УДК 004.6
DOI: http://dx.doi.org/10.21686/1818-4243-2017-6-14-22

В.А. Катаева, И.С. Пантюхин, И.В. Юрин
Университет ИТМО, Санкт-Петербург, Россия, Москва

Метод оценки степени связанности профилей пользователей социальной сети на основе открытых данных

Целью исследования являлось изучение существующих методов определения степени связанности двух пользователей социальной сети, определение их недостатков и разработка нового метода. В ходе исследования были выявлены недостатки существующих методов и предложен новый метод оценки степени связанности профилей социальной сети на основе открытых данных из социальной сети. Под степенью связанности профилей пользователей понимается вероятность связи (взаимодействия) владельцев профилей в реальной жизни, она рассчитывается для двух пользователей социальной сети и выражается в процентах. Работа метода демонстрируется на примере социальной сети «Вконтакте». Данный метод включает в себя последовательность следующих этапов: на первом этапе происходит сбор данных о пользователях социальной сети с помощью API и формирование кортежей признаков профилей пользователей. Кортеж признаков профилей социальной сети – это собранные для каждого из пользователей данные, хранящиеся в структурированном виде. Следующий этап – анализ собранной информации. Для каждого признака из кортежа профилей, т.е. возможного элемента взаимодействия пользователей в социальной сети, рассчитывается коэффициент связанности по признаку. Также для каждого признака рассчитывается его информативность, т.е. на сколько важен тот или иной признак в данной социальной сети. На заключительной этапе происходит формирование результатов с помощью выведенной в процессе исследования формулы вероятности связи двух пользователей. Полученная в результате применения метода вероятность связи двух пользователей может применяться для оптимизации деятельности оперативно-розыскных служб и специальных органов. Также полученная степень связанности двух пользователей может интерпретироваться как вероятность возникновения канала утечки информации между ними. В роли пользователя метода может выступать любая частная или государственная организация, заботящаяся о безопасности корпоративных данных и коммерческой тайне, оперативно-розыскная служба, а также организация, исследующая кибер-преступления и инциденты информационной безопасности.

Ключевые слова: информационная безопасность, метод, информативность, метод накопленных частот, социальная сеть, связь профилей социальной сети, открытые данные, анализ данных.

Valentina A. Kataeva, Igor S. Pantyukhin, Igor V. Yurin
ITMO University, Saint Petersburg, Russia

Estimation method of the cohesion degree for the users' profiles of social network based on open data

The purpose of research was to study the existing methods of determining the degree of cohesion of two users of social network, identifying their shortcomings and developing a new method. The research identified shortcomings of existing methods and proposed a new method for assessing the degree of cohesion of social network profiles based on open data from a social network. Under the degree of cohesion of users' profiles is understood the probability of communication (interaction) of profile owners in real life, it is calculated for two users of the social network and expressed in percent. The work of the method is demonstrated on the example of the social network "In contact". This method includes the sequence of the following stages: the first stage is data collection about users of the social network with API and the formation of tuples of users' profile characteristics. A tuple of characteristics of social network profiles is the data, collected for each user, stored in a structured form.
The next step is the analysis of the collected information. For each characteristic of the tuple of profiles, i.e. the possible element of interaction of users in the social network, the coefficient of cohesion by the characteristic is calculated. In addition, for each feature, its informativeness is calculated, i.e. how important is this or that feature in this social network. At the final stage, the results are generated, using the formula for the probability of communication between two users, derived during the investigation. Obtained as a result of the application of the method, the probability of communication between two users can be used to optimize the activities of the operative-search services and special bodies.
In addition, the received degree of cohesion of two users can be interpreted as the probability of a channel of information leakage between them. The role of the user of the method can be any private or state organization that cares about the security of corporate data and commercial secrets, the operative-search service, as well as an organization that investigates cybercrimes and information security incidents.

Keywords: *information security, method, information, method of accumulated frequencies, social network, and communication of social network profiles, open data, data analysis.*

