Symulacja w Badaniach i Rozwoju
Vol. 6, No. 2/2015

Michał ZABIELSKI, Rafał KASPRZYK, Zbigniew TARAPATA
Wojskowa Akademia Techniczna, Wydział Cybernetyki
ul. Kaliskiego 2, 00-908 Warszawa
E-mail: michal.zabielski@wat.edu.pl, rafal.kasprzyk@wat.edu.pl, zbigniew.tarapata@wat.edu.pl

Metody wykrywania naruszeń prywatności w internetowych sieciach społecznych

1 Wprowadzenie

Obecna popularność internetowych sieci społecznych (ang. *Online Social Networks, OSN*) nie pozostawia złudzeń co do tego, iż właśnie te narzędzia mają istotny wpływ na kształtowanie się relacji międzyludzkich we współczesnym świecie. Jest to związane z wieloma czynnikami. Przede wszystkim z szeroką dostępnością sieci Internet, która dała możliwość szybkiego kontaktowania się osób fizycznie bardzo od siebie oddalonych. Dodatkowo poczucie anonimowości w sieci pozwala niektórym osobom czuć się pewniej w nawiązywaniu relacji z innymi. Z drugiej strony ciągły rozwój internetowych sieci społecznych wprowadza mechanizmy, które w stopniu coraz lepszym pozwalają na wiarygodną identyfikację ich uczestników. W ostatnim czasie dużego znaczenia nabrało zagadnienie ochrony **prywatności**, będącej istotnym aspektem w budowaniu relacji międzyludzkich, dzięki której możliwe jest wykorzystanie internetowej sieci społecznej jako wiarygodnej społecznej struktury.

W zależności od kontekstu rozważań prywatność może być rozumiana różnie, gdyż nie istnieje jej jednoznaczna definicja. Na potrzeby niniejszego opracowania jest ona traktowana jako *możliwość jednostki lub grupy osób do utrzymania swych danych oraz osobistych zwyczajów i zachowań nieujawnionych publicznie* [1]. Szczególnie istotnym aspektem z punktu widzenia internetowej sieci społecznej jest kwestia ukrywania atrybutów swojego profilu społecznego przed niepowołaną grupą osób. W sytuacji gdy dochodzi do udostępnienia danych osobie nieautoryzowanej, mamy do czynienia ze zjawiskiem naruszenia prywatności.

Celem opracowania jest prezentacja zestawu modeli i metod pozwalających na wykrywanie naruszeń prywatności w internetowych sieciach społecznych W rozdziale drugim omówione zostaną problemy związane z faktem zakłócania prywatności oraz pomysły na to, jak można je rozwiązać. Wynikiem tych rozważań będą mechanizmy ujawniające „miejsca" w internetowej sieci społecznej potencjalnie zagrożone naruszeniem prywatności, które opisuje rozdział trzeci. W rozdziale czwartym zaprezentowana zostanie koncepcja symulatora opracowywanego na potrzeby testowania i praktycznego wykorzystania prezentowanych modeli i metod.

2 Problem naruszania prywatności w internetowych sieciach społecznych

Problem naruszania prywatności w Internecie występuje w zasadzie od momentu, gdy pojawiły się mechanizmy pozwalające na przechowywanie danych osobowych o użytkownikach [2]. Zagadnienie to, z racji swej natury, w sposób oczywisty zostało przeniesione na płaszczyznę internetowych sieci społecznych ze względu na fakt, iż te bazują w większości na sieci WWW.

Wyniki analiz wielu naukowców wskazują na szeroki zakres technik pozwalających na wykrywanie informacji niejawnych o użytkowniku sieci społecznej. W ogólności możemy je podzielić na dwie klasy: lokalne i globalne. Mechanizmy lokalne skupiają się głównie na naruszaniu prywatności poprzez badanie odpowiedniego profilu użytkownika i jego bezpośredniego sąsiedztwa. W ten sposób możemy na przykład dokonywać odkrywania ukrytych atrybutów profilu użytkownika – technika ta została przedstawiona w [3]. Odmiennym podejściem do problemu charakteryzują się natomiast modele globalne. Opierają się one na całej pozyskanej strukturze sieci społecznej. Takie podejście pozwala zaangażować dodatkowe grupy algorytmów, od uczenia maszynowego po techniki oparte na podobieństwie grafów i sieci. W wyniku takiego działania możemy pozyskiwać dane wynikające nie tylko bezpośrednio z profilu użytkownika i znajomości atrybutów innych profili w sieci, ale również wywodzące się z podobieństwa struktur sieciowych między wieloma internetowymi sieciami społecznymi. Ten sposób daje możliwość detekcji zjawisk trudnych do zbadania metodami lokalnymi, takich jak na przykład klonowanie profili. Przykład wykorzystania techniki globalnej do odnajdywania faktu wykradania tożsamości został zaprezentowany w [4].

