УДК 27.45.17 DOI: 10.18698/2541-8009-2019-03-452

ОЦЕНКА ДОСТОВЕРНОСТИ ИНФОРМАЦИИ, СОБИРАЕМОЙ ИЗ СОЦИАЛЬНЫХ СЕТЕЙ

Н.Л. Харламов wencyre@gmail.com
SPIN-код: 3450-4547

Л.А. Латышева lubalee.work@gmail.com
SPIN-код: 7510-3315

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

Представлен краткий обзор существующих методов обнаружения ботов, основанных на статистическом и семантическом анализе текстов, поведенческом анализе и теоретико-графовом подходе. Приведен пример применения алгоритма выделения сообществ для решения сопутствующей задачи — поиска лидеров общественного мнения. Предложен новый подход к обнаружению ботов, основанный на анализе сообществ графа ближайшего окружения пользователя. Предложенный метод был опробован на двух выборках виртуальных пользователей из социальной сети ВКонтакте: управляемых ботов и ботов, собранных вручную одним из пользователей. Для проверки была использована выборка из 700 легитимных пользователей.

Ключевые слова

Социальная сеть, социальный граф, выделение сообществ, поиск ботов, лидеры общественного мнения, виртуальные пользователи, алгоритм выявления ботов

Поступила в редакцию 21.01.2019

Введение. Сегодня социальные сети стали альтернативными СМИ. Многие пользователи доверяют информации, полученной из социальных сетей, больше, чем традиционным источникам новостей — телевидению и газетам [1, 2]. Косвенно данный факт подтверждается принятым Федеральным законом № 97-ФЗ от 5 мая 2014 г., приравнивающим крупных блогеров к СМИ.

В настоящее время возникает необходимость анализа социальных сетей, в том числе как сетей оказания влияния на пользователей — выявления ботов (специальных программ, выполняющих автоматически или по заданному расписанию какие-либо действия через интерфейсы, предназначенные для людей). С развитием информационно-телекоммуникационных технологий в мире существенно возросла важность ресурсов нового типа — социальных сетей в сети Интернет — как средств, влияющих на действия пользователей сети и качество информации, циркулирующей в социальных сетях.

Посредством эксплуатации ботов в социальных сетях сильно искажается информация о действительных предпочтениях и интересах пользователей социальных групп. Одна из основных целей использования ботов — распространение информации, как положительной, так и отрицательной относительного

продвигаемой идеи или явления. На основе данных, собранных из социальных сетей, проводится оценка популярности и востребованности продукта или услуги, а также реакции потребителей на проведенные акции и специальные предложения. Это мешает проведению SMM-анализа (Socal Media Marketing — процесс привлечения трафика или внимания к бренду или продукту через социальные платформы), искажая информацию об интересе пользователей к проекту за счет накрутки количества участников в сообществах. Поэтому следует определять, какие пользователи социальной сети реальны, а какие могут оказаться ботами.

Многие политики имеют свои страницы в социальных сетях, а некоторые их записи вызывают большой резонанс. Социальные сети могут повлиять на исход выборов благодаря влиянию на общественное мнение [3].

Последствиями массового распространения ботов в социальных сетях могут быть массовое хищение персональных данных, снижение уровня доверия к информации из социальных сетей, информационные вбросы, создание ложных новостей и необъективность результатов проведения SMM-анализа, социальных исследований.

Определение ботов и лидеров общественного мнения. Выделяют две задачи, которые можно решить с помощью анализа структуры социальной сети. Это задачи выявления лидеров общественного мнения — персон, которые имеют наибольшее влияние на других пользователей социальной сети — и виртуальных пользователей (ботов), которые могут влиять на численные показатели систем мониторинга путем создания большого количества сообщений с определенной тематикой. Обе задачи являются задачами бинарной классификации.

