ОБНАРУЖЕНИЕ КИБЕРУГРОЗ С ПОМОЩЬЮ АНАЛИЗА СОЦИАЛЬНЫХ СЕТЕЙ

Людмила Кириченко, Тамара Радивилова, Алексей Барановский

***Аннотация**: В работе проведен обзор основных методов анализа социальных сетей, которые используются для обнаружения киберугроз. Представлены основные типы угроз в социальных сетях и описаны некоторые методы защиты для их предотвращения. Рассмотрены типичные задачи анализа социальных сетей, направленные на выявление киберугроз, такие как обнаружение сообществ в сети, обнаружение лидеров и экспертов в сообществах, анализ стабильности сообществ, кластеризация текстовой информации и другие. Описаны основные классы методов теории графов и интеллектуального анализа данных, которые широко применяются в анализе социальных сетей. Показано применение методов фрактального анализа для исследования рядов показателей активности пользователей сетей.*

***Ключевые слова**: анализ социальных сетей, киберугрозы, data mining, методы обнаружения лидеров, методы обнаружения экспертов, фрактальный анализ.*

***ITHEA Keywords**: G.3 Probability and statistics - Time series analysis, Stochastic processes, G.1 Numerical analysis, G.1.2 Approximation - Wavelets and fractals, I.2 Artificial intelligence - I.2.0 General, K.6 Management of computing and information systems – K.6.5 Security and Protection.*

Введение

Быстрое развитие социальных сетей и способность собирать информацию из них привели к заметному повышению интереса к анализу социальных сетей (АСС) и появлению новых методов, которые становятся все более популярными и используются в различных областях: для поиска экспертов, набора команды специалистов, социальных рекомендаций, маркетинга, коммуникаций, рекламы и многих других. В настоящее время АСС используется для изучения ряда экономических и организационных явлений и процессов, для борьбы с отмыванием денег, кражей личности, онлайн-мошенничеством, кибер-атаками, при расследовании незаконных операций с ценными бумагами и инвестициями, для предотвращения беспорядков и др. [Carley, 2002, Stohl, 2007, Easley, 2010, Russell, 2011].

Социальные сети все более и более широко используются в интересах информационного и психологического воздействия. Они предоставляют возможности с точки зрения влияния на формирование общественного мнения, принятие политических, экономических и военных решений, влияют на информационные ресурсы противника и распространение специально подготовленной информации (дезинформации) [Додонов, 2013, 2014]. Исходя из этого многие страны создают национальные центры кибербезопасности и документы по стратегии кибербезопасности [National Cyber Security Strategy 2016-2021 for United Kingdom, National cyber security strategy for Ukraine, The NATO Cooperative Cyber Defence Centre].

Таким образом, задача сбора информации, мониторинга и анализа социальных сетей для обеспечения информационной безопасности важна и актуальна. Целью данной работы является обзор и анализ основных задач и методов анализа социальных сетей, используемых для обнаружения, предотвращения и борьбы с угрозами в социальных сетях.

Угрозы в социальных сетях и профилактические меры

В наши дни в социальных сетях существует целый ряд различных угроз. В [Palo, 2017] описаны угрозы для социальной сети деловых кругов, в [Carley, 2002; Дзюндзюк, 2011; Матвиенко, 2011; Shantanu, 2017] описаны угрозы для социальных сетей и некоторые профилактические меры.

Спам в социальных сетях. Спам является одним из самых классических атак всех времен. Злоумышленниками генерируются фиктивные учетные записи, и тысячи запросов друзей автоматически отправляются в надежде, что кто-то их примет. После принятия атакующий может начать отправлять спам-сообщения. Даже запрос на соединение с другом позволяет отправлять короткие сообщения внутри него без какого-либо предыдущего соединения между пользователями. Для борьбы со спамом многие сообщества внедряют тесты CAPTCHA, которые решают, когда отправляется слишком много сообщений. Это должно остановить или, по крайней мере, замедлить автоматическое распространение сообщений. Кроме того, у злоумышленников часто есть возможность использовать несколько учетных записей параллельно, пока каждая из них не блокируется дневным лимитом. Большинство социальных сетей предлагают функцию для обозначения сообщений как спама и блокирования их в будущем, что помогает при условии не частого переключения записей злоумышленниками.

Угрозы социальной инженерии. Самая популярная тактика для киберпреступников. Социальные сети позволяют злоумышленникам находить конфиденциальную информацию, которая может быть использована для имущественного и морального ущерба.

Размещение приманок в социальных сетях. Идея проста: использовать ключевые слова и ссылки таким образом, чтобы спам-сообщения получали лучший список учетных записей. Некоторые нападавшие даже начали манипулировать доброжелательными сообщениями Twitter перед пересылкой. Злоумышленники ищут новые сообщения, содержащие горячие ключевые слова. Это может быть сообщение о сомнительной цели офсайда в последнем футбольном матче, с сокращенным URL-адресом, связанным с соответствующей новостной статьей. Затем мошенник принимает это сообщение, заменяет исходный сокращенный URL своей собственной ссылкой, указывающей на вредоносный сайт, и повторно читает сообщение. Это делает практически невозможным для обычных посетителей отличать хорошие и вредоносные сообщения. Следовательно, шансы, что невинный пользователь, который что-то ищет, наткнется на злонамеренную ссылку относительно высоки. Конечно, мы также видим типичные заманчивые сообщения, предлагающие ссылки на видео голых знаменитостей или взломанные программные средства, рассылаемые спамом в надежде, что кто-то найдет и нажмет на них.

Друзья. Доверие к тем, кто входит в список «друзей», всегда выше, чем случайным людям. С одной стороны, это хорошо, поскольку формируется лояльная аудитория вокруг компании, бренда или человека. С другой стороны, это возможность для злоумышленников.

Олицетворение друзей. Почти во всех социальных сетях олицетворение является реальной проблемой для всех. Поскольку сообщение, кажется, исходит от учетной записи друга, люди склонны доверять ему. Это присущее доверие и обычное любопытство приводит к высокой скорости кликов на вредоносных ссылках, что делает атаки получения пароля очень успешными. Эти сообщения об обновлениях часто содержат ссылки на другие вредоносные сайты, чтобы получить больше паролей к учетным записям.

Возможность замены человека или маскарад: наверняка, не совсем ясно, кто скрывает свои действия за именем друзей или прячется за фотографиями друзей в профиле социальной сети. Возможно по IP-адресу отправителя собрать о нем по крайней мере некоторую информацию в корреспонденции по электронной почте, которая не работает в социальной сети. Этот маскарад

возможен и на корпоративном уровне. Результатом такого вредоносного сценария может быть фишинг, организация «черного PR» или «AntiPR». Уже было множество случаев, когда было непонятно, кто создал сайт от имени какой-либо компании, и это создает проблему для оригинального бренда.

Кража паролей и фишинг. Для аутентификации в социальных сетях, использующих пароли, достаточно знать последовательность символов, и станет возможным отправлять рекламу, некоторую информацию от имени других или мотивировать получателей на любые негативные действия, в частности, на передачу ссылок на вредоносный сайт или запускать вредоносный код и выполнять другие (часто нелегальные) действия. Кроме того, некоторые компании используют социальную сеть для продвижения своих собственных продуктов, а кража паролей группы администраторов позволяет украсть саму группу. А для получения конфиденциальной информации традиционно используются фишинг, фиктивные сайты, социальная инженерия и многие другие. Защита от этих методов атаки рассматривается как DLP-система (Data Loss Prevention) и технологии репутации, которые интегрированы в различные антивирусные продукты.

