç Torra y Jesús Manjón por su ayuda en la redacción de este artículo. Se agradece el apoyo parcial del MEC en los proyectos ARES (CONSOLIDER INGENIO 2010 CSD2007-00004) y RIPUP (TIN2009-11689) y de la Generalitat de Catalunya con el proyecto 2009 SGR 1135. La autora está financiada por una beca FPU (BOEs 17/11/2009 y 11/10/2010). Es investigadora de la Cátedra UNESCO en Privacidad de Datos, pero sus opiniones no reflejan necesariamente aquellas de la UNESCO ni comprometen a esta organzación.

Referencias

[1] Campan, A., Truta, T. M. (2009) Data and Structural K-Anonymity in Social Networks, F. Bonchi et al. (Eds.): PinKDD 2008, Lecture Notes in Computer Science, Volume 5456, Springer, 33-54.
[2] Dalenius, T. (1986) Finding a needle in a haystack, Journal of official statistics, 2:3 329 –336.
[3] Feder, T., Nabar, S. U. and Terzi, E. (2008) Anonymizing graphs, CoRR abs/0810.5578, October.
[4] Liu, K. and Terzi, E. (2008) Towards identity anonymization on graphs, Proc. SIGMOD 2008.
[5] Samarati, P. (2001) Protecting Respondents' Identities in Microdata Release, IEEE Trans. on Knowledge and Data Engineering, 13:6 1010-1027.
[6] Samarati, P. and Sweeney, L. (1998) Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression, SRI Intl. Tech. Rep.
[7] Sweeney, L. (2002) Achieving k-anonymity privacy protection using generalization and suppression, Int. J. of Unc., Fuzz. and Knowledge Based Systems 10:5 571-588.
[8] Sweeney, L. (2002) k-anonymity: a model for protecting privacy, Int. J. of Unc., Fuzz. and Knowledge Based Systems 10:5 557-570.
[9] Zhou, B. and Pei. J. (2008) Preserving privacy in social networks against neighborhood attacks, Proc. ICDE 2008.
[10] Zhou B, Pei J, Luk WS (2008) A brief survey on anonymization techniques for privacy preserving publishing of social network data. SIGKDD Explor 10(2):1222