

DOI: 10.26794/2220-6469-2020-14-2-44-50
УДК 519.17(045)
JEL C61, C63

Выявление ботов в социальных сетях на примере LiveJournal

А.А. Кочкаров^а, Н.В. Калашников^б, Р.А. Кочкаров^с

^{а, б, с} Финансовый университет Москва, Россия

^а <https://orcid.org/0000-0002-3232-5331>; ^б <https://orcid.org/0000-0002-6111-8999>;

^с <https://orcid.org/0000-0003-3186-3901>

АННОТАЦИЯ

Социальные сети прочно вошли в жизнь миллиардов пользователей Интернета по всему миру. В социальных сетях общаются, играют в онлайн-игры, совершают покупки, организуют онлайн-мероприятия — обмениваются контентом из всех сфер жизни [1, 2]. Наиболее популярные и известные в России сервисы — ВКонтакте (vk.com), Youtube.com, Facebook.com, Одноклассники (Ok.ru) и др. Интерфейсы подобных платформ позволяют создавать специальные приложения (боты), выполняющие действия в роли псевдопользователей — фейковых аккаунтов. В данной работе авторы предлагают подход к выявлению ботов на примере социальной сети LiveJournal. Для этого исследуются характеристики эгографа пользователя. Также в статье проведен сравнительный анализ результатов работы алгоритмов классификации.

Ключевые слова: выявление ботов; анализ структуры эгографа; классификация пользователей; социальная сеть; выявление сообществ

Для цитирования: Кочкаров А.А., Калашников Н.В., Кочкаров Р.А. Выявление ботов в социальных сетях на примере LiveJournal. *Мир новой экономики*. 2020;14(2):44-50. DOI: 10.26794/2220-6469-2020-14-2-44-50

ORIGINAL PAPER

Identifying Bots in Social Networks Using the Example of LiveJournal

A.A. Kochkarov^а, N.V. Kalashnikov^б, R.A., Kochkarov^с

^{а, б, с} Finance University, Moscow, Russia

^а <https://orcid.org/0000-0002-3232-5331>; ^б <https://orcid.org/0000-0002-6111-8999>;

^с <https://orcid.org/0000-0003-3186-3901>

ABSTRACT

Social networks have firmly entered the lives of billions of global Internet users worldwide. They communicate in social networks, play online games, make purchases, organise online events — exchange content from all walks of life [1, 2]. The most popular and well-known services in Russia are Vkontakte (vk.com), Youtube.com, Facebook.com, Odnoklassniki (Ok.ru), etc. The interfaces of such platforms allow — fake accounts. In this paper, we propose an approach to detect bots using the LiveJournal social network as an example. For this, we investigated the characteristics of the user's egograph and performed a comparative analysis of the results of the classification algorithms.

Keywords: bots identification; egograph structure analysis; users classification; social network; community identification

For citation: Kochkarov A.A., Kalashnikov N.V., Kochkarov R.A. Identifying bots in social networks using the example of LiveJournal. *Mir novoi ekonomiki = The World of the New Economy*. 2020;14(2):44-50. DOI: 10.26794/2220-6469-2020-14-2-44-50



БОТЫ В СОЦИАЛЬНЫХ СЕТЯХ

Сегодня социальные медиаресурсы с сотнями миллионов пользователей способствуют созданию большого количества ботов — искусственных профилей, имитирующих человеческое поведение.

Широкое определение понятия «бот» (сокращение от «робот») — это специализированное приложение или программа, выполняющая действия по заранее заданному алгоритму, через интерфейсы, предназначенные для реальных пользователей. Есть большое количество полезных ботов, которые используются ИТ-специалистами в своей ежедневной деятельности, например, обслуживающие серверную инфраструктуру для многократного повторения рутинной работы с высокой скоростью. Также популярны боты для общения с клиентами (чат-боты), отвечающие на простые вопросы, либо осуществляющие простой набор действий по заявке пользователя. Такие чат-боты демонстрируют внешнее наличие интеллекта, как будто пользователь общается с настоящим человеком.