Использование сервисов сокращения URL-адресов. В последние годы особенно популярны услуги сокращения URL-адресов, позволяющие маскировать нежелательный адрес сайта под короткой ссылкой. Фактически, домен перенаправляет посетителя. Сегодня идет активная борьба с этими рисками – служба сокращения URL-адресов начала использовать усовершенствованные механизмы для обнаружения спама и других угроз. Однако для пользователей социальных сетей эта угроза сохраняется: соблазнительные сообщения и предложения от уже известных контактов, которые были взломаны, часто приводят к загрузке вредоносного программного обеспечения или отображению нежелательных веб-страниц.

Использование тех же имен пользователей и паролей в корпоративной сети и внешних социальных ресурсах – эта атака также известна как «Daisy Chain». В результате взлом профилей социальной сети пользователей значительно повышает риск проникновения на корпоративные ресурсы от имени одного из сотрудников компании.

Веб-атака. Социальные сети могут использоваться хакерами для организации атак через уязвимости в браузерах, а также XSS / CSRF-атаки. Инструментами для таких атак могут быть троянские программы, поддельные антивирусы, социальные черви, вредоносные JavaScript и HTML-код, которые используются для распространения собственных списков друзей и других. Их главная цель – войти в информационную систему посетителя социальной сети, его рабочей станции или устройства и закрепиться в ней. Для защиты используются такие традиционные инструменты, как антивирусное программное обеспечение, способное работать в реальном времени и блокировать загрузку вредоносного кода.

Утечка информации и компрометация поведения сотрудников компании. Социальные сети могут использоваться для организации утечек важной информации компании, а также для подрыва ее репутации. Такая атака может вестись от внутренних сотрудников, которые недовольны руководством или специально встроенными инсайдерами. В социальных сетях люди часто ведут себя совершенно иначе, чем в корпоративной среде общения, и, возможно, шокирующая публикация и грубые реплики могут нанести определенный ущерб репутации их работодателей. DLP-системы и продукты для анализа публикаций в Интернете, предназначены для защиты от этих угроз.

Advanced Persistent Threat (APT)-атака. Исходя из вышесказанного, социальные сети могут использоваться как шлюз или источник угрозы организации, службе или другому подразделению для квалифицированных хакеров с самым современным и продвинутым вредоносным кодом, методами атак и хакерской методологией вообще. Есть много случаев, когда социальные сети стали источником информации для злоумышленников. Согласно трем хорошо известным методикам атаки: Lockheed Martin Cyber Kill Chain, Mandiant APT Atack Life Model и ISSP ThreatSCALE Model социальные сети могут использоваться на первых шагах всех этих моделей. Первый этап каждой модели - «Reconnaissance» - сбор информации о компании, ее сотрудниках, их позициях, контактах и коммуникациях, домашних адресах и т.д. С этой точки зрения социальные сети являются лучшим источником и инструментом для этого этапа. Далее, социальные сети могут быть легко использованы в качестве инструмента доставки на других этапах моделей: «Вторжение» в ThreatSCALE, «Поставка» в Cyber Kill Chain, «Проникновение» в Mandiant APT Model. Например: злоумышленник может отправить злонамеренную URL-ссылку жертве, MS Word или PDF-документ с вредоносным кодом и т.д.

Так же к основным угрозам в социальных сетях относятся следующие: содержание с признаками подстрекательства к расовой, этнической или религиозной ненависти, пропаганда тоталитарных сект; пропаганда и публичное оправдание терроризма; кибер-унижение и кибер-запугивание; популяризация и распространение наркотиков [Stohl, 2007].

Для защиты от представленных угроз службами национальной информационной безопасности, органами государственной власти, предприятиями государственного и частного сектора экономики и индивидуумами решаются следующие задачи:

- обнаружение информационных атак: определение узлов, из которых выполняется атака, оптимальное размещение сигнальных точек;
- предотвращение информационных атак: сметная стоимость нападения на объект и затраты на оборону;
- формирование и разрушение различных сетей социальной информации;
- обнаружение сообществ злоумышленников, таких как террористы, отслеживание вредоносной деятельности.

Можно выделить следующие направления противодействия информационному и психологическому воздействию виртуальных сообществ [Гриненко, 2012; Пелещишин, 2013]:

– силовые методы: закрытие серверов, формирование трафика;
– правовая и нормативная практика: уголовная ответственность организаторов и участников виртуальных сообществ;
– интернет-цензура;
– мониторинг и анализ социальных сетей.

Рассмотрим преимущества и недостатки каждого метода. Первые два метода эффективны в краткосрочной перспективе, но у них есть недостатки: отсутствие географических границ и ограничений для мгновенного распространения, сбора, обработки и использования информации – вне сферы действия законов, регулирующих правовое регулирование любого правительства; анонимность; легкодоступная изменчивость информации в электронной форме. Цензура плохо работает в демократических государствах, основанных на свободе слова.

Мониторинг и анализ социальных сетей. Методы мониторинга и анализа социальных сетей более эффективны в долгосрочной перспективе, но требуют участия специалистов в различных областях науки. Поскольку виртуальные социальные группы обладают способностью реорганизовать, основной задачей мониторинга и анализа виртуальных сообществ, представляющих угрозу для национальной безопасности информации, является не их уничтожение, а управление и контроль их деятельности с помощью разнообразных методов.

На данный момент разработано большое количество специального программного обеспечения для мониторинга и анализа интернет-среды. Основными функциями этих систем являются: мониторинг, который обеспечивает автоматизированный поиск информации в интернет-среде и позволяет определять и изменять ключевые слова для поиска информации с использованием языков поиска информации и анализ, который включает в себя автоматическую обработку информационных потоков, выявление фактов и событий, визуализацию аналитических данных в форме дайджестов, графов, графиков и других типов отчетов.

Мониторинг относится к процессу непрерывного сбора информации из социальных сетей в целях дальнейшего анализа. Так что в научных исследованиях рассматривалось проведение поиска рассматриваемого глобальной поисковой системой для социальных сетей [Григорьев, 2007; Пелещишин, 2010; 2012; 2013; Тимовчак-Максимець, 2010], и разработка коммерческих поисковых систем для специальных приложений [Горбулін, 2009; List of social networking websites, 2017; Shantanu, 2017; Смирнов, 2014], которые не учитывают особенности функционирования дискуссионных страниц.

Задачи анализа социальных сетей

В процессе анализа социальных сетей в первую очередь решаются следующие основные задачи, которые в дальнейшем декомпозируются на более подробные и точные, в зависимости от полученного результата [Fortunato, 2010, Aggarwal, 2011, Charu 2012, Батура, 2013].

Обнаружение сообществ в сети. Сообщества в сети характеризуются наличием большого числа связей между их участниками и значительно меньшим количеством связей с другими членами сети. Сообщество может соответствовать группам веб-страниц, которые имеют похожие темы [Flake, 2002], группы связанных лиц в социальных сетях [Girvan, 2002] и т. д. Простейшим случаем сообщества является такое, где каждый участник связан с каждым, а другие члены сети не общаются с членами сообщества (клика). Обнаружение сообществ (явных

и неявных) является важной задачей анализа сетей, включающей в себя классификацию членов сообщества, и, как результат, идентификацию однородных групп, групп лидеров или экспертов [Coscia, 2011, Бузун 2012, Kolomeychenko, 2014]. Обнаружение сообщества во многих случаях является кластеризацией, традиционной задачей Data Mining по отношению к различным социальным сетям. Подходы к распределению целевых групп путем выявления сообществ позволяют построить математические модели, а затем использовать модели информационного влияния и управления [Губанов, 2010]. В то же время анализ сетей исследует структуру отношений между участниками в различных областях, и обнаруживает неявные связи между ними с использованием теории графов [Ehrlich, 2005]. Более подробный обзор методов обнаружения сообщества можно найти в [Fortunato, 2010].