Введение

С увеличением количества пользователей социальных сетей возникает и увеличение объема информации, связанной с этими пользователям [7]. При анализе такой информации можно узнать не только интересы пользователя, но и круг лиц, с которым различными способами взаимодействует пользователь. Т.к. пользователем социальный сети в наше время может являться сотрудник любой организации, имеющий доступ к конфиденциальным данным, кибер-преступник или просто частное лицо, появляется вероятность утечки информации через социальную сеть. В данном случае источником угрозы утечки информации будет являться пользователь социальной сети, а каналом утечки информации – социальная сеть и всевозможные взаимодействия между ее пользователями. Утечка конфиденциальных данных от одного пользователя социальной сети к другому – проблема, которую необходимо решать. В связи с этим целью исследования является изучение существующих методов оценки связанности пользователей социальной сети, определение их недостатков и разработка критериев для создания нового метода. В настоящее время существует такое понятие, как «Социальный граф», который тесно связан с целью исследования, однако он учитывает только явные социальные связи между пользователями социальной сети. Также Социальный граф не предоставляет оценки найденных с помощью социального графа связей [10] – [11]. Социальный граф может только визуализировать открытые связи (с помощью ребер) между пользователями (т.е. вершинами графа) [1]. В связи с этим, встает вопрос о создании нового метода, который мог бы быстро и удобно предоставлять информацию о степени связанности профилей пользователей без явного «указания дружбы» в различных социальных сетей, но учитывать другие количественные или качественные показатели взаимодействия пользователей, что и является главной целью исследования. Также существует потребность в создании универсального алгоритма анализа данных из социальных сетей, т.е. метода, адаптируемого под большинство популярных социальных сетей. Практической значимостью разработки данного метода является оптимизации и автоматизации деятельности оперативно-розыскных служб и других органов, а также оценка вероятности образования канала утечки конфиденциальной информации (например, корпоративных данных) любой организации через социальные сети.

Описание метода

Метод состоит из следующих этапов:

1. Сбор данных.
2. Анализ данных.
3. Формирование и вывод результатов.

Сбор данных из социальной сети осуществляется с помощью открытого API (набор готовых функций, методов, классов и т.д. для написания приложений) социальной сети. У большинства популярных сетей, таких как «Вконтакте», «Facebook», «Instagram», «Twitter» и др. имеется открытое API, что существенно упрощает сбор данных и позволяет разрабатываемому методу быть адаптированным под различные социальные сети. Входными данными для анализа являются два профиля пользователей социальной сети. Профиль пользователя – это определенный набор (кортеж) признаков. Во многих социальных сетей этот кортеж имеет схожую структуру. Например, в «Facebook», «Вконтакте» и «Одноклассники» – у пользователя есть список его друзей в открытом доступе. В социальной сети «Друзья» – это люди, об изменениях и обновлениях в профилях которых сообщается пользователю, который находится в друзьях у другого пользователя. Аналогично, в «Instagram» – есть подписчики пользователя (люди, которые подписаны на обновления пользователя) и подписки пользователя (люди, на обновления которых подписан пользователь). В «Вконтакте» и «Facebook» так же есть подписки (люди, сообщества или публичные страницы, на обновления которых подписан пользователь). Практически во всех социальных сетях есть возможность оставить комментарий под каким либо объектом на странице другого пользователя, а также поставить отметку «Мне нравится». Многие социальные сети имеют такую функцию, как «отметка» геолокации, т.е. пользователь может прикрепить к какому-либо объекту (элементу) социальный сети метку геолокации. Также во всех социальных сетях пользователю предлагается заполнить поля общей информации, личной и контактной информации.

В данном методе решено было анализировать 4 основных элемента (признака), имеющихся в профилях пользователей большинства социальных сетей, анализ которых необходим в определении степени связанности двух пользователей, это:

– Друзья.

– Группы (публичные страницы).

– Комментарии.

– Отметки «Мне нравится».

После сбора данных, т.е. формирования кортежа признаков обоих профилей, данные обрабатываются с помощью математических функций. Т.к. после первого этапа данные уже были организованы в кортеж признаков профиля, то метод поочередно анализирует

каждый из признаков кортежа. После проведения алгоритмов обработки каждого признака рассчитывается определенный коэффициент этого признака, называемый «Коэффициент связанности по признаку» (S_i). За основу расчета коэффициента связанности по признаку была взята формула расчета вероятности, где положительные исходы события делятся на всевозможные исходы события. В каждом признаке положительные исходы – это связи, подтверждающие взаимодействия пользователей между собой, а все возможные исходы – это все связи, имеющиеся в профилях пользователей социальной сети. Рассчитывать коэффициенты вышеописанным способом было решено, т.к. этот коэффициент учитывает не только количество связей (взаимодействия пользователей или объекты, подтверждающих связь между пользователями), но и полную социальную активность пользователя. В данном случае под социальной активностью подразумевается насколько полно профиль социальной сети отражает активность пользователя в реальной жизни, а также насколько часто и тщательно пользователь использует социальную сеть для общения с людьми, общения в социальных группах или в любых других целях (рекламы, работы и т.д.). Ниже представлены формулы расчета коэффициентов связанности по каждому признаку S_i:

1. Признак «Друзья» (S_1).

Для расчета коэффициента связанности по признаку «Друзья» в методе используется список друзей каждого пользователя и сформированный список общих друзей двух пользователей. Формула расчета:

$$S_1 = \frac{n_{fr}}{N_{fr}}, \qquad (1)$$

где n_{fr} – количество общих друзей двух пользователей;

N_{fr} – количество друзей двух пользователей без повторений.