В то же время боты могут проявляться как вредоносные или условно-вредоносные программы. Интернет-боты могут использоваться для координации групповых сетевых атак на компьютеры (например, DoS-атаки), для мошеннических действий с целью хищения информации или денежных средств. Также популярны игровые боты для «зарабатывания» игровых денег с целью их дальнейшего обмена на реальные. Широко применяемые и известные спам-боты распространяют рекламную информацию на сайтах сети Интернет или в социальных сетях. Боты применяются для выкупа дешевых билетов на различные мероприятия, авиа- и другие виды транспортного сообщения, для последующей продажи по уже увеличенной рыночной цене.

Основным способом борьбы с вредоносными ботами выступает одна из форм теста машины Тьюринга — графические задания, простые для человека, но невыполнимые для компьютерных программ. Тем не менее, с развитием технологий машинного обучения, в скором будущем потребуются более сложные или специальные методы для выборочного блокирования ботов.

Таким образом, при обсуждении ботов в основном рассматривается программная среда глобальной сети Интернет. Исследование ботов закрытых или локальных сетей не представляет интерес

в силу малого количества клиентов и достаточного уровня защищенности локальных сетей. В таких сетях обычно используется «полезные» боты для решения ежедневных задач.

Переход в виртуальную среду (в частности — в социальные сети) большого количества реальных сервисов, услуг, продуктов создает новые вызовы, с созданием ботов, имитирующих поведение реальных пользователей: автоматическое создание контента, распространение достоверного или недостоверного мнения, пропаганда противоправных действий и пр. Такие боты создают контент и взаимодействуют с людьми в социальных сетях, подражая поведению людей и влияя на их поведение.

Как было сказано ранее, некоторые могут быть безобидными или даже полезными, например, способны агрегировать информацию из различных новостных лент и предоставлять ее в одном месте или выступать в качестве онлайн-консультанта коммерческих компаний для первичного общения с клиентами.

К сожалению, за появлением новых массовых информационных технологий следует развитие новых способов для их злоупотребления и использования в противоправных действиях. В случае социальных сетей появилась категория ботов, представляющая собой вредоносное программное обеспечение, разработанное с целью нанесения ущерба пользователям сети. Такие боты вводят пользователей в заблуждение, манипулируют мнением с помощью слухов, спама, дезинформации, клеветы или создают информационный шум для нивелирования достоверной информации. Такая деятельность ботов может нанести вред обществу на самых разных уровнях. Например, боты могут искусственно раздуть поддержку политических кандидатов, влияя на результаты выборов. Подобное было зафиксировано во время президентских выборов в США в 2016 г. и президентских выборов во Франции в 2017 г.

На *рис. 1* приведено соотношение интернет-трафика, сгенерированного в 2015 г. реальными пользователями, а также «хорошими» и «плохими» ботами.

ВЫЯВЛЕНИЕ БОТОВ В СЕТИ LIVEJOURNAL

В исследовании изучалась выборка из 2700 пользователей LiveJournal, среди которых 700 — реальные, размеченные вручную,



и 2000 — пользователи-боты [3]. Это так называемые боты из «фабрики троллей» — организованной группы пользователей, занимающихся формированием общественного мнения и его манипуляцией в глобальной сети Интернет, зачастую за денежное вознаграждение. В свободном доступе имеется информация о пользователях — структура эгографов, т.е. графов друзей.

В исследовании было обучено несколько алгоритмов классификации (машинное обучение) с целью отделения ботов от реальных пользователей. В качестве признакового описания пользователя выступает информация, агрегированная из его эгографа: количество вершин, коэффициент кластеризации, количество ребер и др. Также был применен алгоритм выделения сообществ для сбора информации о структуре сообществ пользователя. В качестве признаков, связанных со структурой, рассматривались количество и средний размер сообществ пользователя. Всего было выделено 16 признаков, среди них: количество друзей, количество сообществ, коэффициент кластеризации, распределение долей вершин по 5 крупнейшим сообществам и др.