Выделению лидеров общественного мнения посвящено множество исследований. Например, в социальной сети Twitter исследователи выделяют более 75 различных метрик влияния пользователей [2], среди которых заслуживают отдельного упоминания PageRank (числовая мера важности страницы пользователя, рассчитываемая в зависимости от количества и качества ссылок на эту страницу — как внешних, так и внутренних) [4], промежуточность (показывает, сколько кратчайших путей между всеми узлами сети проходит через определенный узел, если у узла высокая промежуточность, то он — единственная связь между различными частями сети) и близостная центральность (демонстрирует, насколько легко достичь определенного узла в сети, и позволяет судить, как пользователь с высокой близостной центральностью взаимодействует с сообществом и может оказать наибольшее влияние на мнение). Стоит обратить внимание на то, какой граф нужно анализировать. Очевидно, что проводить анализ графа всей сети затратно и нецелесообразно, как минимум потому, что эксперт в одной области может совершенно не разбираться в другой. Обычно рассматривают подграф, соответствующий некоторой теме или событию, вершинами которого являются участники обсуждения. Известно, что мнения агентов в рамках одной группы (сообщества) сходятся и практически не поддаются внешнему влиянию [5]. Поэтому в некоторых задачах помимо лидеров темы имеет

смысл предварительно определить в графе темы группы пользователей, имеющих одно мнение, с помощью алгоритма выделения сообществ и определить лидеров мнений в каждом сообществе независимо. Например, если тема биполярна, таким образом можно обнаружить лидеров мнения противника, на которых следует попробовать оказать влияние.

Разработка универсального метода обнаружения ботов, по крайней мере в ближайшее время, не представляется возможной, поскольку виртуальные пользователи, управляемые человеком, могут быть практически неотличимы от реальных пользователей. В свою очередь, поведение новых или малоактивных пользователей порой очень схоже с поведением ботов. Кроме того, иногда реальные пользователи (обычно несовершеннолетние или малообеспеченные) готовы за небольшую плату или безвозмездно выполнять задачи ботов, используя свой реальный аккаунт в социальных сетях: накрутку постов, накрутку количества подписчиков, троллинг других пользователей и т. п. Тем не менее необходимо постоянное совершенствование методов обнаружения ботов, что позволит существенно сократить количество виртуальных пользователей и повысить стоимость их создания, оставив возможность поддержания таких пользователей лишь единичным субъектам [6]. Отметим некоторые из основных таких подходов.

Во-первых, это поведенческий анализ. Его используют преимущественно сами социальные сети как один из внутренних механизмов. В основном так происходит обнаружение пользователей, нарушающих тем или иным образом пользовательское соглашение социальной сети. Резкое изменение активности обычно ведет к предложению пройти специальный тест CAPTCHA; при получении неверного ответа происходит блокировка или заморозка аккаунта. Точные механизмы обнаружения подозрительной активности обычно составляют тайну, хотя при планомерном исследовании методом проб и ошибок злоумышленник может найти методы обхода данных ограничений. Извне проводить подобный анализ довольно затруднительно, однако так можно вычислить очень простых автоматических ботов. Например, некоторые из них проявляют свою активность по расписанию и их легко обнаружить по временному профилю активности [7].

Другой подход заключается в статистическом и семантическом анализе текстов [8]. В случае если группа ботов управляется одним человеком и идет активное обсуждение некоторой темы, тексты управляемых пользователей могут быть схожи между собой или даже идентичны друг другу. Кроме того, у каждого человека есть характерные особенности речи и лексикона, что позволяет на основе анализа корпуса текстов пользователя ставить ему в соответствие уникальный идентификатор. При мониторинге, охватывающем достаточно большую часть социальной сети, легко найти пользователей с одинаковым идентификатором. С большой вероятностью можно утверждать, что часть или все такие пользователи будут ненастоящими (ботами или «фейками»). В некоторых исследованиях с помощью анализа тональности сообщений составляется профиль эмоций пользователя [8]. Боты, предназначенные для

«вбросов», обычно не обладают целостным профилем эмоций. Кроме этого проводится анализ количества ссылок в сообщениях и количества ответов другим пользователям [9].