2. Признак «Группы» (S_2).

Для расчета коэффициента связанности по данному признаку используется формула, аналогичная формуле (1):

$$S_2 = \frac{n_{gr}}{N_{gr}}, \qquad (2)$$

где n_{gr} – количество общих групп (публичных страниц) двух пользователей;

N_{gr} – количество групп (публичных страниц) двух пользователей без повторений.

Однако, при анализе признака «Группы» нужно учитывать один важный момент – это количество человек, состоящих в группе. Чем больше количество человек, состоящих в группы – тем меньше вероятность связи между ее участниками в реальной жизни, и аналогично наоборот. Экспериментально было получено, что группа или публичная страница с количеством участников (подписчиков) более 700 человек никак не подтверждает связь ее участников (подписчиков) в реальной жизни. В связи с этим, при обнаружении общей группы или публичной страницы с количеством участников менее 700 человек, решено было использовать другой расчет коэффициента связанности по признаку:

$$S_2 = F(X_i) = \sqrt[100]{\left(\frac{e}{2}\right)^{-X_i}}, \qquad (3)$$

где $F(X_i)$ – функция экспоненциального распределения;

X_i – количество участников в i-ой общей группе двух пользователей [4].

На рис. 1 приведен график функции $F(X_i)$ при $0 < X < 700$. Значения по оси Y соответствуют значениям коэффициентов связанности по признаку «Группы (публичные страницы)» (S_2). Значения по оси Y уже являются нормированными (от 0 до 1) и отражают вероятность связанности двух пользователей с наличием об-

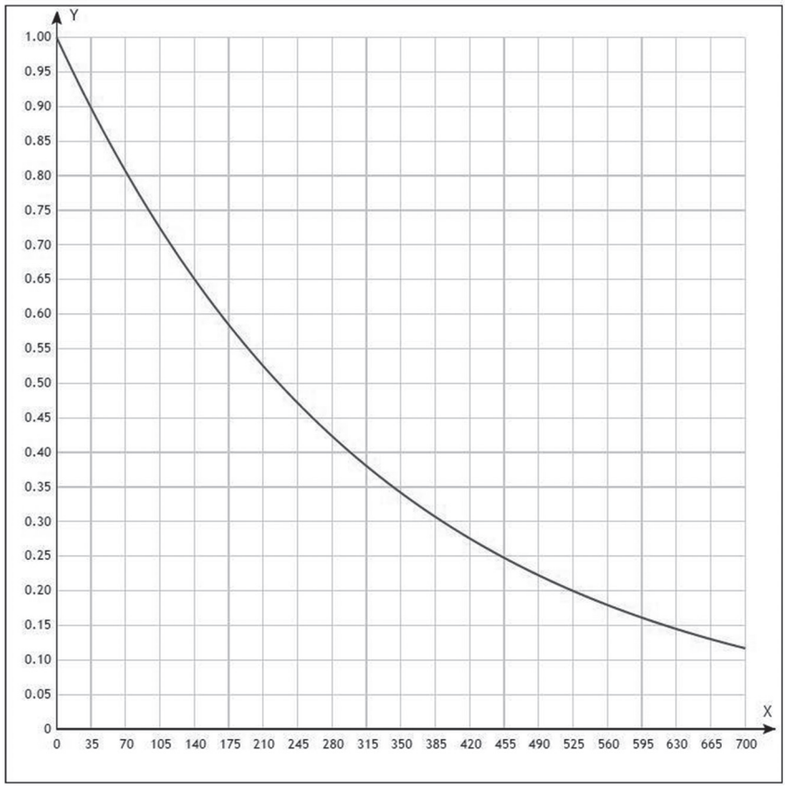

Рис 1. Распределение зависимости коэффициента связанности по признаку S_2 от количества участников группы или публичной страницы

щей группы или публичной страницы с небольшим (< 700) количеством участников.

Соответственно, если при анализе не обнаружены общие группы (публичные страницы) с небольшим количеством участников – коэффициент S_2 считается по Формуле (2), если же такие группы нашлись, выбирается группа с наименьшим количеством участников и коэффициент S_2 считается по Формуле (3). Таким образом, при анализе признака «Группы» учитывается не только наличие общих групп и публичных страниц, но и важность общей группы или публичной страницы.

3. Признак «Комментарии» (S_3).

Т.к. комментарии могут быть оставлены первым пользователем на странице второго, и вторым пользователем на странице первого, то оба профиля поочередно проверяются на наличие комментариев. Для расчета коэффициента связанности по признаку «Комментарии» (S_3) сначала рассчитывается коэффициент S_{31} по формуле:

$$S_{31} = \frac{d_1}{Nk_1}, \quad (4)$$

где d_1 – количество комментариев первого пользователя на странице (стене) второго пользователя;

Nk_1 – количество всех комментариев на странице (стене) второго пользователя.

