Medidas contra ataques activos a la privacidad de una red social

Núria Busom
Dept. de Matemàtica
Universitat de Lleida
C. Jaume II, 69, E-25001 Lleida
Email: nuria.busom@udl.cat

Nacho López
Dept. de Matemàtica
Universitat de Lleida
C. Jaume II, 69, E-25001 Lleida
Email: nlopez@matematica.udl.cat

Francesc Sebé
Dept. de Matemàtica
Universitat de Lleida
C. Jaume II, 69, E-25001 Lleida
Email: fsebe@matematica.udl.cat

***Resumen*—En este trabajo se presenta una técnica que permite evitar que el ataque activo, denominado *del camino*, contra la privacidad de una red social propuesto en un artículo de Backstrom *et al.* tenga éxito. Concretamente, se propone y analiza un método basado en la eliminación de un conjunto aleatorio de aristas. Los resultados analíticos y experimentales permiten afirmar que la propuesta es efectiva siendo muy reducido su impacto sobre la calidad de los datos.**

I. Introducción

En la actualidad, existe una gran cantidad de plataformas en Internet donde cualquier persona puede crear una cuenta y establecer relaciones (de amistad, laborales, aficiones comunes, etc.) con otros usuarios. De este modo se crea una red social cuya estructura puede modelarse mediante un grafo donde cada nodo representa a un participante de la red y cada arista se corresponde con una relación entre dos usuarios. Aunque no serán consideradas en este trabajo, hay que tener en cuenta que existen relaciones que se modelan mejor mediante un arco dirigido tal como ocurre con relaciones del tipo "confiar en" o "ser seguidor de".

El análisis de este tipo de datos [5] y su evolución [4] proporciona información de gran valor en algunas áreas de investigación tales como la sociología, la psicología o la economía, entre otras. Dado que la información contenida en una plataforma de este tipo contiene datos de carácter personal, es muy importante que antes de cederla se tomen medidas para que no sea posible inferir información personal de sus participantes. Una medida que siempre debe tomarse consiste en anonimizar la red mediante la supresión de los datos que permiten identificar a los nodos (tales como el nombre).

Una vez los datos han sido cedidos o publicados, la privacidad de sus usuarios puede verse comprometida de dos formas:

- *Reidentificación de nodos:* un analista deshonesto es capaz de descubrir la identidad del usuario asociacido a un nodo concreto del grafo (anónimo) publicado. Cuando esto sucede se revela información tal como su número de relaciones.
- *Reidentificación de relaciones:* un analista descubre que existe una relación entre dos usuarios concretos de la red. Esto sucede, por ejemplo, cuando se ha podido reidentificar varios nodos. En este caso, comprobando la existencia de una arista entre los nodos correspondientes del grafo se deduce si existe o no una relación entre ellos.

Un analista deshonesto intentará descubrir los nodos que corresponden a determinados usuarios de la red, o si esto no es posible, acotar al máximo la lista de candidatos. Para ello, puede utilizar distintos tipos de información estructural conocida con anterioridad [9]. Así pues, si de antemano se conociera la cantidad de amistades de un participante de la red, esta información podría ser usada para acotar, a partir de su grado, los nodos que pueden corresponderle, o llegar a una reidentificación si existiera un único nodo con esta característica. Para evitar este riesgo, en [4] se propone un proceso que perturba los datos para que en el momento de su publicación, exista siempre un mínimo de k nodos compartiendo cada grado existente en la red. La propuesta [8] aborda el mismo problema considerando que el atacante conoce las relaciones existentes entre los nodos relacionados con un determinado usuario. En general, las medidas a tomar para evitar los ataques de reidentificación a partir del conocimiento previo de algún tipo de información estructural siempre requieren la aplicación de algún tipo de perturbación sobre los datos. Esta perturbación debe ser lo más reducida posible para evitar que el resultado de los estudios se vea alterado de forma significativa.