Предположение о том, что, обладая информацией лишь об эгографе пользователя, возможно классифицировать его как бота или настоящего пользователя, связано с числом Данбара [4]. Число Данбара ограничивает количество постоянных социальных связей для одного человека в диапазоне от 100 до 230 (чаще используется среднее значение, равное 150). Также было обнаружено, что пользователям социальной сети Facebook свойственно иметь не более 6 крупных сообществ [5]. Школа, университет, работа, друзья — примеры основных сообществ, присущих большинству людей. Количество сообществ, превышающее значение 6, может свидетельствовать об активной социальной жизни пользователя или о подозрительной деятельности профиля.

На рис. 2 представлен сравнительный анализ количественных показателей эгографов ботов и реальных пользователей. Реальные пользователи LiveJournal и Facebook в большинстве своем имеют 6 сообществ.

Для ботов такое распределение бимодально и имеет максимумы в точках 3 и 7 (количество сообществ). Среднее количество друзей у реальных пользователей оказалось равным 148, что согласуется с числом Данбара. Только 3% пользователей имеют количество друзей больше 250.



Рис. 1 / Fig. 1. Глобальный интернет-трафик, сгенерированный реальными пользователями и ботами в 2015 г. / Global Internet traffic generated by real users and bots in 2015

Источник / Source: Отчет компании Imperva Incapsula (Bot Traffic Report 2016) / Imperva Incapsula 'Bot Traffic Report 2016'. URL: <https://www.imperva.com/blog/bot-traffic-report-2016/>.

Аналогичное распределение для ботов, несмотря на то, что достигает максимума эмпирической плотности в районе 150, имеет длинный «хвост». У 30% ботов количество друзей больше 250, что согласуется с выводами о стремлении ботов создать больше связей для увеличения социального веса. Величина модулярности демонстрирует, насколько структура сообществ для данного графа отличается от случайно сгенерированного графа с теми же характеристиками. Логично предположить, что боты, стремясь получить как можно больше друзей, создают связи довольно хаотично. Данный процесс развития эгосети бота в некотором смысле схож со случайным процессом генерации графа и как следствие имеет низкое значение модулярности.

Следующим шагом было обучение алгоритма классификации пользователей: имеющуюся выборку пользователей разделили на обучающее и тестовое множества в пропорции: 75% на 25%. Воспользовавшись алгоритмами: логистическая регрессия, случайный лес, метод опорных векторов и наивный байесовский классификатор мы получили следующие результаты (см. таблицу).