Анализ стабильности сообщества. Анализ явных и неявных сообществ позволяет исследовать устойчивость социальных структур. Для анализа временной стабильности групповой структуры обычно используется следующая методика. Создается первая трехмерная матрица, где строки представляют оценки взаимодействий участника со всеми другими участниками, представленные самими участниками; столбцы - это собственные оценки взаимодействия участника; временные периоды расположены на третьей оси. После этого применяются методы уменьшения размерности (например, анализ главных компонент), то есть проектирование вершин в евклидово пространство уменьшенной размерности для описания отношений между строками и столбцами матрицы. В результате возможна визуализация изменения статуса пользователя сети на фоне изменений в статусе подгрупп [Johnson, 1994]. Полученная проекция может быть сгруппирована с использованием стандартных алгоритмов кластеризации [Koren, 2003].

Обнаружение лидеров в сообществах. Поиск лидеров в сообществе является важной задачей ACC, поскольку в исследовании и моделировании информационного влияния важно иметь данные о характере взаимодействий членов сообщества, связи между ними и законами распределения информационных потоков. Согласно [Goyal, 2008], некоторый участник является лидером, если после совершения определенного действия значительное число других повторяет одно и то же действие в заданный интервал времени. Задачи обнаружения лидеров широко распространены во многих областях. Например, в [Watts, 2007] «гипотеза влиятельных членов» рассматривается в связи с маркетинговыми задачами; выбор многих лиц для предложения какого-либо продукта или инноваций [Kempe, 2003]; распространение и максимизация влияния в

конкурентных социальных сетях и привлечение последователей, вирусный маркетинг [Carnes, 2007]; распространение социального влияния [Dodds 2005, Слабченко 2013] и др.

Обнаружение экспертов в сетях. Социальная сеть может быть инструментом для поиска экспертов в конкретной области. Обнаружение экспертов связано с проблемами определения доверия и распределения влияния, а также с проблемой распространения информации в сети. С этой точки зрения распространение экспертного влияния транзитивно, т.е. влияние передается от одного узла к другому, уменьшаясь с каждым вовлеченным узлом экспертов [Bonchi 2011, Укустов, 2013]. Более подробный обзор методов экспертов по обнаружению можно найти, например, в [Charu, 2012].

Эволюция в динамических социальных сетях. С течением времени в социальных сетях появляются новые участники, некоторые участники прекращают общение, создаются новые ссылки, некоторые ссылки устаревают, так как участники больше не общаются. Это приводит к изменениям в структуре социальных сетей в целом и некоторых сообществах в частности. Таким образом, возникают важные вопросы: в соответствии с какими правилами происходят долгосрочные изменения между основными сообществами в социальных сетях; как развиваются сообщества с течением времени; как узнать изменения, которые могут произойти, какие существуют возможности их отслеживания и представления. Для исследования динамики сети используются подходы, описанные в [Bonchi 2011, Докука 2015]. Большую роль играет моделирование эволюции сетевого графа, в котором исследуются различные стратегии построения сети. Например, в [Leskovec, 2005] было обнаружено, что со временем плотность сети возрастает по степенному закону. Среди работ, представляющих алгоритмические инструменты для анализа эволюции сети, можно выделить [Tantipathananandh, 2007], где предложены алгоритмы оценки сообщества пользователей и их изменение со временем. Основное внимание уделяется определению приблизительных кластеров пользователей и эволюционных кластеров. Более подробный обзор моделей и методов эволюции социальных сетей можно найти в [Charu 2012, Aggarwal 2014, Saoussen 2014].

Прогнозирование формирования связей. Исследования, направленные на выявление и прогнозирование возможных связей между участниками или сообществами в будущем, полезны для извлечения интересующей информации из социальной сети. Связи динамичны и могут со временем сильно изменяться. Структура сети и информация об особенностях различных

участников могут быть задействованы в процессе прогнозирования связей. Задача прогнозирования связи состоит в определении того, будут ли связаны два конкретных участника друг с другом через определенный интервал времени. Эта вычислительная задача, которая основана на анализе эволюции социальной сети во времени, и называется задачей прогнозирования формирования связей. Для ее решения используется автоматическое моделирование процесса развития социальной сети с привлечением некоторых сетевых характеристик, таких как количество общих соседей, кратчайший путь, влияние вершин, время первого входа в социальную сеть. Для решения этих задач предлагается построить множество структурных и реляционных моделей. Существуют модели предсказания ссылок на основе машинного обучения, используя личную информацию пользователей сети для повышения точности прогнозирования [Liben-Nowell, 2003]. Иногда для выявления связей между пользователями используются иерархические, стохастические (марковские) и реляционные модели. В других моделях [Kumar, 2004] в качестве основы предлагается использовать свойства пользователей, например, наличие числа многих ссылок в блогосфере, которое можно объяснить сравнением демографических групп, общих интересов или географической близостью. Обзор моделей и методов прогнозирования ссылок представлен в [Hasan, 2011, Kushwah, 2016].

Кластеризация текстовой информации на основе частотного анализа. Собирая и группируя текстовые данные из социальной сети, можно определить основные темы и события, обсуждаемые пользователями социальных сетей в разных городах и странах. В настоящее время существует множество методов, позволяющих решить проблему классификации и кластеризации текстов. На этой основе реализовано много систем с использованием семантической обработки текста. Одним из основных методов частотного анализа является подсчет количества вхождений каждого слова в документ. На основе полученной информации можно сделать так называемое «облако тегов» – визуальное представление веса слов в документе. Для корректного вычисления веса слова необходимо использовать меры, которые не только посчитают количество появлений слова в документе, но также учитывать количество вхождений слова в другие документы. Примером таких мер является TF-IDF [Ramos, 2003]. В работе исследователей, занимающихся кластерным анализом текстовой информации в различных поисковых системах, часто используется индуктивная мера Word2vec [Wang, 2014, Yu, 2014]. Принцип действия меры состоит в том, чтобы найти отношения между контекстом слова в соответствии с предположением, что слова, которые находятся в сходных контекстах,

имеют тенденцию означать сходные вещи, то есть быть семантически близкими. Word2vec анализирует использование контекстов слов и приходит к заключению, что они близки по смыслу или нет. Алгоритмы, лежащие в основе word2vec, подробно описаны в работах [Mikolov, 2013, 2013 (1)].

Основные классы методов, используемых для АСС

Среди основных классов методов, используемых в АСС, можно выделить следующие: методы анализа графов, статистические методы, интеллектуальный анализ данных, методы теории оптимизации и теории алгоритмов. Также удобно выделить отдельно методы семантического анализа и анализа текстов. В этом случае необходимо обратить внимание на поддержку системой языка, на котором общаются пользователи анализируемой социальной сети.

Методы анализа графов.