Existen propuestas que pretenden proporcionar anonimato contra cualquier tipo de información previa que pueda ser conocida por el atacante. Las propuestas k-*symmetry* [6] y k-*automorphism* [7] proporcionan seguridad contra la reidentificación de nodos mientras que la propuesta k-*isomorphism* [2] también proporciona seguridad contra la reidentificación de relaciones.

Todos los trabajos citados hasta el momento pretenden evitar los denominados ataques *pasivos* contra la privacidad. En ellos, el atacante intenta desvelar información confidencial de la red social haciendo uso solamente de la información que tiene a su disposición, sin efectuar ninguna manipulación.

II. Ataques activos contra la privacidad

En un ataque *activo*, el atacante realiza acciones previas orientadas a facilitar su tarea posterior de reidentificación una vez los datos sean cedidos. En [1] se describe una técnica que funciona del siguiente modo:

1. Antes de que la plataforma genere el grafo anonimizado G con los datos de la red social, un atacante crea k cuentas nuevas en la plataforma.
2. A continuación, crea relaciones entre estas nuevas cuentas creando así un subgrafo H que se hallará dentro del grafo global G.
3. Finalmente, desde estas nuevas cuentas, se establecen relaciones con determinados usuarios cuyos nodos en G se desea reidentificar.

Más adelante, cuando la plataforma publique el grafo anonimizado G, el atacante buscará el subgrafo H, y mediante los enlaces creados será capaz de saber qué nodos de G corresponden a los usuarios investigados. La forma en que H es creado debe facilitar su posterior localización. En [1] se propone una técnica llamada *del camino*. En la descripción que damos a continuación, la creación de un nodo nuevo se realiza mediante la creación de una cuenta en la plataforma, mientras que una arista se crea estableciendo una relación entre los usuarios asociados a los dos nodos involucrados.

Proceso para la creación de H (véase la Figura 1):

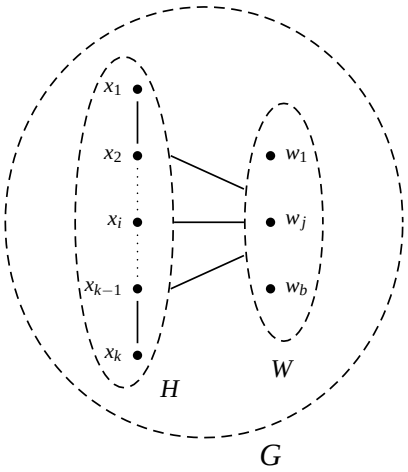

Figura 1. Creación de H como subgrafo de G.

1. Dado un valor pequeño $\delta > 0$, se calcula el valor constante $k = (2+\delta)\log n$ (n es el número de usuarios de la red) y se crean k nodos nuevos $\{x_1, \ldots, x_k\}$. Después, se elige $d_0 \leq d_1 = O(\log n)$, y para cada $i = 1, 2, \ldots, k$ se escoge un *grado externo* $\triangle_i \in [d_0, d_1]$ correspondiente al número de aristas entre x_i y nodos de $G - H$.
2. Dado el conjunto $W = \{w_1, \ldots, w_b\}$ de usuarios a investigar ($b = O(\log^2 n)$), para cada w_j se escoge un subconjunto $N_j \subseteq \{x_1, \ldots, x_k\}$ tal que cada N_j sea distinto y con la restricción adicional de que cada x_i aparezca como máximo en $\triangle_i$ conjuntos N_j. Después, se crean aristas entre w_j y cada $x_i \in N_j$.
3. Se añaden enlaces aleatorios entre nodos de H y nodos de $G - H$ hasta conseguir que cada nodo x_i tenga exactamente $\triangle_i$ aristas que lo relacionen con nodos de $G - H$.
4. Finalmente, se crean las aristas internas de H. Para cada $i = 1, \ldots, k-1$, se crea la arista $(x_i, x_i + 1)$ (formando el denominado *camino principal*) y se añade cada una de las aristas restantes (x_i, x_j) con probabilidad $1/2$. Denotaremos $\triangle'_i$ al grado de cada x_i en el grafo G una vez el proceso ha concluido.