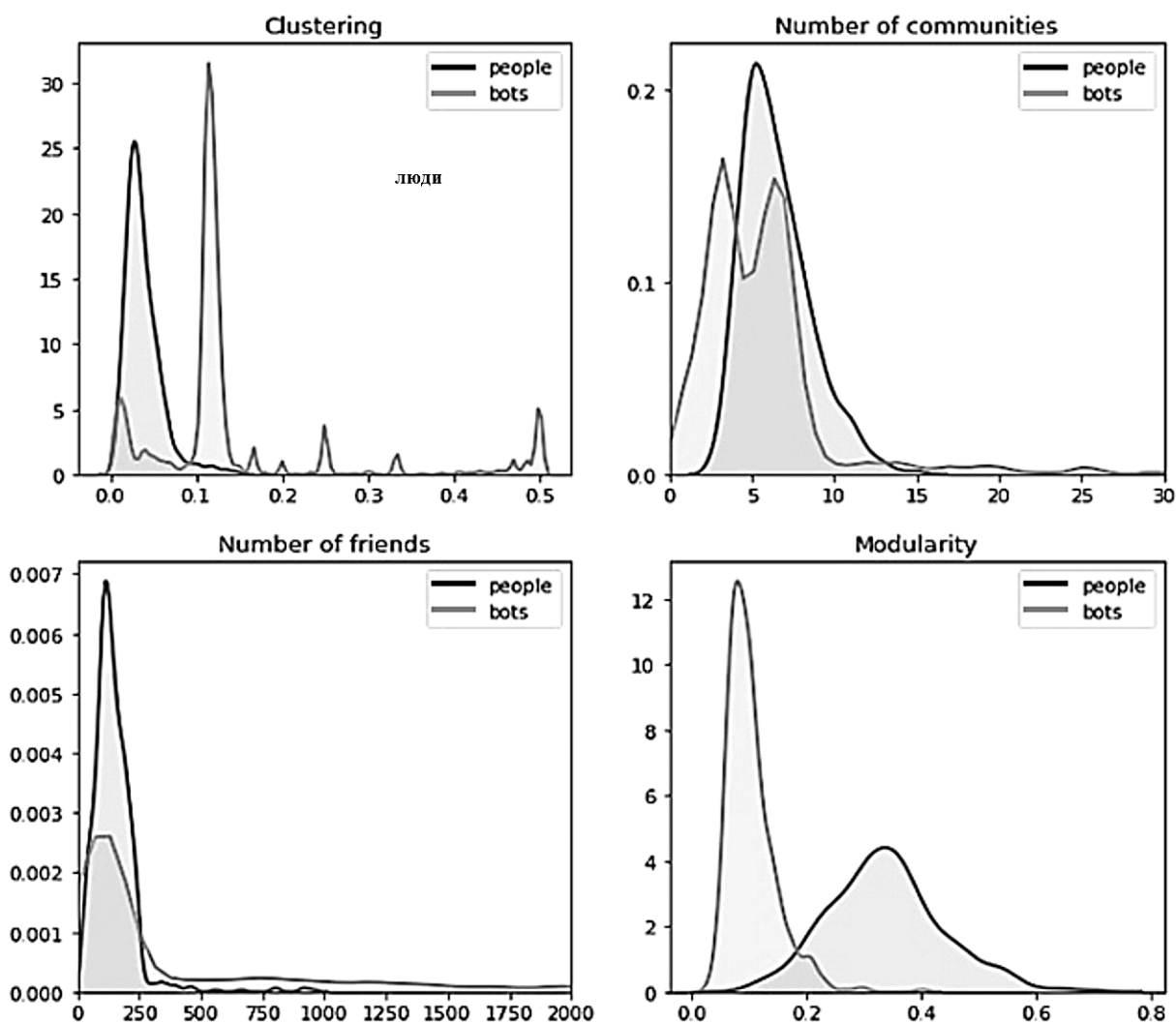


Рис. 2 / Fig 2. Распределение количественных показателей эгографов ботов и реальных людей. Вверху слева – коэффициент кластеризации, вверху справа – количество сообществ, внизу слева – количество друзей, внизу справа – значение модулярности / Distribution of quantitative indicators of egographs of bots and real people. Top left – clustering coefficient, top right – number of communities, bottom left – number of friends, bottom right – modularity value

Источник / Source: составлено авторами / compiled by the authors.

Таблица / Table

Результаты работы алгоритмов классификации / The results of the classification algorithms

	Точность	Полнота	F1-мера
Логистическая регрессия	0,966	0,963	0,965
Случайный лес	0,990	0,988	0,989
Метод опорных векторов (SVM)	0,983	0,973	0,978
Наивный байесовский классификатор (Naive bayes)	0,939	0,663	0,778

Источник / Source: составлено авторами / compiled by the authors.

Примененные алгоритмы показали высокую точность выявления ботов. Случайный лес справился с задачей чуть лучше линейных моделей. Оптимальные параметры моделей подбирались в ходе перекрестной проверки по 5 интервалам (K-fold cross validation).