Далее рассчитывается коэффициент S_{32} по формуле:

$$S_3 = \frac{d_2}{Nk_2}, \quad (5)$$

где d_2 – количество комментариев второго пользователя на странице (стене) первого пользователя;

Nk_2 – количество всех комментариев на странице (стене) первого пользователя.

Для получения конечного коэффициента связанности по признаку «Комментарии» (S_3) выбирается максимальный полученный коэффициент из S_{31} и S_{32}:

$$S_3 = \max(S_{31};\ S_{32}) \quad (6)$$

4. Признак «Отметка «Мне нравится» (S_4).

Коэффициент связанности по данному признаку рассчитывается аналогично предыдущему:

$$S_{41} = \frac{e_1}{Nl_1}, \quad (7)$$

где e_1 – количество отметок «Мне нравится» первого пользователя на странице (стене, под фотографиями профиля) второго пользователя;

Nl_1 – количество всех отметок «Мне нравится» на странице (стене, под фотографиями профиля) второго пользователя.

Далее рассчитывается коэффициент S_{32} по формуле:

$$S_{42} = \frac{e_2}{Nl_2}, \quad (8)$$

где e_2 – количество отметок «Мне нравится» второго пользователя на странице (стене, под фотографиями профиля) первого пользователя;

Nl_2 – количество всех отметок «Мне нравится» на странице (стене, под фотографиями профиля) первого пользователя.

Для получения конечного коэффициента связанности по признаку «Комментарии» (S_4) выбирается максимальный полученный коэффициент из S_{41} и S_{42}:

$$S_4 = \max(S_{41};\ S_{42}) \quad (9)$$

Для расчета вероятности связи $P(com)$ двух пользователей в методе суммируются рассчитанные для каждого признака коэффициенты связанности по признаку S_i (Формулы 1–9). Однако, для расчета вероятности связи нужно учитывать еще один показатель – на сколько важен тот или иной признак в данной социальной сети, т.е. учитывать его информативность $I(S_i)$ [2].

Для расчета информативности признака был выбран Метод накопленных частот (МНЧ) [18–20], т.к. результаты, полученные с помощью данного метода, удобнее всего интерпретировать в данном случае: они являются ненормированными и о большей или меньшей информативности того или иного признака говорится в относительном плане – более высокая или более низкая по сравнению с информативностью другого признака [3]. Для расчета информативности $I(S_i)$ с помощью МНЧ необходимо иметь две выборки распределения признака $(x_i)S_i$, принадлежащим двум различным классам. Выборками по каждому признаку являются связи и взаимодействия пользователей в социальной сети, т.е. общие друзья, общие группы и публичные страницы, комментарии и отметки “Мне нравится”. У автора данной работы имеется профиль в социальной сети, и он опытным путём может обозначить два класса взаимодействия себя и других пользователей социальной сети: одногруппники (A_1) и одноклассники (A_2). Для каждого признака S_i было получено 10 значений признака x_i для объектов (пользователей) из двух выбранных классов: A_1 и A_2. Ниже в табл. 1 приведен расчет информативности $I(S_1)$ признака «Друзья» (S_1):

В табл. 1 приведены значения распределения признака «Друзья» для двух выбранных классов: $x_{1,i}$ –количество об-

Таблица 1

Значения распределения признака «Друзья» $(x_i)S_1$

Номер объекта i		1	2	3	4	5	6	7	8	9	10
Класс A_1	$x_{1,i}$	82	65	30	10	28	52	111	7	83	13
Класс A_2	$x_{2,i}$	22	41	11	17	127	16	37	43	25	47

Таблица 2

Расчет информативности признака «Друзья» с помощью МНЧ

Интервалы	Класс A_1 Частоты $m_{1,j}$	Накопленные частоты $M_{1,j}$	Класс A_2 Частоты $m_{2,j}$	Накопленные частоты $M_{2,j}$	$\lvert M_{1,j} - M_{2,j}\rvert$
0–20	3	3	3	3	0
20–40	2	5	3	6	1
40–60	1	6	3	9	3
60–80	1	8	0	9	1
80–100	2	10	0	9	1
100–120	1	10	0	9	1
120–127	0	10	1	10	0

щих друзей с каждым объектом из класса A_1, $x_{2,i}$ – количество общих друзей с каждым объектом из класса A_2.

Оценкой информативности является модуль максимальной разности накопленных частот. Расчет накопленных частот и определение информативности для признака «Друзья» приведен в табл. 2:

Как видно из табл. 2 максимальным модулем разности накопленных частот является число 3, следовательно, информативность признака «Друзья» равна 3. Для остальных признаков (S_2, S_3, S_4) был произведен аналогичный расчет информативности с выборками из тех же классов (A_1 и A_2) методом накопленных частот.