Графовые модели и методы их анализа играют важнейшую роль в анализе социальных сетей, поскольку любую социальную сеть можно математически представить в виде графа G=(V,E), где V – множество вершин графа, E – множество ребер графа, N – количество вершин в графе. Графовые модели социальных сетей используются для моделирования экономических и коммуникационных связей людей, анализа процессов распространения информации, нахождения сообществ и связанных подгрупп, на которые можно разбить всю социальную сеть. В графе социальной сети вершинами являются, условно говоря, участники, а ребра означают наличие отношений между ними. Отношения могут быть как направленными, так и ненаправленными. Как правило, выделяют два типа отношений: «дружба» (люди знакомы друг с другом) и «интересы» (есть общие интересы, люди входят в одну группу по интересам).

Для анализа графовых моделей социальных сетей удобно использовать коэффициент плотности, определенный как отношение числа ребер в анализируемом графе к числу ребер в полном графе с тем же числом вершин (полный граф – это граф, в котором все вершины соединены между собой). Кроме этого, сеть могут характеризовать такие величины, как число путей заданной длины (путь – последовательность вершин, связанных между собой), минимальное число ребер, удаление которых разбивает граф на несколько частей, и другие. [Чураков, 2001]

Анализ центральности и других локальных свойств. Чтобы определить относительную важность (вес) вершин графа (т. е. насколько участник в рамках конкретной сети является влиятельным), вводят понятие центральности – меры близости к центру графа. Центральность можно определить разными способами, поэтому существуют различные меры центральности [Hanneman, 2005].

Центральность по степени (Degree centrality) определяется как количество связей, инцидентных вершине (инцидентность -связь между вершиной и ребром) Выделяют входящие и исходящие связи. Входящие связи характеризуют популярность человека, выходящие – его общительность. Полученную величину можно нормировать, разделив на общее число участников в сети. Другими словами, центральность по степени предполагает, что среди участников сети более влиятельным является тот, у кого больше друзей, либо тот, кто входит в большее количество сообществ. Тем не менее участник сети, имеющий большое количество друзей, может быть связан с остальным графом маленьким количеством ребер.

Центральность по близости (Closeness centrality) является показателем, насколько быстро распространяется информация в сети от одного участника к остальным. В качестве меры расстояния между двумя участниками используется кратчайший путь по графу. Так, непосредственные друзья участника находятся на расстоянии 1, друзья друзей – на расстоянии 2, и т. д. Далее берется сумма всех расстояний и нормируется. Полученная величина называется удаленностью вершины от других вершин. Близость определяется как величина, обратная удаленности. Центральность по близости позволяет понять, насколько близок рассматриваемый участник ко всем остальным участникам сети.

Центральность по посредничеству (Betweenness centrality). Еще одной характеристикой участника является его важность при распространении информации. Именно в этом контексте центральность по посредничеству оценивает участника. Она рассчитывается как число кратчайших путей между всеми парами участников, проходящих через рассматриваемого участника.

Центральность по собственному вектору (Eigenvector centrality) демонстрирует зависимость между центральностью участника и центральностями его друзей. Участник, который

имеет много связей с теми, у кого тоже много связей, имеет высокую центральность по собственному вектору. Таким образом, чем больше у участника друзей и чем они центральнее, тем больше его центральность. Мера центральности по собственному вектору сложна для вычисления и возможна только с помощью специализированных компьютерных программ.

Центральность можно вычислить при помощи алгоритма ссылочного ранжирования (PageRank), который используется в поисковой системе Google. В основу положен принцип «важности» веб-страницы: чем больше ссылок на страницу, тем она «важнее». Кроме того, вес самой страницы определяется весом ссылки передаваемой на нее страницы. Таким образом, PageRank – это метод вычисления веса страницы путем подсчета важности ссылок на нее [Langville, 2006].

Важными характеристиками связей сети являются *сбалансированность и транзитивность.* Сбалансированность – это отсутствие ситуаций типа «позитивное взаимодействие (дружба, партнерство) между 1-м и 2-м участниками, а также между 1-м и 3-м, но негативное взаимодействие (вражда, соперничество) между 2-м и 3-м». Утверждается, что сбалансированные сети психологически более комфортабельны для участников и более устойчивы по сравнению с несбалансированными [Johnson, 1994]. Транзитивность – это выполнение условий вида «если есть взаимодействие между 1-м и 2-м участниками, а также между 2-м и 3-м, то имеет место взаимодействие между 1-м и 3-м. Данные характеристики описывают локальные связи участников и часто используются при анализе диад и триад.

Полезной характеристикой при анализе социальных сетей является *уровень доверия*. Алгоритм вычисления уровня доверия (TrustRank) предложен в [Gyöngyi, 2004]. Изначально он был создан для отделения информативных веб-страниц от спама. Если говорить в терминах сайтов, то сначала эксперты выбирают некоторую выборку надежных сайтов, которые принимаются за эталон. Дальнейшие действия базируются на положении, что хорошие сайты редко ссылаются на плохие, а плохие очень часто ссылаются на хорошие. TrustRank – величина, которая дает оценку того, можно ли доверять конкретному сайту, считая, что он не содержит спама. Чем больше ссылок на сайте, тем меньше доверия «передается» по каждой такой ссылке. Степень доверия сайту (TrustRank) убывает с увеличением расстояния между ним и первоначальной выборкой.

Среди наиболее важных оптимизационных задач, связанных с графами, можно выделить задачу коммивояжера [Travelling salesman problem]. Это одна из самых известных комбинаторных задач оптимизации, которая заключается в нахождении наиболее прибыльного пути, проходящего через заданные вершины хотя бы один раз, а затем возвращении к исходному. В условиях

задания указываются критерии прибыльности маршрута (кратчайшие, самые дешевые, кумулятивные критерии и т. д.) и соответствующие матрицы расстояний, стоимости и тому подобное. Одним из эффективных методов для нахождения приближённых решений задачу коммивояжера, а также решения аналогичных задач поиска маршрутов на графах является муравьиный алгоритм. Суть подхода заключается в анализе и использовании модели поведения муравьёв, ищущих пути от колонии к источнику питания и представляет собой метаэвристическую оптимизацию [Ant colony optimization algorithms].

Интеллектуальный анализ данных

Многие компании стремятся проанализировать огромное количество данных социальной сети, чтобы воспользоваться этим социальным феноменом. Анализ данных в социальных сетях - одна из самых горячих тем исследования в области интеллектуального анализа данных. Применение эффективных методов интеллектуального анализа данных позволяет пользователям находить ценные, точные и полезные данные из данных социальных сетей [Cortizo, 2009, Aggarwal, 2011, Russell, 2011 Adedoyin-Olowe, 2014].

Интеллектуальный анализ данных (Data Mining) – это мультидисциплинарная область, возникшая и развивающаяся на базе таких наук как прикладная статистика, распознавание образов, искусственный интеллект, теория баз данных и др.

Основная особенность интеллектуального анализа данных - это сочетание широкого математического инструментария (от классического статистического анализа до новых кибернетических методов) и последних достижений в сфере информационных технологий. К методам и алгоритмам интеллектуального анализа относятся следующие: искусственные нейронные сети, деревья решений, символьные правила, методы ближайшего соседа и k-ближайшего соседа, метод опорных векторов, байесовские сети, линейная регрессия, корреляционно-регрессионный анализ; иерархические методы кластерного анализа, неиерархические методы кластерного анализа, в том числе алгоритмы k-средних и k-медианы; методы поиска ассоциативных правил, в том числе алгоритм Apriori; метод ограниченного перебора, эволюционное программирование и генетические алгоритмы, разнообразные методы визуализации данных и множество других [Data mining]. Большинство аналитических методов, используемые в технологиях интеллектуального анализа – это известные математические алгоритмы и методы. Новым в их применении является возможность их использования при

решении тех или иных конкретных проблем, обусловленная появившимися возможностями технических и программных средств.