Третий подход представляет собой анализ связей пользователя. Простейшие методы основаны на анализе количества связей пользователя и отношения количества входящих и исходящих связей [10]. Методы, используемые для определения ботнетов, подходят и для анализа социального графа. Зараженные сегменты сети можно определить с помощью алгоритма выделения сообществ [11, 12] примененного к графу взаимодействия узлов; боты определяются по аномальному трафику.

Объединяет данные подходы комбинированный метод, заключающийся в использовании нескольких одновременно. Зачастую он строится на машинном обучении [8, 9] и позволяет на основе многих факторов выносить решение об обнаружении бота с некоторой долей уверенности. Основная задача заключается в отборе значимых факторов, подборе классификаторов и обучении модели. Даже если не использовать машинное обучение, применение нескольких подходов повышает вероятность обнаружения поддельных аккаунтов.

Отметим, что описанные подходы можно использовать для выявления недобросовестных СМИ, которые занимаются цитированием друг друга и публикуют недостоверную информацию.

Метод обнаружения ботов с помощью выделения сообществ. Известно, что социальные связи человека имеют свои особенности. Например, число связей, которые человек может эффективно поддерживать, ограничено парой сотен. Данный предел был выведен антропологом Р. Данбаром [12] и назван в его честь. Человек естественным образом участвует в нескольких сообществах, соответствующих его сферам деятельности и интересам: у него есть одноклассники, коллеги по работе, родственники и т. д.

Таким образом, степень каждой вершины в графе социальной сети ограничена сверху. Поэтому целесообразно вместо анализа графа всей социальной сети провести анализ графов ближайшего окружения пользователей. Скорее всего, графы реальных людей будут сильно отличаться от графов виртуальных пользователей, ведь у последних граф будет сформирован искусственным образом.

Для проверки гипотезы были использованы три группы пользователей социальной сети ВКонтакте:

– 668 пользователей полученного из открытых источников сети Интернет списка управляемых ботов (управляемые боты);

– боты, отобранные вручную одним из пользователей сети с помощью аккаунта-приманки (1366 профилей) (боты, отобранные вручную одним из пользователей сети);

– выборка из 700 случайно выбранных пользователей, которые считались не ботами (обычные пользователи).

Для каждого из пользователей был составлен граф ближайшего окружения, состоящий из всех друзей пользователя и связей между ними. Был использован

граф сети ВКонтакте. После этого к каждому из графов был применен алгоритм выделения сообществ, основанный на максимизации модулярности.

На рис. 1 представлен график количества сообществ для каждой из трех групп. В основном обычные пользователи имеют около пяти сообществ, в то время как для многих ботов их либо меньше четырех, либо больше десяти. Некоторые боты вообще не имеют сообществ либо имеют только одно. При этом средний размер сообществ (рис. 2) у ботов обычно либо больше, чем у нормальных пользователей, либо представляет величину, меньшую 5. У большинства ботов из числа управляемых средний размер сообщества превышает 100 человек, при этом количество их «друзей» может исчисляться несколькими сотнями пользователей, как видно на рис. 3. У большинства нормальных пользователей количество друзей лежит в интервале от 50 до 200, что согласуется с числом Данбара.