В процессе анализа данных была получена следующая формула вероятности связи двух пользователей $P(com)$:

$$P(com) = \sum_{1}^{n} (S_i \cdot I(S_i)), \quad (10)$$

где n – количество анализируемых признаков социальной сети;
S_i – коэффициент связанности по i-му признаку;
$I(S_i)$ – информативность i-го признака;
i – номер признака; $1 < i < n$.

Формула вероятности связи является количественной оценкой связанности двух профилей пользователей социальной сети на основе открытых данных из этой социальной сети. Значения вероятности связи $P(com) \in [0;1]$. При выводе значение вероятности связи домножается на 100, чтобы вероятность связи была представлена в процентной шкале.

Результаты

Для программной реализации метода были выбраны следующие средства: язык программирования – Python, фреймфорк для разработки веб-приложения – Flask. Пользовательская часть создана с помощью стандартных средств веб-разработки (HTML, CSS, JS). В данном методе используется множество различных стандартных и встроенных библиотек. В большей степени используется библиотека vk_api.

Входными данными для работы методы являются два профиля пользователей социальной сети «Вконтакте». Для проверки работоспособности метода решено было использовать тестовую выборку из пользователей социальной сети «Вконтакте» в количестве 1000 человек. Все выбранные для проверки пользователи находятся в списке друзей эксперта (автора). Тестовая выборка была сформирована данным образом, т.к. эксперт эмпирически может подтвердить или опровергнуть определенную с помощью данного метода вероятность связи. Для проверки точности метода было разработан вспомогательный алгоритм, который автоматически создает список из *id* проверяемых пользователей, и поочередно рассчитывает коэффициенты связанности по признакам и конечную вероятность связи для каждого проверяемого пользователя и автора. Результат заносится в текстовый файл, который в дальнейшем проверяется на наличие ошибок, т.е. верную и неверную (слишком/недостаточно большую или слишком/недостаточно малую) вероятность связи.

В результате поочередной проверки 1000 пользователей на определение вероятности связи с одним неизменяемым пользователем (автором), было получено:

– 209 пользователей (20,9%) – метод верно подтвердил достаточную вероятность связи;

– 658 пользователей (65,8%) – метод верно определил малую вероятность связи;

– 98 пользователей (9,8%) – метод не подтвердил потенциально существующую связь.

– 35 пользователей (3,5%) – метод подтвердил потенциально несуществующую связь.

Под достаточной вероятностью связи подразумевается связь > 40%. Под малой вероятностью связи подразумевается связь < 40%. Критерий достаточной вероятности связи введен экспертом и может быть изменен в зависимости от цели анализа, т.е. полученная в результате работы метода вероятность связи может трактоваться по-разному (например, наличие даже малого процента вероятности связи может приравнивается к наличию канала утечки информации). В случаях, когда метод не подтвердил существующую связь – отсутствуют или являются непоказательными все взаимодействия пользователей. В случаях, когда метод подтвердил потенциально несуществующую связь – у анализируемых пользователей имелась в наличие общая группа с небольшим числом участников. Однако, расценить связь, как несуществующую, возможно только опытными

Таблица 3

Ошибки первого и второго рода для тестовой выборки

		Верное предположение	
		H_0	H_1
Результат применения метода	H_0	H_0 верно принято 20,9%	H_1 неверно принято (ошибка второго рода) 3,5%
	H_1	H_1 неверно отвергнуто (ошибка первого рода) 9,8%	H_0 верно отвергнуто 65,8%

путем, и при отсутствии эксперта, в данном случае автора, эта связь считалась бы верно подтвержденной.

В табл. 3 приведен разбор ошибок первого и второго рода, выявленных при проверке тестовой выборки. H_0 – предположение, согласно которому метод верно подтверждает связь, H_1 – предположение, согласно которому метод верно не подтверждает связь. Таким образом, с помощью проверки разработанного метода с помощью ошибок первого и второго рода было получено, что данный метод способен определить вероятность связи профилей пользователей социальной сети на основе открытых данных с точностью 86,7% [6].

Эффективность метода при оценке каналов утечки информации

Канал утечки информации – методы (способ, путь) утечки информации из информационной системы; последовательность носителей информации, один или несколько из которых могут быть нарушителем или его специальной техникой [5]. Основными элементами описания угроз утечки информации являются: источник угрозы, среда распространения информации и носитель защищаемой информации. Технический канал утечки информации – среда распространения информации и носитель защищаемой информации совместно с техническими средствами, с помощью которых может быть получена защищаемая информация [9]. Пользователь социальной сети в данном случае будет являться источником угрозы утечки информации и носителем информации одновременно. Каналы утечки информация классифицируются по различным признакам, например по принципам функционирования, физическим свойствам, уровню доступа к системе и т.д.