Наиболее распространенными задачами Data Mining являются классификация, кластеризация, объединение, прогнозирование и визуализация. Классификация является наиболее простой и общей задачей Data Mining. В результате решения проблемы классификации обнаруживаются признаки, характеризующие группы объектов исследуемого набора данных (классов). Новый объект может быть отнесен к тому или иному классу, основанному на этих признаках. Для решения проблемы классификации используются методы ближайшего соседа и k-ближайшего соседа, байесовские сети, индукция деревьев решений и нейронных сетей [Методы классификации].

Кластеризация является логическим продолжением идеи классификации. Эта задача более сложная, функция кластеризации состоит в том, что классы объектов изначально не предопределены. Результатом кластеризации является разбиение объектов на группы. В отличие от задач классификации, кластерный анализ не требует априорных предположений о наборе данных, не накладывает ограничений на представление объектов, он позволяет анализировать различные типы данных (интервальные, частотные, двоичные). Анализ кластеров позволяет уменьшить размер данных, сделать их видимыми. Методы кластерного анализа можно разделить на две группы: иерархические и неиерархические. Каждая группа включает в себя множество подходов и алгоритмов. [Методы кластерного анализа]

Суть иерархической кластеризации состоит в последовательном объединении меньших кластеров в крупные или разделении больших кластеров на более мелкие. Преимуществом иерархических методов кластеризации является их наглядность. Иерархические алгоритмы, связаны со строительством дендрограмм, которые являются результатом иерархического кластерного анализа. Дендрограмма описывает близость отдельных точек и кластеров друг к другу, это графическая последовательность слияния (разделения) кластеров. При большом числе наблюдений методы иерархического кластерного анализа не подходят. В таких случаях используются неиерархические методы, основанные на разделении, которые являются итерационными методами фрагментации исходного набора. Во время разделения новые кластеры формируются до тех пор, пока не будет выполнено правило остановки.

Одним из самых популярных методов анализа данных является анализ основных компонентов, который исходит из прикладного статистического анализа [Principal component analysis]. Это один из основных способов уменьшить размерность пространства наблюдения, потеряв наименьшее

количество информации. Он используется во многих областях, в том числе в эконометрике, биоинформатике, обработке изображений, сжатии данных, в социальных науках.

Применение фрактального анализа для АСС

Если наблюдать во времени динамику показателей, полученных с помощью АСС, то от статических данных мы перейдем к временным рядам, к которым применимы все инструменты анализа и прогнозирования временных рядов. В настоящее время стало общепризнанным, что многие временные ряды имеют долгосрочную зависимость и фрактальные свойства. Одними из первых реальных стохастических процессов, у которых были обнаружены самоподобные свойства, были информационные потоки данных в телекоммуникационных сетях. Существует большое количество публикаций, посвященных анализу самоподобных и мультифрактальных свойств трафика и их влияния на функционирование и качество обслуживания телекоммукационной сети (см., например, [Шелухин, 2011]). Многочисленные исследования показали, что многие биоэлектрические сигналы обладают фрактальной структурой [Stanley, 1999]. Другим примером фрактальных стохастических структур являются современные финансовые рынки. Анализируя динамику возникновения участков с различной фрактальной структурой, можно диагностировать и прогнозировать нестабильные состояния (кризисы) рынка [Peters, 1996, Соловьев, 2015]. В последние годы появились исследования динамики сообществ в социальных группах, которые показывают, что соответствующие временные ряды обладают свойствами самоподобия [Cosoi, 2009, Yang, 2009, Rybski, 2012, Qingyun, 2016].

Самоподобие случайных процессов заключается в сохранении вероятностных характеристик при изменении масштаба времени. Стохастический процесс $X(t)$ является самоподобным с параметром H, если процесс $a^{-H}X(at)$ описывается теми же законами конечномерных распределений, что и $X(t)$. Параметр H, $0<H<1$, называемый показателем Херста, представляет собой степень самоподобия процесса. Наряду с этим свойством, показатель $H>0.5$ характеризует меру долгосрочной зависимости стохастического процесса. Это означает, что если временной ряд в течении какого-то времени возрастал (убывал), то с вероятностью, близкой к показателю Херста, ряд сохранит эту тенденцию в течение аналогичного промежутка времени [Feder, 1988]. Начальные моменты самоподобного случайного процесса имеют скейлинговое соотношение $\mathrm{E}\left[|X(t)|^q\right] \propto t^{qH}$, где величина t – интервал времени.

Мультифрактальные объекты являются статистически неоднородными самоподобными объектами. Для мультифрактальных временных рядов статистическая неоднородность объекта выражается в неоднородности распределения данных ряда, т.е. наличии тяжелых хвостов плотности распределения вероятностей временного ряда. В сравнении с самоподобными, мультифрактальные процессы проявляют более сложное скейлинговое поведение: для моментов мультифрактальных процессов выполняется отношение $\mathrm{E}\left[\left|X(t)\right|^q\right] \propto t^{qh(q)}$, где $h(q)$ – обобщенный показатель Херста, являющийся в общем случае нелинейной функций. Значение $h(q)$ при $q=2$ совпадает со значением степени самоподобия H. Для монофрактальных процессов обобщенный показатель Херста не зависит от параметра q: $h(q)=H$. [Reidi, 2002].

В работе был проведен сравнительный фрактальный анализ для двух групп в социальной сети Facebook, найденным по ключевым словам, связанных с киберугрозами. Для проведения исследований в сети были выбраны две группы ThreatPost и Threat Signal, каждая численностью 13980 и 84375 пользователей соответственно. Для каждой их этих социальных групп были сняты данные за последние пять лет о количестве лайков, комментариев и уровне вовлеченности (показатель уровня вовлеченности аудитории в активности группы измеряется в процентном соотношении действий пользователей к охвату аудитории).

На рис.1 слева показаны временные ряды ежедневного количества лайков для обеих групп. Фрактальный анализ показал, что временные ряды показателей этих групп обладают сильными мультифрактальными свойствами. На рис. 1 справа представлены значения обобщенного показателя Херста для этих рядов. Оба ряда обладают фрактальными свойствами, персистентностью $(H_1, H_2 > 0.5)$, однако ряд лайков для группы Threat Signal обладает значительно большей статистической неоднородностью, что проявляется в значительно большем диапазоне величины $\Delta h(q)$.

Исследования показали, что ряды лайков достаточно сильно коррелируют с рядами уровня вовлеченности, поэтому их соответствующие мультифрактальные характеристики очень близки. Ряды количества комментариев (рис.2, слева), в отличие от рядов лайков, также имеют фрактальные свойства, но обладают достаточно близкой между собой мультифрактальной структурой (рис.2, справа).

Можно предположить, что механизм генерации комментариев, который формирует фрактальную структуру временного ряда, достаточно сильно отличается от механизма проставления лайков. Таким образом, проведенные в работе исследования подтвердили, что многие временные ряды показателей активности пользователей социальных сетей обладают фрактальными свойствами

и применение фрактального анализа позволяет выявить различия и обнаружить характерные черты динамики разных социальных групп.