Más adelante, la plataforma va a generar y publicar el grafo anonimizado G. En este momento, el atacante va a aplicar el siguiente proceso que tiene por objetivo la localización de H:

1. Se crea un arbol $\mathcal{T}$ formado por un único nodo raíz virtual α^*. Durante el proceso, cada nodo $\alpha \in \mathcal{T}$ (distinto de α^*) corresponderá a un nodo de G que denotaremos $f(\alpha)$.
2. En un primer paso, se añade al árbol una hoja para cada nodo de G cuyo grado sea $\triangle'_1$.
3. Durante la ejecución del algoritmo, dada una hoja α cuyo camino hacia la raíz es $\alpha^* = \alpha_0, \alpha_1, \ldots, \alpha_\ell = \alpha$, se consideran los vecinos v de $f(\alpha)$ cuyo grado es $\triangle'_{\ell+1}$ y tales que la arista $(f(\alpha_i), v)$ existe en G si y sólo si $(x_i, x_{\ell+1}) \in H$, para cada $i \leq \ell$. Para cada uno de estos v se añade una nueva hoja al árbol que cuelga de α.
4. Al final, si existe un único camino de longitud k en $\mathcal{T}$, éste se corresponde con los nodos de H.

El proceso de localización de H es eficiente siempre que el árbol $\mathcal{T}$ no crezca en exceso durante el proceso de búsqueda. Este es el caso (tal como se afirma en [1]), ya que las condiciones que debe cumplir un nodo para ser añadido al árbol son muy estrictas: debe tener un grado determinado, y las aristas hacia los nodos que se hallan en el camino hacia la raíz deber tener una estructura conocida muy concreta.

III. Medidas contra ataques activos

Una plataforma que va a ceder o publicar sus datos, debe velar por la privacidad de sus usuarios, por tanto, debe tomar medidas ante la posibilidad de que se haya realizado un ataque activo como el descrito en la Sección II. El objetivo de la plataforma consiste en conseguir que el atacante no sea capaz de localizar el subgrafo H.

Centrándonos en la técnica *del camino* descrita en la Sección II, para que el atacante consiga localizar H, es necesario que ni el subgrafo H ni el conjunto de aristas entre él y $G-H$ hayan sufrido ninguna modificación desde su creación. Si este fuera el caso (por ejemplo, alguna arista del tipo (x_i, x_{i+1}) ha sido suprimida), la localización de H fracasaría. Este es, por tanto, el objetivo de la plataforma.

La forma de perturbar el grafo G con el objetivo de evitar la posterior localización del subgrafo H puede hacerse de distintas formas. En este trabajo vamos a estudiar una estrategia basada en la eliminación de un conjunto aleatorio (y reducido) de aristas.

III-A. Eliminación aleatoria de aristas

Esta sección comienza planteando el Lema 3.1 que será la base del análisis posterior.

Lema 3.1: Sea G un grafo con m aristas, de las cuales distinguimos un subconjunto S cuyo cardinal es $|S| = d$. La

probabilidad de que al escoger ℓ aristas de G al azar se tome al menos una de las d distinguidas es

$$p = 1 - \frac{\binom{m-d}{\ell}}{\binom{m}{\ell}}. \quad (1)$$

Demostración: Un conjunto de m aristas, tiene un total de $\binom{m}{\ell}$ subconjuntos de ℓ aristas distintos, $\binom{m-d}{\ell}$ de los cuales no contiene ninguna de las d aristas de S. Por tanto, la probabilidad de no haber escogido ninguna de las aristas distinguidas es $\binom{m-d}{\ell} / \binom{m}{\ell}$. Así pues, la probabilidad de su suceso complementario (haber tomado al menos una de ellas) es 1 menos dicha cantidad, tal como se expresa en la ecuación (1). □

En el problema que estamos estudiando, las d aristas distinguidas son aquellas del grafo G que fueron creadas durante la inserción del grafo H. El valor concreto d es desconocido para la plataforma, aunque puede analizarse su distribución.