Каналом утечки информации может являться несанкционированный обмен информацией между физическими лицами (например, между персоналом организации), т.к. человек является одним из основных источников угроз утечки информации, помимо физических (бумажных) носителей информации, технических средств хранения и обработки информации, специальных средств коммуникации (средств передачи информации) и конфиденциальных сообщений, передаваемых непосредственно по каналам связи.

Говоря об источнике утечки информации и подразумевая под ним физическое лицо нужно отметить, что главными факторами образования соответствующего канала утечки информации является человеческий фактор и несоблюдения правил обращения с конфиденциальной информацией. Под оценкой канала утечки информации подразумевается получение вероятности образования канала утечки информации. Эта вероятность пропорциональна вероятности связи анализируемых пользователей, полученной в результате применения разработанных алгоритмов обработки пользовательских данных. Применив разработанный метод пользователь метода получит оценку вероятности утечки информации от одного лица к другому. Если обнаружена большая вероятность утечки (вероятность связи), то, возможно, потребуется последующий анализ, и далее могут быть приняты меры для получения доступа к закрытым данным. Также, к примеру, если имеются какие-либо подозрения в отношении определенного лица (например, сотрудника), можно проанализировать его взаимодействие с полным списком его друзей.

В роли пользователя метода может выступать любая частная или государственная организация, заботящаяся о безопасности корпоративных данных и коммерческой тайне, оперативно-розыскная служба, а также любая организация, исследующая кибер-преступления и инциденты информационной безопасности [12–17].

Важным моментом в утечке информации через социальные сети является то, что информация, переданная или опубликованная в социальных сетях, в будущем не может быть удалена оттуда, даже если пользователь якобы «удалил» ее со своей страницы. Это связано с тем, что вся информацию, медиа-контент, данные геолокации, переписки и прочие объекты социальной сети сохраняются и не всегда удаляются «владельцами» социальной сети. Т.е. удаление пользователем, например, отправленного сообщения не гарантирует полного удаления данных этого сообщения с серверов социальной сети. И, следовательно, если присутствует факт передачи какой-либо информации через социальную сеть, присутствует и факт ее перехвата, поэтому обсуждение между работниками организации текущих «дел» фирмы через социальную сеть часто

становится источником дохода для фирм-конкурентов [8].

Таким образом, данный метод оптимизирует деятельность по определению потенциально возможных каналов утечки информации, которые образуются при использовании социальных сетей человеком, являющимся одновременно источником угрозы утечки информации и носителем конфиденциальной информации.

Заключение

В настоящее время в социальных сетях хранится большой объем открытых пользовательских данных. Анализ этих данных может существенно помочь в определении потенциально возможных каналов утечки информации, источником которых могут быть пользователи социальной сети, связанные определенным образом с конфиденциальной информацией. Также анализ данных из социальных сетей может быть полезен в оптимизации деятельности оперативно-розыскных служб и других специальных органов, т.к. при анализе пользовательских данных можно получить круг возможных знакомых анализируемого пользователя. Главной целью исследования являлась разработка метода оценки степени связанности профилей пользователей социальных сетей на основе открытых данных, и цель была достигнута:

1. Был проведен обзор существующих методов оценки степени связанности профилей пользователей социальных сетей на основе открытых данных, определены недостатки существующих методов для достижение цели и сформулирован набор требований для создания нового метода оценки степени связанности профилей пользователей социальных сетей.

2. Был разработан метод оценки степени связанности профилей пользователей социальных сетей на основе открытых данных, а также проверена его работоспособность на примере социальной сети «Вконтакте». Точность метода была проверена на тестовой выборке, состоящей из 1000 профилей пользователей социальной сети «Вконтакте» и составила 86,7%. Разработанный метод может быть адаптирован под другие социальные сети.

Литература

1. Алексеев В.Е., Теория графов. Электронное учебно-методическое пособие. Нижний Новгород: ННГУ им. Лобачевского, 2012. 60 с.

2. Бессонова Е.Е., Метод идентификации пользователей в сети Интернет с использованием компонентного профиля // Материал диссертационной работы на соискание ученой степени кандидата технических наук. Санкт-Петербург, 2014. 115 с.

3. Голованова И.С., Выбор информативных признаков. Оценка информативности // Методические указания к лабораторной работе по дисциплине «Методы обработки биомедицинских данных». Томск, ТПУ. 2003. 18 с.

4. Тарасов В.Н., Экспоненциальный закон распределения. Математическая теория надежности. Учебно-методический комплекс по дисциплине «Математическая теория надежности». Самара, ПГУТИ. 2012. 204 с.