Poza wspomnianymi technikami naruszania prywatności istnieje jeszcze szereg innych mechanizmów. Do najpopularniejszych należą: odkrywanie profili społecznych ukrytych w sieci czy przechwytywanie znajomych. Wszystkie te techniki opierają się na elemencie stanowiącym podstawę istnienia osoby w sieci społecznej: profilu użytkownika. Profil użytkownika w internetowej sieci społecznej możemy opisać jako zestaw cech i powiązań między innymi profilami, które opisują osobę w rozpatrywanej sieci. W zależności od typu OSN zestaw atrybutów i powiązań składających się na profil może być inny. Fakt ten powoduje, że obecnie w Internecie nie istnieje jeden unikalny profil sieci społecznej dla danej osoby, co daje możliwość podszywania się pod daną osobę w różnych OSN-ach.

Poziom skomplikowania rozważań na temat naruszania prywatności w internetowych sieciach społecznych potęguje zjawisko *agregowania sieci społecznych* (ang. *social network aggregation*), które wskazuje na fakt, iż dana osoba może posiadać zupełne odmienne profile w różnych sieciach społecznych. Zwiększa to szansę na potencjalne sklonowanie profilu czy podszycie się pod inną osobę, utrudniając jednocześnie możliwość wiarygodnego porównywania profili użytkowników między internetowymi sieciami społecznymi. Te i inne problemy spowodowały powstanie mechanizmów wykrywania naruszeń prywatności.

3 Modele wykrywania naruszeń prywatności

Znając dokładnie problem naruszania prywatności w Internecie, a w szczególności w internetowych sieciach społecznych, możliwe jest opracowanie mechanizmów pozwalających na zapobieganie zakłócaniu prywatności. W niniejszej pracy zaprezentowane zostaną trzy modele, z których jeden kwalifikuje się do grupy modeli lokalnych, podczas gdy pozostałe dwa są modelami globalnymi. Są to: model wyszukiwania ukrytych atrybutów profilu użytkownika, model wyszukiwania strukturalnych wzorców w sieci oraz model prognozowania rozwoju struktury sieci społecznej.

Model wyszukiwania ukrytych atrybutów profilu użytkownika opiera się na mechanizmie uczenia częściowo nadzorowanego. Badania wskazują na to, że takie podejście daje efekty lepsze zarówno od uczenia nadzorowanego, jak i nienadzorowanego [5]. Istotnym czynnikiem takiego modelu jest odpowiedni dobór klasyfikatora, którego zadaniem jest odpowiedni przydział wartości atrybutów profilu użytkownika w sieci społecznej do profili badanych osób. Istnieje wiele rodzajów klasyfikatorów – do przykładowych należą funkcja Gaussa czy maszyna wektorów wspierających [6]. Analiza prac innych badaczy w tym zakresie wskazuje, że dla struktur grafowo-sieciowych najlepszym klasyfikatorem okazuje się funkcja Gaussa [7], mechanizm globalnej i lokalnej spójności w sieci [3] oraz specjalistyczna funkcja ważona, której odpowiedni wariant jest wykorzystywany w analizowanym modelu.

Do opisania opracowanego podejścia potrzebne będą następujące oznaczenia:
W - zbiór węzłów w sieci, reprezentujących osoby w internetowej sieci społecznej;
U- zbiór krawędzi w sieci, odwzorowujących relacje między poszczególnymi osobami w sieci społecznej, $U \subset \{\{x, y\} \subset W\}$;
A- liczba możliwych atrybutów węzłów w sieci społecznej;
K_a- zbiór wartości atrybutu $a \in \overline{\{1, \dots, A\}}$;
k_a^n - wartość atrybutu $a \in \overline{\{1, \dots, A\}}$ dla węzła $n \in \overline{W}$, $k_a^n \in K_a$;
$\alpha_a: W \to K_a$ - funkcja przyporządkowująca węzłom wartość a-tego atrybutu;
$\alpha_a(n) = k_a^n$;
$b_a^n(x)$ – zmienna binarna określająca występowanie wartości $x \in K_a$ dla atrybutu a węzła n:

$$b_a^n(x) = \begin{cases} 1 & gdy\ k_a^n = x \\ 0 & w\ przeciwnym\ przypadku \end{cases};$$

$i_a^n(x)$ - istotność wartości $x \in K_a$ atrybutu $a \in \overline{\{1, \dots, A\}}$ dla $n \in \overline{W}$, $i_a^n(x) \in [0,1]$;
$d: U \to (0,1]$ - funkcja symetryczna określająca siłę relacji między połączonymi węzłami sieci;
d_{nm}- siła relacji n-tego węzła z m-tym, $d_{nm} \in (0,1]$;
s_n- zbiór bezpośrednich sąsiadów węzła n-tego.