El subgrafo H, por tener k nodos, tiene un máximo de $\frac{k(k-1)}{2}$ aristas internas. De ellas, (paso 4 del proceso de construcción) las $(k-1)$ que forman el *camino principal* siempre existen mientras que cada una de las $\frac{k(k-1)}{2} - (k-1) = \frac{k^2-3k+2}{2}$ restantes existe con probabilidad $1/2$. Por tanto, el número esperado de aristas internas de H es

$$(k-1) + \frac{1}{2}\frac{(k^2-3k+2)}{2} = \frac{k^2+k-2}{4}.$$

En el paso 3 del proceso de construcción, desde cada nodo $x_i \in H$ se añade un total de $\triangle_i$ aristas hacia nodos de $G-H$, donde $\triangle_i$ es un valor aleatorio dentro del intervalo $[d_0, d_1]$ (escogido de forma uniforme). El valor esperado de aristas desde cada x_i hacia nodos de $G-H$ es, por tanto, $\frac{d_0+d_1}{2}$. Como hay k nodos x_i en H, la cantidad esperada de aristas creadas durante la inserción de H es

$$\hat{d} = \frac{k^2+k-2}{4} + k\frac{(d_0+d_1)}{2}.$$

La plataforma no puede calcular este valor ya que desconoce los valores d_0, d_1 escogidos por el atacante. Lo que sí puede hacer es calcular una cota inferior suya correspondiente a la situación más desfavorable para ella (y para sus usuarios), que es aquella en la que el número de aristas creadas por el atacante es menor (un subgrafo de tamaño reducido es más difícil de localizar). Por ello, se considera el menor tamaño posible para k, es decir $k = 2\log n$ ($\delta = 0$ en el paso 1 del proceso de construcción), y lo mismo para d_0, d_1 considerando que ambas constantes tienen valor 1. Con esto se obtiene que,

$$\hat{d}_{\text{mín}} = \frac{(2\log n)^2 + 2\log n - 2}{4} + 2\log n = \\ = \frac{4\log^2 n + 10\log n - 2}{4}. \quad (2)$$

A partir de este valor, empleando la ecuación (1), la plataforma puede determinar la cantidad ℓ de aristas que debe eliminar de G para que la probabilidad de que el atacante fracase en su posterior búsqueda de H esté por encima de un determinado valor. El cuadro I muestra, para un hipotético grafo G de orden n con m aristas, la estimación del número mínimo de aristas $\hat{d}_{\text{mín}}$ creadas durante la inserción de H junto con el número ℓ de aristas (y el porcentaje que representan respecto del total m) que deben ser eliminadas para que la probabilidad de que H sea no recuperable sea superior a 0.95. Podemos observar que el porcentaje de aristas a eliminar depende únicamente del número de nodos del grafo G y que éste se reduce al aumentar el orden del grafo.

n	m	$\hat{d}_{\text{mín}}$	ℓ	porcentaje
1,000	5,000	64	228	4.56 %
1,000	10,000	64	456	4.56 %
1,000	50,000	64	2286	4.56 %
10,000	50,000	107	1,379	2.76 %
10,000	100,000	107	2,760	2.76 %
10,000	500,000	107	13,804	2.76 %

Cuadro I
CÁLCULO DEL NÚMERO ℓ Y DEL PORCENTAJE DE ARISTAS A ELIMINAR EN UN GRAFO FORMADO POR n NODOS Y m ARISTAS PARA QUE LA PROBABILIDAD DE ELIMINAR H SEA SUPERIOR A 0.95.

III-B. Eliminación dirigida de aristas

En la sección anterior se ha estudiado la cantidad de aristas que deben ser eliminadas al azar para alcanzar un cierto grado de certeza de que el subgrafo H no podrá ser localizado.