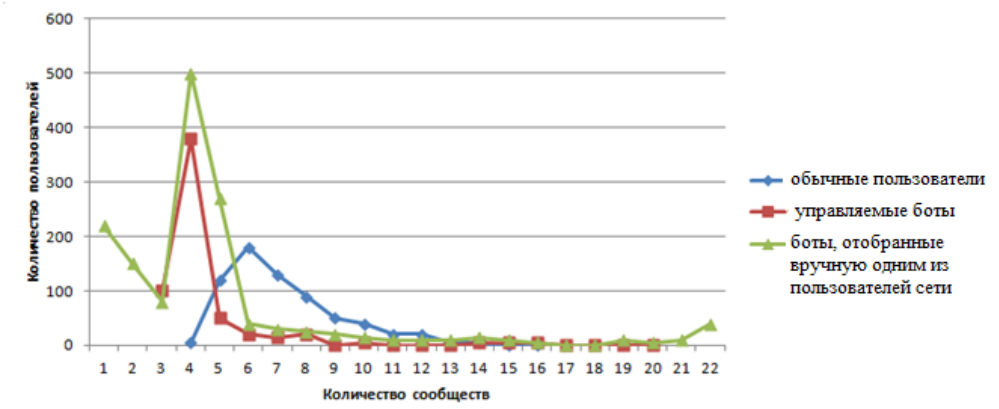

Рис. 1. Количество сообществ у пользователей

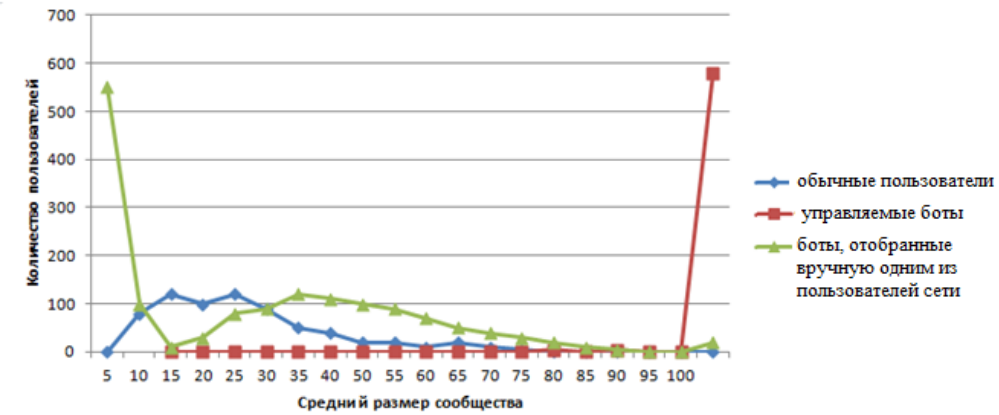

Рис. 2. Средний размер сообществ у пользователей

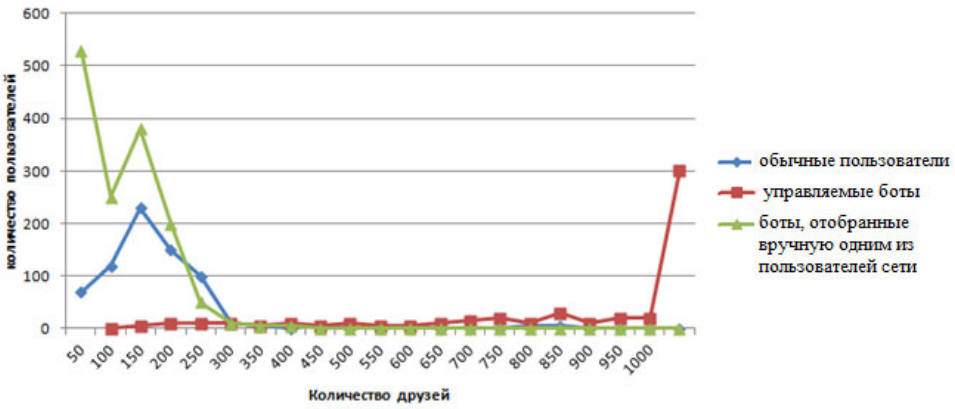

Рис. 3. Количество друзей у пользователей

Для каждой из выборок были рассчитаны следующие значения:

$$A = \frac{(T_P + T_N)}{T_P + T_N + F_P + F_N};$$

$$P = \frac{T_P}{T_P + F_P};$$

$$R = \frac{T_P}{T_P + F_N};$$

$$F_1 = \frac{2P_R}{R + P} = \frac{2T_P}{2T_P + F_P + F_N},$$

где T_P и T_N — количество правильно определенных ботов и обычных пользователей соответственно; F_P и F_N — ошибки первого и второго рода; A — доля правильных ответов (accuracy); P — точность; R — полнота; F_1 — F-мера.

Результаты для всех выборок представлены в таблице.