Internetowa sieć społeczna formalnie zdefiniowana została następująco:

$$S = \langle \langle W, U \rangle, \{\alpha_a\}_{a=\overline{1,A}}, d \rangle \ . \tag{1}$$

Węzły opisane są zestawem atrybutów, które łącznie budują profil tej osoby w sieci społecznej. Zakładamy, że znane są zbiory możliwych wartości każdego z atrybutów. Aby poznać wartość nieznanego atrybutu a' danej osoby n, badamy wartości

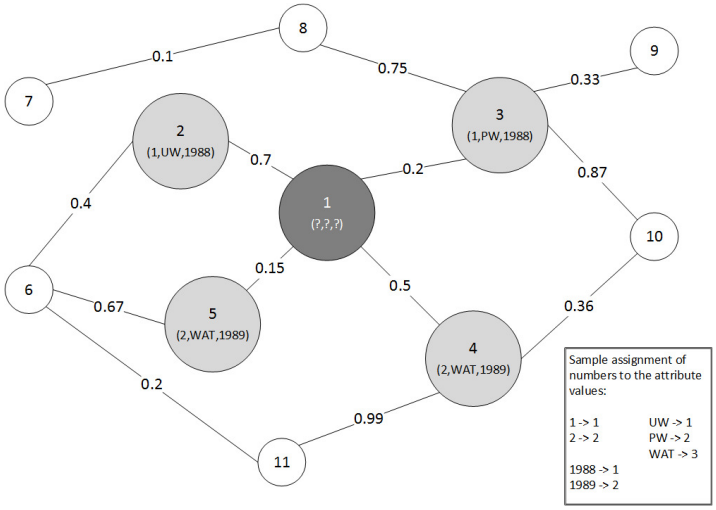
poszczególnych atrybutów u bezpośrednich znajomych rozpatrywanego osobnika. Dodatkowo uwzględniamy siłę relacji między nimi, wychodząc z założenia, że większa siła relacji wzmacnia wiarygodność danych przeznaczonych do analizy. Wyznaczamy wielkość $i_{a'}^{n}(x)$ daną wzorem:

$$i_{a'}^{n}(x) = \frac{\sum_{m \in s_n} b_{a'}^{m}(x) d_{nm}}{|s_n|} \quad . \tag{2}$$

Spośród wszystkich wyznaczonych istotności wybieramy wartość największą, tzn. taką wartość x^* atrybutu, dla której istotność jest największa - jest ona poszukiwaną wartością atrybutu rozpatrywanego węzła:

$$i_{a'}^{n}(x^*) = \max_{x \in K_{a'}} i_{a'}^{n}(x) \Rightarrow k_{a'}^{n} = x^* \tag{3}$$

Przykład zastosowania proponowanej metody oparty został na sieci przedstawionej na rysunku 1. Założono, że poszukiwanym atrybutem jest atrybut nr 3 w wierzchołu czarnym (węzeł nr 1).



Rys. 1. Przykładowa sieć społeczna zbudowana zgodnie z definicją (1); dla uproszczenia, w dalszych rozważaniach rozpatrywane są tylko węzły szare i czarny [opracowanie własne]

Fig. 1. Sample social network build according to definition (1). For simplicity, we will only consider gray and black nodes [own proposition]

Ponieważ w sieci możemy wyróżnić dwie możliwe wartości atrybutu numer trzy, dokonujemy wyznaczenia istotności dla dwóch możliwych wartości atrybutu zgodnie z zależnością (2):

$$i_3^1(1988) = \frac{1 * 0.7 + 1 * 0.2 + 0 * 0.5 + 0 * 0.15}{4} = 0.225$$

$$i_3^1(1989) = \frac{0 * 0.7 + 0 * 0.2 + 1 * 0.5 + 1 * 0.15}{4} = 0.163$$

Ponieważ największą wartość istotności uzyskujemy dla wartości 1988 atrybutu trzeciego, zatem ta właśnie wartość zostanie przypisana do węzła numer 1.