A continuación se estudian algunas técnicas que ayudarán a conseguir el mismo resultado eliminando menos aristas, reduciendo así la perturbación sufrida por G.

III-B1. Consideración del tamaño de las componentes conexas: Por construcción, el subgrafo H es conexo y tiene $k \geq 2\log n$ nodos. Además, cada nodo de H está conectado con al menos d_0 nodos de $G-H$. Por tanto, el subgrafo H se halla en una componente conexa de G cuyo orden no será nunca inferior a $k+d_0$ (situación que se da cuando todos los nodos x_i de H tienen grado externo $\triangle_i = d_0$ y apuntan hacia los mismos nodos (aislados) de $G-H$). El valor mínimo se obtiene para $k = 2\log n$ y $d_0 = 1$, lo que lleva a la conclusión de que H no puede hallarse en una componente conexa de tamaño inferior a

$$c = 2\log n + 1. \quad (3)$$

Las aristas que se hallen en componentes conexas más pequeñas que c no es necesario que sean perturbadas.

III-B2. Consideración del grado de los nodos: Dados los k nodos de H, $\{x_1, \ldots, x_k\}$, tomando cualquier $x_i \in \{x_2, \ldots, x_{k-1}\}$, el proceso de construcción garantiza que las aristas $(x_{i-1}, x_i), (x_i, x_{i+1})$ siempre existen en H. Además, cualquier otra arista interna (x_i, x_j), existe con probabilidad $1/2$. También se sabe que x_i tiene $\triangle_i$ enlaces hacia nodos de $G-H$, por tanto, el grado de x_i es,

$$\deg(x_i) = 2 + g_1 + g_2, \quad (4)$$

siendo

$$g_1 \sim U(d_0, d_1), \qquad g_2 \sim B(k-3; \frac{1}{2}), \tag{5}$$

donde $U(n_1, n_2)$ denota una variable aleatoria que toma, de forma uniforme, valores enteros en el intervalo delimitado por n_1 y n_2, y $B(n, p)$ denota una distribución binomial con n ensayos y probabilidad de éxito p. Para los nodos x_1, x_k de H (los extremos del camino principal), se tiene que $\deg(x_i) = 1 + g_1 + g_2$, con g_1 distribuido tal como se indica en la ecuación (5) pero con $g_2 \sim B(k-2; \frac{1}{2})$. Dada la poca relevancia de la diferencia entre ambas distribuciones, se va a considerar que todos los nodos de H tienen el grado distribuido del mismo modo según se ha expresado en las ecuaciones (4) y (5).

La distribución exacta que toma $\deg(x_i)$ no puede ser determinada por la plataforma ya que los valores d_0, d_1 solamente son conocidos por el atacante. A pesar de ello, el conocimiento que hay sobre ella permite identificar algunos nodos que es muy poco probable que pertenezcan a H. Por un lado, dado que $d_0 \geq 1$, se cumple que $g_1 \geq 1$. Y por otro lado, sabiendo que $k \geq 2 \log n$, si se toma $k = 2 \log n$ y $d_0 = 1$ se puede determinar el mayor valor r para el que la probabilidad $P(g_2 < r)$ sea menor o igual que ϵ, siendo ϵ un valor pequeño. Con ello se llega a que con una probabilidad superior a $1 - \epsilon$, se cumple $\deg(x_i) \geq g$ siendo

$$g = 3 + r. \tag{6}$$

Los nodos de G cuyo grado esté por debajo de esta cantidad serán etiquetados como no pertenecientes a H. De este modo, las aristas que unen nodos así etiquetados podrán quedar sin perturbación.