Точность определения ботов

Боты	A	P	R	F_1
Управляемые	0,93	0,88	0,98	0,93
Отобранные вручную одним из пользователей сети	0,93	0,94	0,95	0,95
Все	0,94	0,96	0,96	0,96

Вместе с тем данный анализ неприменим к блогерам, имеющим более тысячи подписчиков, список которых заранее известен.

Таким образом, даже простой анализ, затрагивающий лишь количество сообществ и их размер, позволяет отделить простых ботов от реальных людей.

Большинство создателей виртуальных пользователей не утруждают себя таким уровнем проработки, поскольку подобный анализ почти невозможен без привлечения автоматизированных средств. Повысить качество обнаружения возможно с помощью качественного анализа полученных сообществ: людей в сообществе обычно объединяет некоторый внешний фактор: общее место учебы или работы, схожие интересы или взгляды. Поскольку социальные сети характеризуются богатыми профилями, заполняемыми пользователями, а также позволяют создавать виртуальные сообщества по интересам, можно предположить, что для большинства выделенных сообществ образующий внешний фактор будет определяться по профилям пользователей. Соответственно, если у большей части сообществ некоторого пользователя невозможно выделить общий атрибут, это повышает вероятность того, что данный пользователь является ботом.

Помимо этого, чтобы повысить достоверность результатов, аналогичные действия стоит выполнить со всеми вершинами графа ближайшего окружения пользователя, т. е. с друзьями рассматриваемого пользователя. Очевидно, что большинство профилей друзей тоже должно соответствовать правилу: иначе, если большинство друзей пользователя боты, то и сам он, скорее всего, бот. Кроме того, если объединение всех графов ближайшего окружения друзей пользователей образует компоненту связности общего графа социальной сети или имеет малое количество связей с ней, то данное множество вершин состоит как минимум из подозрительных аккаунтов. Таким образом, сложность создания правдоподобного виртуального пользователя $O(N)$ для N пользователей возрастает на несколько порядков: злоумышленник должен не только создать правдоподобный профиль одного пользователя, но и для всех его «друзей», а также установить двусторонние связи со многими существующими пользователями социальной сети. Если изначально злоумышленнику нужно создать N профилей (один «главный» профиль и N – 1 друзей), то теперь ему придется создать $O(N_d)$ друзей друзей, где N — среднее количество друзей пользователя, а d — диаметр анализируемого графа. При этом предлагаемый способ анализа не несет существенных дополнительных расходов при условии полномасштабного мониторинга социальной сети.

Заключение. Социальные сети завоевали доверие пользователей и стали для многих основным источником информации. В настоящее время возникает необходимость выработки средств защиты от влияния и «вбросов». Одна из основных задач заключается в поиске автоматических или автоматизированных виртуальных пользователей. В данной статье был предложен метод выявления ботов на основе анализа сообществ графов их ближайшего окружения. Разработанный метод был опробован на двух выборках ботов из социальной сети ВКонтакте и показал высокие значения численных мер оценки качества алгоритмов бинарной классификации. Метод может быть использован как один из элементов подсистемы обнаружения виртуальных пользователей в системе мониторинга социальных сетей.

Литература

[1] Базенков Н.И, Губанов Д.А. Обзор информационных систем анализа социальных сетей. *Управление большими системами: сборник трудов*, 2013, № 41, с. 357–394.

[2] Riquelme F. Measuring user influence on Twitter: a survey. *Inf. Process. Manag.*, 2016, vol. 52, no. 5, pp. 949–975. DOI: 10.1016/j.ipm.2016.04.003 URL: https://www.sciencedirect.com/science/article/abs/pii/S0306457316300589

[3] Ratkiewicz J., Conover M., Meiss M. et al. Detecting and tracking political abuse in social media. *Proc. 5th AAAI ICWSM*, 2011. URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2850/0

[4] Tang L., Liu H. Community detection and mining in social media. *Synth. Lect. Data Mining Knowl. Discov.*, 2010, vol. 2, no. 1, pp. 1–137. DOI: 10.2200/S00298ED1V01Y201009DMK003 URL: https://www.morganclaypool.com/doi/abs/10.2200/S00298ED1V01Y201009DMK003

[5] Губанов Д.А., Новиков Д.А., Чхартишвили А.Г. Модели информационного влияния и информационного управления в соцсетях. *Проблемы управления*, 2009, № 5, с. 28–35.