5. Каналы утечки информации. Википедия: свободная энциклопедия / URL: https://ru.wikipedia.org/wiki/Каналы_утечки_информации (дата обращения: 11.04.2016)

6. Проверка статистических гипотез. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализ данных / URL: http://www.machinelearning.ru/wiki/index.php?title=Проверка_статистических_гипотез# (дата обращения: 20.04.2016)

7. Социальная сеть. Википедия: свободная энциклопедия / URL: https://ru.wikipedia.org/wiki/Социальная_сеть (дата обращения: 17.03.2016).

References

1. Alekseev V.E., Teoriya grafov. Elektronnoe uchebno-metodicheskoe posobie. Nizhniy Novgorod: NNGU im. Lobachevskogo, 2012. 60 p. (In Russ.)

2. Bessonova E.E., Metod identifikatsii pol'zovateley v seti Internet s ispol'zovaniem komponentnogo profilya. Material dissertatsionnoy raboty na soiskanie uchenoy stepeni kandidata tekhnicheskikh nauk. Sankt-Peterburg, 2014. 115 p. (In Russ.)

3. Golovanova I.S., Vybor informativnykh priznakov. Otsenka informativnosti. Metodicheskie ukazaniya k laboratornoy rabote po distsipline «Metody obrabotki biomeditsinskikh dannykh». Tomsk, TPU. 2003. 18 p. (In Russ.)

4. Tarasov V.N., Eksponentsial'nyy zakon raspredeleniya. Matematicheskaya teoriya nadezhnosti. Uchebno-metodicheskiy kompleks po distsipline «Matematicheskaya teoriya nadezhnosti». Samara, PGUTI. 2012. 204 p. (In Russ.)

5. Kanaly utechki informatsii. Vikipediya: svobodnaya entsiklopediya / URL: https://ru.wikipedia.org/wiki/Kanaly_utechki_informatsii (accessed: 11.04.2016) (In Russ.)

6. Proverka statisticheskikh gipotez. Professional'nyy informatsionno-analiticheskiy resurs, posvyashchennyy mashinnomu obucheniyu, raspoznavaniyu obrazov i intellektual'nomu analiz dannykh / URL: http://www.machinelearning.ru/wiki/index.php?title=Proverka_statisticheskikh_gipotez# (accessed: 20.04.2016) (In Russ.)

7. Sotsial'naya set'. Wikipedia: svobodnaya entsiklopediya / URL: https://ru.wikipedia.org/wiki/Sotsial'naya_set' (accessed: 17.03.2016) (In Russ.).

8. Утечка корпоративных данных через социальные сети. Пресс-центр компании Serchinform / URL: http://searchinform.ru/press/articles/777/ (дата обращения 20.04.2016)

9. Базовая модель угроз безопасности персональных данных при их обработке в информационных системах персональных данных (выписка) // Федеральная служба по техническому и экспортному контролю. М.: 2008. 69 с.

10. Mislove A., Measurement and Analysis of Online Social Networks / Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi // Proceedings of the 5th ACM/USENIX Internet Measurement Conference. 2007. P. 29–42.

11. William H. Hsu, Structural Link Analysis from User Profiles and Friends Networks: A Feature Construction Approach / William H. Hsu, Joseph Lancaster, Martin S.R. Paradesi, Tim Weninger // International Conference on Weblogs and Social Media. 2007.

12. Пантюхин И.С., Зикратов И.А. Методика проведения постинцидентного внутреннего аудита средств вычислительной техники // Научно-технический вестник информационных технологий, механики и оптики. 2017. Т. 17. № 3. С. 467–474. doi: 10.17586/2226-1494-2017-17-3-467-474

13. Пантюхин И.С., Зикратов И.А., Левина А.Б. Метод проведения постинцидентного внутреннего аудита средств вычислительной техники на основе графов // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 3. С. 506–512. doi: 10.17586/2226-1494-2016-16-3-506-512

14. Юрасов Д.С., Зикратов И.А. Различение пользователей на основе их поведения в сети Интернет // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 6 (88). С. 148–151.

15. Бессонова Е.Е., Зикратов И.А., Росков В.Ю. Анализ способов идентификации пользователей в сети Интернет // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 6 (82). С. 128–130.

16. Воробьева А.А. Методика идентификации интернет-пользователя на основе стилистических и лингвистических характеристик коротких электронных сообщений // Информация и космос. 2017. № 1. С. 127–130.

17. Воробьева А.А. Анализ возможности применения различных лингвистических характеристик для идентификации автора анонимных коротких сообщений в глобальной сети Интернет // Информация и космос. 2014. № 1. С. 42–46.

18. В.В. Быкова, А.В. Катаева Методы и средства анализа информативности признаков при обработке медицинских данных // Программные продукты и системы. 2016. № 2 (114). С. 172–178.

19. Shannon, C.E. The Mathematical Theory of Communication / C.E.Shannon and W.Weaver.

8. Utechka korporativnykh dannykh cherez sotsial'nye seti. Press-tsentr kompanii Serchinform / URL: http://searchinform.ru/press/articles/777/ (accessed 20.04.2016) (In Russ.)