III-B3. Consideración del número de triangulos de los nodos: Dados los nodos de H, $\{x_1, \ldots, x_k\}$, consideremos cualquier $x_i \in \{x_2, \ldots, x_{k-1}\}$. El número de nodos de H que son vecinos de x_i viene dado por $\deg_H(x_i) = 2 + g_2$, con $g_2 \sim B(k-3; \frac{1}{2})$. Dados dos vecinos de x_i en H, si existe una arista entre ellos, entonces ellos dos junto a x_i formaran parte de un ciclo de orden 3 (es decir, un triángulo). Considerando que la probabilidad de que esta arista exista es $1/2$, el número de triángulos internos a H en los que se espera que x_i participe son

$$t = \sum_{j=0}^{k-3} P(g_2 = j) \cdot \left(\begin{array}{c} j+2 \\ 2 \end{array} \right) \cdot \frac{1}{2} \tag{7}$$

En efecto, como $\deg_H(x_i) = 2 + g_2$, la probabilidad de que x_i tenga exactamente $j+2$ vecinos viene dada por $P(g_2 = j)$. Entre estos $j+2$ vecinos hay un total de $\left(\begin{array}{c} j+2 \\ 2 \end{array} \right)$ posibles aristas, cada una de las cuales existe con probabilidad 1/2. Sumando para cada j tomando valores entre 0 y $k-3$ llegamos a la fórmula que nos da el valor esperado t. La forma de calcular este valor para los nodos x_1 y x_k es ligeramente distinta, aunque al ser esta diferencia muy poco significativa (igual que en la Sección III-B2), para ellos se empleará igualmente la fórmula (7). En este cálculo no se ha considerado que las aristas que forman el *camino principal* de H siempre existen, con lo cual el valor esperado real sería un poco superior a t.

Aquellos nodos que participen en un número inferior a $t/2$ (calculado con $k = 2 \log n$) serán considerados como muy poco probables de pertenecer a H. Las aristas entre nodos considerados de este modo no será necesario que sean perturbadas.

IV. Resultados experimentales

Para comprobar el correcto funcionamiento de estas técnicas se ha tomado el conjunto de datos "coautores" generado a partir de información bibliográfica extraida de la colección *Computer Science Bibliographies*[1] donde a cada autor se le asigna un nodo y las aristas representan relaciones 'ser coautor'. Del grafo resultante, se han eliminado los nodos aislados (sin ninguna arista) obteniendo así un grafo con 8,936 nodos y 12,126 aristas. Conjuntos de datos construidos del mismo modo se han empleado previamente en trabajos como por ejemplo [3], [7].

Simulando las acciones que llevaría a cabo un atacante, se ha tomado el grafo *coautores* y se le ha creado un subgrafo H tomando $k = 18$ y $d_0 = 1, d_1 = 10$ (valores recomendados en [1]). Durante la creación de H se han añadido 181 aristas.

De este modo, la plataforma publicaría un grafo G con $8,954$ nodos y $12,307$ aristas. Aplicando la fórmula (2), la plataforma puede estimar que durante la creación de un hipotético subgrafo H se ha añadido un mínimo de $\hat{d}_{\text{mín}} = 105$ aristas con lo cual, si quiere que, con una probabilidad superior a 0.95, H no sea recuperable, deberá eliminar $\ell = 345$ aristas (que representan el 2.8 % del total).

Veamos ahora como se puede reducir la cantidad de aristas a eliminar. En un principio, la plataforma puede estimar que el número mínimo de nodos del subgrafo H presuntamente introducido en G tendrá un mínimo de $k = 2 \log n = 18$ nodos. A partir de la fórmula (3) también estima que no es posible que H se halle en una componente conexa de tamaño inferior a $c = 19$. A partir de este dato, es capaz de descartar un total de $4,993$ aristas del grafo G que son aquellas que se hallan en componentes conexas pequeñas.

Después, a partir de la fórmula (6) se deduce que los nodos de H es muy poco probable (probabilidad por debajo de $\epsilon = 0.05$) que tengan grado inferior a $g = 6$. En nuestro ejemplo, considerando las aristas que unen nodos con un grado inferior a esa cantidad, se pueden descartar otras $2,164$ aristas.