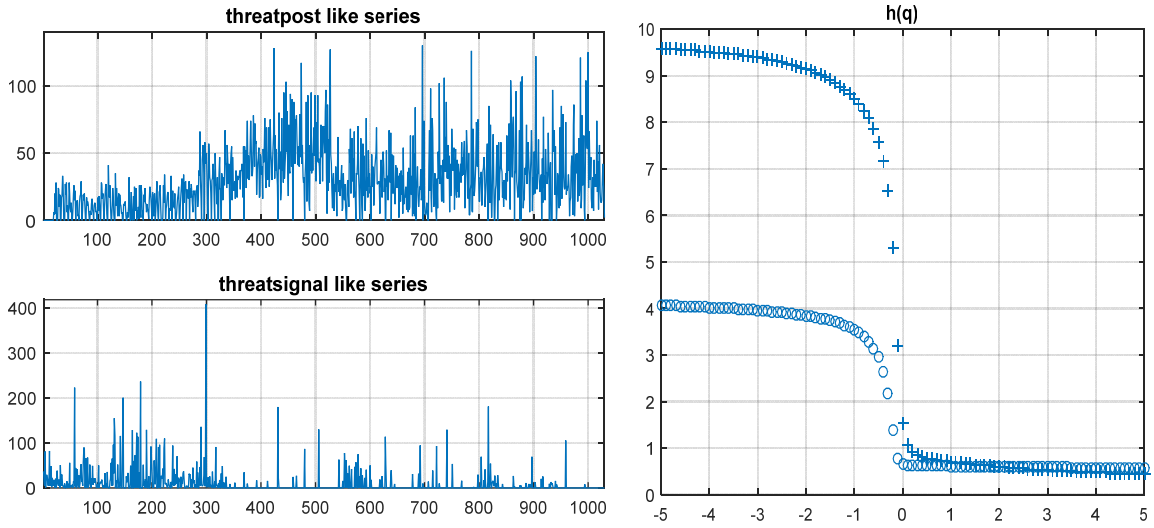

Рисунок 1. Слева: временные ряды лайков для групп; справа: соответствующие значения $h(q)$ (+ – группа Threat Post, o – Threat Signal)

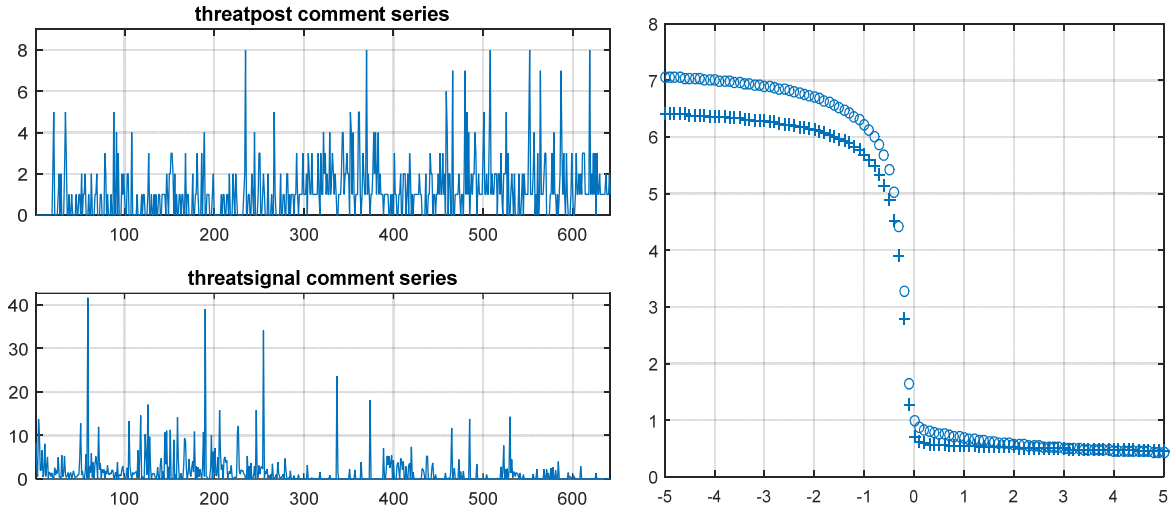

Рисунок 2. Слева: временные ряды комментариев для групп; справа: соответствующие значения $h(q)$ (+ – группа Threat Post, o – Threat Signal)

Выводы

В работе представлен обзор основных методов анализа социальных сетей, используемых для обеспечения информационной безопасности. Рассмотрены основные типы угроз и основные профилактические меры для безопасности социальных сетей. Представлены типичные задачи анализа социальных сетей, направленные на выявление киберугроз, такие как обнаружение сообществ в сети, обнаружение лидеров в сообществах, обнаружение экспертов в сетях, кластеризация текстовой информации на основе частотного анализа и другие. Кратко описаны основные методы и алгоритмы теории графов и интеллектуального анализа данных, которые используются в анализе социальных сетей. Поведенные исследования подтвердили, что многие ряды показателей активности пользователей сетей обладают фрактальными свойствами. Фрактальный анализ выявил различия в мультифрактальной структуре рядов различных показателей и социальных групп.

Bibliography

[Adedoyin-Olowe, 2014] Mariam Adedoyin-Olowe, Mohamed Medhat Gaber, Frederic Stahl. A Survey of Data Mining Techniques for Social Network Analysis. Journal of Data Mining & Digital Humanities, 2014. pp.1-25.

[Aggarwal, 2011] Aggarwal C. Introduction to social network data analytics. Springer US, 2011. doi: 10.1007/978-1-4419-8462-3_2

[Aggarwal, 2014] Aggarwal C., Karthik S. Evolutionary Network Analysis: A Survey ACM Computing Surveys, Vol. 47, No. 1, Article 10, 2014.

[Ant colony optimization algorithms]: https://en.wikipedia.org/wiki/Ant_colony_optimization_algorithms

[Bonchi, 2011] Bonchi F., Castillo C., Gionis A., Jaimes A. Social Network Analysis and Mining for Business Applications, ACM TIST, vol. 2, iss. 3, 2011. pp.22–58.

[Carley, 2002] Carley K., Lee J., Krackhardt D. Destabilizing networks. Connections, vol. 24, №.3, 2002. pp.79 – 92.

[Carnes, 2007] Carnes T., Nagarajan R., Wild S.M., A. Van Zuylen. Maximizing influence in a competitive social network: a follower's perspective . Proceedings of the ninth international conference on Electronic commerce, Minneapolis, USA, 2007. 351–360.

[Charu, 2012] Charu C. Social network data analytics. Springer Science & Business Media, 2012. 486 p.

[Cortizo, 2009] Cortizo, J., Carrero, F., Gomez, J., Monsalve, B., Puertas, E. Introduction to Mining SM. Proceedings of the 1st International Workshop on Mining SM, 2009. pp.1 – 3.

[Coscia, 2011] Coscia M., Giannotti F., Pedreschi D. A classification for community discovery methods in complex networks. Statistical Analysis and Data Mining, 2011. pp.512–546

[Cosoi, 2009] Alexandru Catalin Cosoi, Carmen Maria Cosoi. A fractal approach to social network spam detection: https://www.virusbulletin.com/conference/vb2009/abstracts/fractal-approach-social-network-spam-detection/

[Data mining]: https://en.wikipedia.org/wiki/Data_mining .

[Dodds, 2005] Dodds P.S., Watts D.J. A generalized model of social and biological contagion. Journal of theoretical biology, iss. 4, 2005. pp.587–604.

[Easley, 2010] Easley D., Kleinberg J. Networks, Crowds, and Markets: Reasoning about a Highly Connected World, Cambridge University Press, 2010. 819 p.