Rozpatrywany model kładzie duży nacisk na siłę relacji między węzłami sieci, zakładając, że jest to jeden z czynników decydujących o skuteczności określanych wartości atrybutów. Formuły tego typu wykorzystuje się w modelach rozprzestrzeniania opinii w sieciach społecznych i mają one swoje uzasadnienie w znanych wynikach badań. W rozpatrywanym modelu istotną zależnością jest (3), która stanowi swego rodzaju klasyfikator. Wymaga ona jednakże weryfikacji na dużym zbiorze rzeczywistych danych testowych (sieci), co jest przedmiotem aktualnych prac autorów. Ważne jest, że wykorzystując przedstawione podejście, nie musimy ograniczać się jedynie do funkcji maksimum. W sytuacji gdy dokładnie znamy strukturę analizowanej sieci społecznej i specyfikę jej tworzenia i rozwijania, może się okazać, że lepszej klasyfikacji dokona się, opierając się na średniej arytmetycznej lub geometrycznej. Stanowi to zatem element modelu, który może pozwolić na jego kalibrację stosownie do potrzeb.

Kolejną metodą opracowaną na potrzeby analiz naruszeń prywatności w internetowych sieciach społecznych jest model wyszukiwania strukturalnych wzorców w sieci. W przeciwieństwie do podejścia opisanego wcześniej rozpatrywany sposób należy do grupy rozwiązań globalnych i przyjmuje jako dane dwa modele sieci społecznych. Jeden z nich stanowi sieć wzorcową, która jest odpowiednikiem zbioru uczącego dla algorytmów uczenia maszynowego. Drugi z nich opisuje z kolei badaną sieć społeczną. Idea rozwiązania jest prosta – stosując mechanizmy badania podobieństwa grafów i sieci, wyszukiwane są takie fragmenty sieci społecznej, które są najbardziej do siebie podobne pod względem strukturalnym. Następnie dokonywane jest porównywanie występujących w znalezionych fragmentach profili – jeśli są one zbieżne, z dokładnością do pewnej wartości progowej, możemy założyć, że mamy do czynienia z takimi samymi profilami. Wskazywać to może na jeden z dwóch scenariuszy: albo użytkownik posiada profil w badanej sieci społecznej, albo ktoś podszywa się pod niego i próbuje przechwycić jego znajomych. Z tego względu technika ta sprawdza się najlepiej do wykrywania mechanizmu klonowania profili oraz przechwytywania znajomych, nie znajdzie jednak zastosowania w sytuacji, gdy będziemy chcieli poznać ukryte wartości atrybutów składających się na profil osoby w sieci społecznej.

W modelu wyszukiwania strukturalnych wzorców ważnym mechanizmem jest algorytm badania podobieństwa między grafami i sieciami. Spośród wielu technik wypracowanych przez lata godnymi uwagi dla tego typu zastosowań są algorytmy opisane w [8, 9]. Nie mniej istotnym elementem modelu jest wartość progowa, po przekroczeniu której możemy uznać, że fragment obu sieci jest wystarczająco podobny, by przejść do porównywania profili. Jeśli wartość ta będzie zbyt mała, wówczas zbyt często będzie dochodziło do kosztownego operacyjnie, zbędnego porównywania profili osób w sieci społecznej. Z kolei zbyt wysoki próg może doprowadzić do sytuacji, w której niemożliwe będzie wykrycie jakiegokolwiek podobieństwa, co tym samym przełoży się na brak użyteczność modelu.

Rozważania nad wcześniej wspomnianymi modelami, w połączeniu ze znajomością problemów wynikających z naruszania prywatności w internetowych sieciach

społecznych, doprowadziły do próby rozwiązania nowej kwestii, a mianowicie: czy możliwa jest taka analiza sieci społecznej, aby wykrywać możliwość naruszenia prywatności, zanim jeszcze ona nastąpi? Ten problem przyczynił się do zastosowania metod prognozowania rozwoju struktury sieci społecznej w obszarze analizy naruszeń prywatności. Punktem wyjścia były prace Kleinberga oraz Liben-Nowella na temat przewidywania tworzenia się połączeń między ludźmi w sieciach społecznych [10]. Wykorzystując zaprezentowane tam podejście i uwzględniając model wyszukiwania strukturalnych wzorców, możemy na podstawie sieci wzorcowej próbować prognozować rozprzestrzenianie się relacji w sieci badanej oraz wykrywać w niej podobieństwo z siecią wzorcową. Jeśli takie podobieństwo zostałoby wykryte, można przypuszczać, że znaleziony profil osoby z sieci badanej jest zbieżny z profilem osoby z sieci wzorcowej. Po raz kolejny pozwala to nam stwierdzić, że albo osoba buduje właśnie swój profil w nowej sieci społecznej, albo ktoś sklonował jej profil i jest w trakcie przechwytywania znajomych. Takie rozwiązanie pozwala na odpowiednio wczesną reakcję minimalizującą skutki naruszania prywatności wynikających z klonowania profili.