9. Bazovaya model' ugroz bezopasnosti personal'nykh dannykh pri ikh obrabotke v informatsionnykh sistemakh personal'nykh dannykh (vypiska). Federal'naya sluzhba po tekhnicheskomu i eksportnomu kontrolyu. Moscow: 2008. 69 p. (In Russ.)

10. Mislove A., Measurement and Analysis of Online Social Networks / Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi. Proceedings of the 5th ACM/USENIX Internet Measurement Conference. 2007. P. 29–42.

11. William H. Hsu, Structural Link Analysis from User Profiles and Friends Networks: A Feature Construction Approach / William H. Hsu, Joseph Lancaster, Martin S.R. Paradesi, Tim Weninger. International Conference on Weblogs and Social Media. 2007.

12. Pantyukhin I.S., Zikratov I.A. Metodika provedeniya postintsidentnogo vnutrennego audita sredstv vychislitel'noy tekhniki. Nauchno-tekhnicheskiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki. 2017. T. 17. No. 3. P. 467–474. doi: 10.17586/2226-1494-2017-17-3-467-474 (In Russ.)

13. Pantyukhin I.S., Zikratov I.A., Levina A.B. Metod provedeniya postintsidentnogo vnutrennego audita sredstv vychislitel'noy tekhniki na osnove grafov. Nauchno-tekhnicheskiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki. 2016. Vol. 16. No. 3. P. 506–512. doi: 10.17586/2226-1494-2016-16-3-506-512 (In Russ.)

14. Yurasov D.S., Zikratov I.A. Razlichenie pol'zovateley na osnove ikh povedeniya v seti Internet. Nauchno-tekhnicheskiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki. 2013. No. 6(88). P. 148-151 (In Russ.)

15. Bessonova E.E., Zikratov I.A., Roskov V.Yu. Analiz sposobov identifikatsii pol'zovateley v seti Internet. Nauchno-tekhnicheskiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki. 2012. No. 6(82). P. 128–130 (In Russ.).

16. Vorob'eva A.A. Metodika identifikatsii internet-pol'zovatelya na osnove stilisticheskikh i lingvisticheskikh kharakteristik korotkikh elektronnykh soobshcheniy. Informatsiya i kosmos. 2017. No. 1. P. 127–130 (In Russ.).

17. Vorob'eva A.A. Analiz vozmozhnosti primeneniya razlichnykh lingvisticheskikh kharakteristik dlya identifikatsii avtora anonimnykh korotkikh soobshcheniy v global'noy seti Internet. Informatsiya i kosmos. 2014. No. 1. P. 42–46 (In Russ.).

18. V.V. Bykova, A.V. Kataeva Metody i sredstva analiza informativnosti priznakov pri obrabotke meditsinskikh dannykh. Programmnye produkty i sistemy. 2016. No. 2 (114). P. 172–178 (In Russ.).

19. Shannon, C.E. The Mathematical Theory of Communication / C.E.Shannon and W.Weaver.

Urbana, IL: University of Illinois Press. ISBN: 0252725484. 1963. 144 p.

20. Глазкова А.В., Математическое моделирование классификации объектов (на примере определения категории поненциальных адресатов текста) // Материал диссертационной работы на соискание ученой степени кандидата технических наук. Тюмень. 2016. 141 с.

Urbana, IL: University of Illinois Press. ISBN: 0252725484. 1963. 144 p.

20. Glazkova A.V., Matematicheskoe modelirovanie klassifikatsii ob»ektov (na primere opredeleniya kategorii ponentsial'nykh adresatov teksta). Material dissertatsionnoy raboty na soiskanie uchenoy stepeni kandidata tekhnicheskikh nauk. Tyumen'. 2016. 141 p. (In Russ.)

Сведения об авторах

Валентина Алексеевна Катаева
Магистрант
Университет ИТМО, Санкт-Петербург, Россия
Эл. почта: kataeva@cit.ifmo.ru

Игорь Сергеевич Пантюхин
Ассистент
Университет ИТМО, Санкт-Петербург, Россия
Эл. почта: zevall@cit.ifmo.ru

Игорь Валентинович Юрин
К.воен.н., доцент
Университет ИТМО, Санкт-Петербург, Россия
Эл. почта: 9402015@mail.ru

Information about the authors

Valentina A. Kataeva
Master student
ITMO University, Saint-Petersburg, Russia
E-mail: kataeva@cit.ifmo.ru

Igor S. Pantyukhin
Assistant
ITMO University, Saint Petersburg, Russia
E-mail: zevall@cit.ifmo.ru

Igor V. Yurin
Cand. Sci. (Military), Assistant professor
ITMO University, Saint Petersburg, Russia
E-mail: 9402015@mail.ru