[6] Лыфенко Н.Д. Виртуальные пользователи в социальных сетях: мифы и реальность. *Вопросы кибербезопасности*, 2014, № 5(8), с. 17–20.

[7] Ghosh R., Surachawala T., Lerman K. Entropy-based classification of 'retweeting' activity on Twitter. *arXiv.org*: веб-сайт. URL: https://arxiv.org/abs/1106.0346 (дата обращения: 25.01.2019).

[8] Dickerson J.P., Kagan V., Subrahmanian V.S. Using sentiment to detect bots on Twitter: are humans more opinionated than bots? *Proc. ASONAM*, 2014, pp 620–627. DOI: 10.1109/ASONAM.2014.6921650 URL: https://ieeexplore.ieee.org/document/6921650

[9] Wang A.H. Detecting spam bots in online social networking sites: a machine learning approach. In *Data and Applications Security and Privacy XXIV*. Springer, 2010, pp. 335–342.

[10] Wang J., Paschalidis I.Ch. Botnet detection using social graph analysis. *arXiv.org*: веб-сайт. URL: https://arxiv.org/abs/1503.02337 (дата обращения: 25.01.2019).

[11] Cao Q., Sirivianos M., Yang X., et al. Aiding the detection of fake accounts in large scale social online services. *Proc. 9th USENIX Symp. NSDI 12*, 2012, pp. 469–493.

[12] Dunbar R.I.M. Neocortex size as a constraint on group size in primates. *J. Hum. Evol.*, 1992, vol. 22, no. 6, pp. 469–493. DOI: 10.1016/0047-2484(92)90081-J URL: https://www.sciencedirect.com/science/article/pii/004724849290081J

[13] Вельц С.В. Моделирование информационного противоборства в социальных сетях на основе теории игр и динамических байесовских сетей. *Инженерный журнал: наука и инновации*, 2013, № 11(23). DOI: 10.18698/2308-6033-2013-11-991 URL: http://engjournal.ru/catalog/it/security/991.html

[14] Khondker H.H. Role of the new media in the Arab spring. *Globalization*s, 2011, vol. 8, no. 5, pp. 675–679. DOI: 10.1080/14747731.2011.621287 URL: https://www.tandfonline.com/doi/abs/10.1080/14747731.2011.621287

[15] Yang C., Harkreader R.Ch., Gu G. Die free or live hard? Empirical evaluation and new design for fighting evolving twitter spammers. *Proc. RAID 2011*. Springer, 2011, pp. 318–337.

Харламов Никита Леонидович — студент кафедры «Защита информации», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Латышева Любовь Андреевна — студент кафедры «Защита информации», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

ASSESSMENT OF THE RELIABILITY OF INFORMATION COLLECTED FROM SOCIAL NETWORKS

N.L. Kharlamov wencyre@gmail.com
SPIN-code: 3450-4547

L.A. Latysheva lubalee.work@gmail.com
SPIN-code: 7510-3315

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

A brief review of existing methods for detecting bots, based on statistical and semantic analysis of texts, behavioral analysis and graph-theoretic approach, is presented. An example of the application of the algorithm for allocating communities to solve the accompanying problem - the search for leaders of public opinion is given. A new approach to the detection of bots, based on the analysis of communities in the graph of the nearest environment of the user, is proposed. The proposed method was tested on two samples of virtual users from the social network VKontakte: managed bots and bots manually collected by one of the users. For verification, a sample of 700 legitimate users was used.

Keywords

Social network, social graph, community highlighting, search for bots, public opinion leaders, virtual users, algorithm for detecting bots

Received 21.01.2019

References

[1] Bazenkov N.I, Gubanov D.A. Information systems for social networks analysis: a survey. *Upravlenie bol'shimi sistemami: sbornik trudov* [Large-Scale Systems Control], 2013, no. 41, pp. 357–394 (in Russ.).