4 Koncepcja symulatora

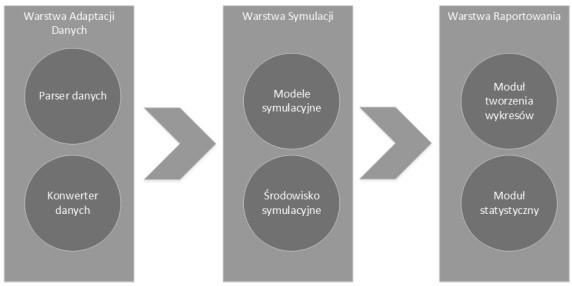
Aby zweryfikować poprawność opracowanych modeli i jednocześnie znaleźć dla nich zastosowanie praktyczne, postanowiono zbudować symulator internetowych sieci społecznych, który w sposób bezpośredni zaangażuje model prognozowania rozwoju struktury sieci społecznej oraz model wyszukiwania ukrytych atrybutów profilu użytkownika. Proponowany schemat działania takiej aplikacji ilustruje rysunek 1.

Rys. 2. Proponowany schemat działania symulatora [opracowanie własne]
Fig. 2. Simulator working scheme proposition [own proposition]

Idea funkcjonowania bazuje bezpośrednio na opracowanych i przedstawionych w poprzednich rozdziałach modelach i metodach. Po pozyskaniu sieci wzorcowej i badanej następowałaby obróbka danych do formatów „czytanych" przez poszczególne usługi symulatora. W kolejnym kroku dochodziłoby do odkrywania ukrytych atrybutów w sieci badanej na podstawie danych o bezpośrednich sąsiadach. Po odkryciu wartości atrybutów uruchamiana byłaby symulacja, w czasie której w sposób zgodny z modelem prognozowania rozwoju struktury sieci społecznej dochodziłoby do rozwoju badanej sieci społecznej i porównywania jej ze wzorcem. W wyniku symulacji otrzymalibyśmy

raport, z którego możliwe byłoby rozpoznanie, gdzie może dochodzić do klonowania profili oraz jakie atrybuty prywatne są możliwe do wykrycia przy takiej strukturze sieci społecznej. Z architektonicznego punktu widzenia rozpatrywany symulator składałby się z modułu adaptacji danych, który dokonywałby obróbki otrzymanych modeli sieciowych do formatu wymaganego do dalszych analiz. Kolejną warstwą byłaby warstwa symulacji, zadaniem której byłoby realizowanie zaimplementowanych modeli naruszania prywatności z wykorzystaniem otrzymanych modeli sieci wzorcowej i badanej. Wszelkie dane uzyskane w wyniku symulacji przekazywane byłyby do modułu raportowania, który zajmowałby się interpretacją otrzymanych wyników i przekształcaniem ich do postaci możliwej do łatwej interpretacji przez eksperta. Proponowaną architekturę rozwiązania przedstawia rysunek 3.

Z podanego schematu działania wyłania się szereg zastosowań, w których można byłoby wykorzystać proponowany symulator. W sposób oczywisty omawiane narzędzie dałoby możliwość zwiększenia skuteczności w wyszukiwaniu naruszeń prywatności. Znając luki bezpieczeństwa w sieci społecznej, możemy przedsięwziąć odpowiednie kroki w miejscu, gdzie jest to rzeczywiście konieczne. Co więcej, dokonując prognozowania przyszłych powiązań w sieci społecznej, bylibyśmy w stanie z pewnym wyprzedzeniem wykrywać potencjalne niedopatrzenia mogące skutkować naruszeniem prywatności danej osoby. Innym scenariuszem wykorzystania symulatora może być możliwość testowania różnych polityk prywatności przed rzeczywistym ich wdrożeniem. Pozwoliłoby to na uprzednie zweryfikowanie poprawności odpowiednich mechanizmów i wybór tych, które w sposób najbardziej adekwatny są w stanie uchronić przed incydentami naruszania prywatności.