Finalmente, el número estimado de triángulos de los que forma parte, de media, cada nodo de H empleando la fórmula (7) es $t = 21.1$. Etiquetando los nodos de G que participan en menos de $t/2 = 10$ triángulos como probablemente no pertenecientes a H se pueden descartar las aristas que unen nodos etiquetados de este modo. Mediante este proceso se han descartado otras $2,834$ aristas.

Finalizado el proceso, queda un total de $2,316$ aristas entre las que se encuentran las $\hat{d}_{\text{mín}} = 105$ que la plataforma cree

[1] http://liinwww.ira.uka.de/bibliography

que, como mínimo, pertenecen a H. Para eliminar alguna de ellas con una probabilidad superior a 0.95 se deben eliminar 64 (representan el 0.52 % del total).

El beneficio de aplicar la técnica de la eliminación dirigida de aristas es claro. En este ejemplo concreto se ha conseguido reducir el número de aristas a eliminar de las 345 iniciales a solamente 64. Es de destacar que, muy posiblemente, la poca modificación que se realiza sobre el grafo (dado que se eliminan pocas aristas) implica que las medidas estructurales del grafo se vean también poco modificadas.

V. Conclusión

En este trabajo se ha presentado una técnica basada en la eliminación de un conjunto aleatorio de relaciones que permite evitar que el ataque activo *del camino*, propuesto en [1], contra la privacidad de una red social tenga éxito. Junto con el método básico se han detallado algunos criterios que permiten reducir el número de relaciones que debe ser suprimido. Los resultados experimentales muestran que la técnica puede aplicarse de forma efectiva sobre una red social con cerca de 9,000 nodos afectando solamente al 0.52 % de las relaciones.

Como trabajo futuro se estudiarán nuevos criterios para reducir más la cantidad de relaciones que deben ser eliminadas.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Ciencia e Innovación mediante los proyectos TIN2010-18978 y "ARES" CSD2007-0004. El Grupo de Investigación en Criptografía y Grafos está reconocido por la Generalitat de Catalunya (2009SGR442).

Referencias

[1] L. Backstrom, C. Dwork, J. Kleinberg, "Wherefore art thou R3579X? anonymized social networks, hidden patterns, and structural steganography", *Proc. of 16th Intl. World Wide Conference*, May 8–12, 2007.

[2] J. Cheng, A.W-C. Fu, J. Liu, "K-Isomorphism: privacy preserving network publication against structural attacks", *Proc. of SIGMOD'10*, June 6–11, pp. 459–470, 2010.

[3] K. Liu, E. Terzi, "Towards identity anonymization on graphs", *Proc. of ACM SIGMOD Conference*, June 9–12, pp. 93–106, 2008.

[4] R. Kumar, J. Novak, A. Tomkins, "Structure and evolution of online social networks", *Proc. of The Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 20–23, pp. 611–617, 2006.

[5] J. Scott, Social network analysis handbook. Sage Publications Inc., 2000.

[6] W. Wu, Y. Xiao, W. Wang, Z. He, Z. Wang, "K-Symmetry model for identity anonymization in social networks", *Proc. of 13th International Conference on Extending Database Technology*, March 22–26, pp. 111–122, 2010.

[7] L. Zou, L. Chen, M.T. Öszu, "K-Automorphism: a general framework for privacy preserving network publication", *Proc. of 35th Intl. Conference on Very Large Data Bases*, August 24–28, pp. 946–957, 2009.

[8] B. Zhou, J. Pei, "Preserving privacy in social networks against neighborhood attacks", *Proc. of the 24th International Conference on Data Engineering Workshops*, April 7–12, pp. 506–515, 2008.

[9] B. Zhou, J. Pei, W-S. Luk, "A brief survey on anonymization techniques for privacy preserving publishing of social network data", *ACM SIGKDD Explorations Newsletter*, 10 (2), 12–22, 2008.