[Ehrlich, 2005] Ehrlich K., Carboni I. Inside Social Network Analysis IBM Watson Research Center. New York, USA, Technical Report, 2005. pp.5–10.

[Feder, 1988] J. Feder. Fractals. Plenum, New York, 1988.

[Flake, 2002] Flake G.W., Lawrence S., Giles C.L., Coetzee F.M. Self-organization and identification of Web communities. Computer. , Iss. 3, 2002. pp.66–70.

[Fortunato, 2010] Fortunato S. Community detection in graphs. Physics Reports, Vol. 486, Iss. 3–5, 2010. pp.75–174.

[Girvan, 2002] Girvan M., Newman M.E. Community structure in social and biological networks, Proceedings of the National Academy of Sciences of the United States of America, Iss. 12, 2002. pp.7821–7826

[Goyal, 2008] Goyal A., Bonchi F., Laks V.S. Lakshmanan. Discovering leaders from community actions. Proceedings of the 17th ACM Conference on Information and Knowledge Management, Napa Valley, California, USA, 2008. pp.499–508.

[Gyöngyi, 2004] Gyöngyi Z., Garcia-Molina H., Pedersen J. Combating Web Spam with TrustRank. Proceedings of the International Conference on Very Large Data Bases, Vol. 30, 2004. pp.576-587.

[Hanneman, 2005] Hanneman R., Riddle M. Introduction to social network methods. Riverside, CA: University of California. 2005. http://faculty.ucr.edu/~hanneman/nettext .

[Hasan, 2011] Mohammad Al Hasan, Mohammed J. Zaki. A Survey of Link Prediction in Social Networks. Social Network Data Analytics, 2011. pp.243-275.

[Johnson, 1994] Johnson J., Ironsmith M. Assessing Children's Sociometric Status: Issues and the Application of Social Network Analysis. Journal of Group Psychotherapy, Psychodrama & Sociometry, vol. 47(1), 1994. pp. 36–49.

[Kempe, 2003] Kempe D., Kleinberg J., Tardos É. Maximizing the spread of influence through a social network. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, USA, 2003. pp.137–146.

[Kolomeychenko, 2014] M. I. Kolomeychenko, A. A. Chepovskiy, A. M. Chepovskiy. An algorithm for detecting communities in social networks, Fundamentalnaya i prikladnaya matematika, Vol. 19, No. 1, 2014. pp. 21—32.

[Koren, 2003] Koren Y. On Spectral Graph Drawing . Proceedings of the 9th International Computing and Combinatorics Conference, Springer, 2003. pp. 496–508

[Kumar, 2004] Kumar R., Novak J., Raghavan P., Tomkins A. Structure and Evolution of Blogspace Commun. ACM, vol. 47, No. 12, 2004. pp.35–39.

[Kushwah, 2016] Ajay Kumar Singh Kushwah, Amit Kumar Manjhvar. A Review on Link Prediction in Social Network International Journal of Grid and Distributed Computing, vol. 9, No. 2, 2016. pp.43-50.

[Langville, 2006] Langville Amy N., Meyer Carl D. Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press, 2006. 224 p.

[Leskovec, 2005] Leskovec J., Kleinberg J., Faloutsos C. Graphs over time: densification laws, shrinking diameters and possible explanations. Proc. 11th ACM SIGKDD Intern. Conf. on Knowledge Discovery in Data Mining, NY, 2005. pp.177–187.

[Liben-Nowell, 2003] Liben-Nowell D., Kleinberg J. The Link Prediction Problem for Social Networks. Proceedings of the 12th International Conference on Information and Knowledge Management, N. Y.: ACM Press, 2003. pp. 556–559.

[List of social networking websites, 2017] List of social networking websites, 2017. URL: https://en.wikipedia.org/wiki/List_of_social_networking_websites

[Mikolov, 2013 (1)] Mikolov T., Sutskever I., Chen K., Corrado G., Dean J. Distributed Representations of Words and Phrases and their Compositionality, 2013. https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf

[Mikolov, 2013] Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space, 2013. https://arxiv.org/pdf/1301.3781.pdf

[National Cyber Security Strategy 2016-2021 for United Kingdom]: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/564268/national_cyber_security_strategy.pdf.

[National cyber security strategy for Ukraine]: http://www.president.gov.ua/documents/962016-19836

[Peters, 1996] Edgar E. Peters. Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility. Edgar E. Peters. Wiley, 2 edition, 1996.

[Principal component analysis]: https://en.wikipedia.org/wiki/Principal_component_analysis.

[Qingyun, 2016] Qingyun Liu, Xiaohan Zhao, Walter Willinger, Xiao Wang, Ben Y. Zhao, Haitao Zheng Self-Similarity in Social Network Dynamics ACM Trans. Model. Perform. Eval. Comput. Syst., Vol. 2, No. 1, Article 5, 2016.

[Ramos, 2003] Ramos J. Using tf-idf to determine word relevance in document queries. Proceedings of the first in-structional conference on machine learning, 2003.

[Russell, 2011] Matthew A. Russell. Mining the Social Web: Analyzing Data from Facebook,Twitter, LinkedIn, and Other Social Media Sites. O'Reilly, 2011. 332 p.

[Rybski, 2012] Diego Rybski, Sergey V. Buldyrev, Shlomo Havlin, Fredrik Liljeros, Herna A. Makse Communication activity in a social network: relation between long-termcorrelations and inter-event clustering Scientific reports, 2012, 2 : 560 | DOI: 10.1038/srep00560

[Saoussen, 2014] Saoussen Aouay, Salma Jamoussi, Faiez Gargouri, Ajith Abraham. Modeling Dynamics of Social Networks: A Survey. Sixth International Conference on Computational Aspects of Social Networks (CASoN), 2014. pp.49-53.

[Shantanu, 2017] Shantanu Ghosh. Top seven social media threats, 2017. URL: http://www.computerweekly.com/tip/Top-seven-social-media-threats

[Stanley, 1999] H. E. Stanley, L. A.N. Amaral, A.L. Goldberger, S. Havlin, P.Ch. Ivanov, C.-K. Peng. Statistical Physics and Physiology: Monofractal and multifractal approaches. Physica A 270, 1999. pp. 309-324.

[Stohl, 2007] Stohl C., Stohl M. Networks of Terror: Theoretical Assumptions and Pragmatic Consequences. Communication Theory, Vol. 17, 2007. pp.93 – 124.

[Tantipathananandh,2007] Tantipathananandh C., Berger-Wolf T., Kempe D. A framework for community identification in dynamic social networks .Proc. 13th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining, NY, 2007. pp.717–726

[The NATO Cooperative Cyber Defence Centre]: https://ccdcoe.org/cyber-security-strategy-documents.html.

[Travelling salesman problem]: https://en.wikipedia.org/wiki/Travelling_salesman_problem.

[Wang, 2014] Wang H. Introduction to Word2vec and its application to find predominant word senses, 2014. http://compling.hss.ntu.edu.sg/courses/hg7017/pdf/word2vec%20and%20its%20application% 20to% 20wsd.pdf.

[Watts, 2007] Watts D.J., Dodds P.S. Influentials, Networks, and Public Opinion Formation. Journal of consumer research, iss. 4,2007. pp.441–458.

[Yang, 2009] Christopher C. Yang, Marc Sageman Analysis of terrorist social networks with fractal views Journal of Information Science, V. 35, Issue 3, 2009. pp. 299–320.