Rys. 3. Proponowana architektura symulatora [opracowanie własne]
Fig. 3. Simulator architecture proposition [own proposition]

5 Podsumowanie

Problem naruszania prywatności w internetowej sieci społecznej, zarówno rozpatrywany lokalnie, jak i globalnie, jest zagadnieniem dość trudnym. Dodatkową kwestią komplikującą ten temat jest fakt, iż różne internetowe sieci społeczne powstają często niejako z zupełnie innych potrzeb jej uczestników, co determinuje różne mechanizmy nawiązywania relacji międzyludzkich. Innym problemem może być ustalenie wiarygodności modelu sieci społecznej stanowiącego wzorzec. W przypadku gdy sieć wzorcowa nie będzie wiernie odwzorowywała relacji zawiązywanych przez osoby, musimy liczyć się z błędnymi klasyfikacjami i wnioskami. Istotnym

utrudnieniem może być również celowe tworzenie różnych struktur sieci społecznych przez daną osobę w celu zachowania anonimowości.

Przedstawione modele i metody stanowią pewien punkt wyjścia do rozważań na temat mechanizmów detekcji naruszeń prywatności w sieciach społecznych i niewątpliwie wymagają rozwoju. W dalszej kolejności konieczne będzie opracowanie klasyfikatorów, które uwzględniają naturę (mechanizm nawiązywania relacji międzyludzkich) różnych internetowych sieci społecznych. Również modyfikacja funkcji istotności i rozszerzenie jej na zasięg globalny może skutkować poprawą klasyfikacji.

Literatura

1. Jeff Smith H.: *Managing Privacy: Information Technology and Corporate America*, UNC Press Books, 1994
2. Faith Cranor L.: Internet privacy, *Communications of the ACM* nr 42, rozdz. 2, 28-38, 1999
3. Mo M., Wang D., Li B., Hong D., King I., Exploit of Online Social Networks with Semi-Supervised Learning, *The 2010 International Joint Conference on Neural Networks (IJCNN)*, 1-8, 2010
4. Khayyambashi M., Rizi F., An approach for detecting profile cloning in online social networks, *e-Commerce in Developing Countries: With Focus on e-Security (ECDC)*, 1-12, 2013
5. Blum A., Chawla Sh,. Learning from Labeled and Unlabeled Data using Graph Mincuts, ICML '01 *Proceedings of the Eighteenth International Conference on Machine Learning*, 19-26, 2001
6. Chen Y., Wang G., Dong Sh., Learning with Progressive Transductive Support Vector Machine, *IEEE International Conference Data Mining*, 67-74, 2002
7. Zhu X., Ghahramani Z., Lafferty J., Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions, *Proceedings of the Twentieth International Conference on Machine Learning*, 912-919, 2003
8. Bunke H., Graph Matching: Theoretical Foundations, Algorithms, and Applications, *International Conference on Vision Interface*, Montreal, 82-88, 2000
9. Tarapata Z., Kasprzyk R.: An application of multicriteria weighted graph similarity method to social networks analyzing, *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining*, 366-368, 2009
10. Liben-Nowell, Kleinberg, The Link-Prediction Problem for Social Networks, *Journal of the American Society for Information Science and Technology*, vol.58 (7), 1019-1031, 2007

Streszczenie

Wraz z pojawieniem się internetowych sieci społecznych znaczenie aspektu prywatności w Internecie wzrosło drastycznie. Stąd ważne jest opracowanie mechanizmów, które uniemożliwią osobom niepowołanym dostęp do prywatnych danych osobowych. W pracy podjęta została próba określenia modeli naruszeń prywatności poprzez analizę wpływu struktury sieci oraz jej atrybutów na możliwości naruszenia prywatności w internetowej sieci społecznej. Wynikiem tych działań jest opracowanie koncepcji symulatora pozwalającego na weryfikację wniosków wypływających z utworzonych modeli.

Słowa kluczowe: internetowe sieci społeczne, prywatność, uczenie półnadzorowane

Models of privacy violation detection in online social networks

Summary

With the arrival of online social networks, the importance of privacy on the Internet has increased dramatically. Thus, it is important to develop mechanisms that will prevent our hidden personal data from unauthorized access. In this paper an attempt was made to present some set of privacy violation detection models defined from local – appropriate person personal data – and global point of view – online social network structure. The result of this activities, despite models, is conception of simulator, which will allow us to verify conclusions from the analysis of online social networks privacy violation.

Keywords: online social networks, privacy, semi-supervised learning