[2] Riquelme F. Measuring user influence on Twitter: a survey. *Inf. Process. Manag*., 2016, vol. 52, no. 5, pp. 949–975. DOI: 10.1016/j.ipm.2016.04.003
URL: https://www.sciencedirect.com/science/article/abs/pii/S0306457316300589

[3] Ratkiewicz J., Conover M., Meiss M. et al. Detecting and tracking political abuse in social media. *Proc. 5th AAAI ICWSM*, 2011.
URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2850/0

[4] Tang L., Liu H. Community detection and mining in social media. *Synth. Lect. Data Mining Knowl. Discov*., 2010, vol. 2, no. 1, pp. 1–137. DOI: 10.2200/S00298ED1V01Y201009DMK003
URL: https://www.morganclaypool.com/doi/abs/10.2200/S00298ED1V01Y201009DMK003

[5] Gubanov D.A., Novikov D.A., Chkhartishvili A.G. Informational influence and informational control models in social networks. *Problemy upravleniya*, 2009, no. 5, pp. 28–35 (in Russ.).
(Eng. version: *Autom. Remote Control*, 2011, vol. 72, no. 7, pp. 1557–1567.
DOI: 10.1134/S0005117911070216
URL: https://link.springer.com/article/10.1134%2FS0005117911070216)

[6] Lyfenko N.D. Virtual users in social networks: myths and realities. *Voprosy kiberbezopasnosti* [Cybersecurity issues], 2014, no. 5(8), pp. 17–20 (in Russ.).

[7] Ghosh R., Surachawala T., Lerman K. Entropy-based classification of 'retweeting' activity on Twitter. *arXiv.org*: website. URL: https://arxiv.org/abs/1106.0346 (accessed: 25.01.2019).

[8] Dickerson J.P., Kagan V., Subrahmanian V.S. Using sentiment to detect bots on Twitter: are humans more opinionated than bots? *Proc. ASONAM*, 2014, pp 620–627. DOI: 10.1109/ASONAM.2014.6921650 URL: https://ieeexplore.ieee.org/document/6921650

[9] Wang A.H. Detecting spam bots in online social networking sites: a machine learning approach. *Data and Applications Security and Privacy XXIV*. Springer, 2010, pp. 335–342.

[10] Wang J., Paschalidis I.Ch. Botnet detection using social graph analysis. *arXiv.org:website*. URL: https://arxiv.org/abs/1503.02337 (accessed: 25.01.2019).

[11] Cao Q., Sirivianos M., Yang X., et al. Aiding the detection of fake accounts in large scale social online services. *Proc. 9th USENIX Symp. NSDI 12*, 2012, pp. 469–493.

[12] Dunbar R.I.M. Neocortex size as a constraint on group size in primates. *J. Hum. Evol.*, 1992, vol. 22, no. 6, pp. 469–493. DOI: 10.1016/0047-2484(92)90081-J URL: https://www.sciencedirect.com/science/article/pii/004724849290081J

[13] Vel'ts S.V. Modelling information warfare in social networks based on game theory and dynamic Bayesian networks. *Inzhenernyy zhurnal: nauka i innovatsii* [Engineering Journal: Science and Innovation], 2013, no. 11(23). DOI: 10.18698/2308-6033-2013-11-991 URL: http://engjournal.ru/catalog/it/security/991.html

[14] Khondker H.H. Role of the new media in the Arab spring. *Globalizations*, 2011, vol. 8, no. 5, pp. 675–679. DOI: 10.1080/14747731.2011.621287 URL: https://www.tandfonline.com/doi/abs/10.1080/14747731.2011.621287

[15] Yang C., Harkreader R.Ch., Gu G. Die free or live hard? Empirical evaluation and new design for fighting evolving twitter spammers. *Proc. RAID 2011*. Springer, 2011, pp. 318–337.

Kharlamov N.L. — Student, Department of Information Security, Bauman Moscow State Technical University, Moscow, Russian Federation.

Latysheva L.A. — Student, Department of Information Security, Bauman Moscow State Technical University, Moscow, Russian Federation.