[Yu, 2014]Yu M., Dredze M. Improving lexical embeddings with semantic knowledge. Association for Computational Linguistics (ACL), 2014. pp.545-550.

[Батура, 2013] Батура Т.В. Модели и методы анализа компьютерных социальных сетей. Программные продукты и системы. - №3 (103), 2013. С.130-137.

[Бузун, 2012] Бузун Н., Коршунов А. Выявление пересекающихся сообществ в социальных сетях М.: Институт системного программирования РАН, 2012. 18 с. http://www.twirpx.com/file/987306

[Горбулін, 2009] Горбулін В. П., Додонов О. Г., Ланде Д. В. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія. Інтертехнологія, 2009. – 164 с.

[Григорьев, 2007] Григорьев А. Н., Ландэ Д. В., Бороденков С. А., Мазуркевич Р. В., Пацьора В. Н. InfoStream. Мониторинг новостей из Интернет: технология, система, сервис. К.: ООО “Старт-98”, 2007. 40 с.

[Гриненко, 2012] Гриненко І, Прокофьєва – Янчиленко Д. Вплив віртуальних спільнот на інформаційну безпеку: сучасний стан та тенденції розвитку. Правове, нормативне та метрологічне забезпечення систем захисту інформації в Україні, № 1 (23), 2012. С. 18 – 23.

[Губанов, 2010] Губанов Д., Новиков Д., Чартишвили А. Социальные сети: модели информационного влияния, управления и противоборства. Москва, 2010. 225 с.

[Дзюндзюк, 2011] Дзюндзюк В. Б. Віртуальні співтовариства: потенційна загроза для національної безпеки / В. Б. Дзюндзюк // Державне будівництво [Електронне видання]. № 1, 2011. Режим доступу до журн. : http:// www.kbuapa.kharkov.ua. – Назва з екрана

[Додонов, 2009] Додонов О. Г., Ланде Д. В., Путятін В. Г. Інформаційні потоки в глобальних комп'ютерних мережах. Київ: Наукова думка, 2009. 295 с.

[Додонов, 2013] Додонов А. Г., Ландэ Д. В., Прищепа В. В., Путятин В. Г. Конкурентная разведка в компьютерных сетях. Киев: ИПРИ НАН Украины, 2013. 248 с.

[Додонов, 2014] Додонов А. Г., Ландэ Д. В., Путятин В. Г. Компьютерные сети и аналитические исследования. Киев: ИПРИ НАН Украины, 2014. 486 с.

[Докука, 2015] Докука С. В., Валеева Д. Р. Статистические модели для анализа динамики социальных сетей в исследованиях образования. Вопросы образования, № 1, 2015. С.201-213

[Матвиенко, 2011] Матвиенко Ю. А. Деструктивные сетевые социальные структуры как средство информационной войны и угроза безопасности России. Информационно-аналитический портал «Геополитика», 2011. [Электронный ресурс]. Режим доступу до журн.: http://old.geopolitica.ru/Articles/1218. – Загл. с экрана.

[Методы классификации]: http://www.intuit.ru/studies/courses/6/6/lecture/176.

[Методы кластерного анализа]: http://www.intuit.ru/studies/courses/6/6/lecture/182.

[Пелещишин, 2010] Пелещишин О. П. Інформаційні технології обліку та пошуку онлайн-спільнот у задачі соціального маркетингу. Вісник Національного університету “Львівська політехніка”. Серія економічна, № 44, 2010. С. 50 – 59.

[Пелещишин, 2012]Пелещишин А. М., Сєров Ю. О., Березко О. Л., Пелещишин О. П., Тимовчак-Максимець О. Ю., Марковець О. В. Процеси управління інтерактивними соціальними комунікаціями в умовах розвитку інформаційного суспільства: монографія. Львів: Видавництво Львівської політехніки, 2012. 368 с.

[Пелещишин, 2013] Пелещишин О. П. Аналіз та протидія загрозам маркетинговій позиції підприємства в онлайн-спільнотах. Захист інформації, № 3 (15), 2013. С. 217 – 224.

[Слабченко, 2013] Слабченко О. О., Сидоренко В. Н., Пономарчук Р. А.. Методы и алгоритмы выявления сообществ потенциальных абитуриентов и их лидеров в социальных сетях. Вестник Кременчугского национального университета, №1 (78), 2013. С.53-61.

[Смирнов, 2014] Смирнов А. И., Григорьев В. Р., Кохтюлин И. Н., Куроедов Б. В., Сандаров О. В. Глобальная безопасность в цифровую эпоху: стратагемы для России. М.: ВНИИ геосистем, 2014. 394 с.

[Соловйов, 2015] Соловйов В.М. Мережні міри складності соціально-економічних систем. Вісник Черкаського університету, сер. «Прикладна математика. Інформатика», № 38 (371), 2015. с.67-79.

[Тимовчак-Максимець, 2010] Тимовчак-Максимець О. Ю. Методи використання розширених можливостей глобальних пошукових систем в задачі пошуку споживацького досвіду в онлайн середовищах. Вісник Національного університету “Львівська політехніка”: Інформаційні системи та мережі, № 689, 2010. с. 323 – 331.

[Укустов, 2012] Укустов С. С., Кравец А. Г. Подход к решению задачи идентификации влиятельных разработчиков в социальной сети гитхаб. Известия Волгоградского Государственного Технического Университета, Выпуск № 15 (102), 2012. с.61-66.

[Чураков, 2001] Чураков А. Н. Анализ социальных сетей. Социальные исследования, № 1, 2001. С. 109–121.

[Шелухин, 2011] Шелухин О. И. Мультифракталы. Инфокоммуникационные приложения. М.: Горячая Линия, Телеком, 2011. 578 с.

Информация об авторах

Людмила Кириченко *– д.т.н., профессор Харьковского национального университета радиоэлектроники; пр. Науки 14, 61166, Харьков, Украина; e-mail: lyudmyla.kirichenko@nure.ua.*

Основные области научных исследований: самоподобные и мультифрактальные временные ряды, фрактальный анализ, вейвлет-анализ, детерминированные хаотические системы.

Алексей Барановский *- к.т.н., старший преподаватель Национального технического университета Украины "Киевский политехнический институт"; пр. Победы 37, 03056, Киев; email: o.baranovskyi@kpi.ua .*

Основные области научных исследований: фрактальный анализ, анализ рекуррентных диаграмм, теория безопасности, информационная безопасность, информационные операции и воздействия.

Тамара Радивилова *– к.т.н., доцент Харьковского национального университета радиоэлектроники; пр. Науки 14, 61166, Харьков, Украина; e-mail:: tamara.radivilova@gmail.com.*

Основные области научных исследований: самоподобные и мультифрактальные временные ряды, телекоммуникационные системы, управление трафиком, информационная безопасность

Detecting cyber threats through social network analysis

Lyudmyla Kirichenko, Tamara Radivilova, Oleksii Baranovskyi

Abstract*: The paper reviews the basic methods of social networks analysis which are used to detect cyber threats. The main types of threats in social networks are presented and some methods for their prevention are described. Typical tasks of social networks analysis aimed at identifying cyber threats, such as community detection in networks, detection of leaders and experts in communities, stability analysis of community, clustering text information, etc. are considered. The basic classes of methods of graph theory and data mining, which are widely used in the analysis of social networks are described. The application of fractal analysis methods to study the series of network users activity indicators is shown.*

Keywords*: analysis of social networks, fractal analysis, cyber threats, data mining, detection leader methods, detection experts methods.*