Боты, стремясь получить как можно больше друзей, создают связи довольно хаотично. Данный процесс развития эгосети бота в некотором смысле схож со случайным процессом генерации графа и как следствие имеет низкое значение модулярности.

ЗАКЛЮЧЕНИЕ

На основе предложенного подхода решена важная задача анализа социальных сетей — выявление в них ботов. Экспериментально исследовано предположение о классификации пользователя (бот — реальный пользователь) по набору доступных показателей — количеству сообществ пользователя и его друзей. Продемонстрировано, что при помощи алгоритмов классификации с использованием информации о структуре сообществ эгосети возможно выявлять аккаунты ботов с высокой точностью.

Представляется важным изучение структурных характеристик эгосетей, социальных се-

тей и сообществ в них. В качестве инструмента моделирования или распознавания структуры социальных сетей используются динамические графы. Социальная сеть — это граф с изменяющейся структурой ребер во времени. Жизненный цикл социальной сети состоит из нескольких выраженных этапов, на первом этапе происходит рост числа вершин (набор количества пользователей) и активно формируются связи между ними. На следующем этапе, в связи с ограничением количества пользователей, рост числа вершин замедляется, но при этом изменяется структура связей — появляются новые ребра и исчезают старые. Эти два этапа органично переходят один в другой и распределены во времени, как топологическом, так и в реальном [6–11].

Для реализации задачи поиска сообществ предлагается использовать инструментарий специального подкласса динамических графов — предфрактальных графов, порождение которых соответствует росту социальной сети, а блоки и затравки рассматриваются в качестве сообществ и эгосетей социальной сети.

Развитие инструментальной базы моделирования, в частности использования динамических и предфрактальных графов, позволит расширить круг задач в социальных сетях, в числе которых многокритериальные (многопараметрические) задачи, задачи с множественными и нечеткими весами, прогнозные задачи с заданным уровнем надежности, разработка параллельных алгоритмов и др. Для специалистов в области информационной безопасности будут интересны задачи структурного разрушения сетей, структурная устойчивость и устойчивость оптимизационных задач в социальных сетях [12, 13].

СПИСОК ИСТОЧНИКОВ

1. Scott J. Social Network Analysis: A Handbook. London: SAGE Publications Ltd; 2000. 224 p.
2. Fortunato S., Castellano C. Encyclopedia of Complexity and Systems Science. Springer. *Rough Sets in Decision Making*. 2009;7753–7786.
3. Чесноков В. О., Ключарёв П. Г. Современные методы выделения сообществ в социальных сетях. *Наука и Образование: Научное издание*. 2017;(4):137–152.
4. Dunbar R. I. M. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*. 1992;22(6):469–493.
5. Калашников Н. В., Анализ социальных графов пользователей Facebook. *Современная математика и концепции инновационного математического образования*. 2018;5(1):408–413.
6. Perepelitsa V. A., Kochkarov A. M., Sergienko I. V. Recognition of fractal graphs. *Cybernetics and Systems Analysis*. 1999;35(4):572–585.
7. Кочкаров А. А., Кочкаров А. М., Салпагарова Л. У. Моделирование разрушения сложных сетевых систем: теоретико-графовый подход. *Известия ЮФУ. Технические науки*. 2009;5(94):234–240.



8. Кочкаров А.А., Кочкаров Р.А., Малинецкий Г.Г. Некоторые аспекты динамической теории графов. *Журнал вычислительной математики и математической физики*. 2015;55(9):1623–1629.
9. Кочкаров А.А., Кочкаров Р.А. Параллельный алгоритм поиска кратчайшего пути на предфрактальном графе *Журнал вычислительной математики и математической физики*. 2004;44(6):1157–1162.
10. Кочкаров Р.А. Многовзвешенные предфрактальные графы с недетерминированными весами. Приложения в экономике, астрофизике и сетевых коммуникациях. М.: Ленанд; 2017. 432 с.
11. Кочкаров А.А., Салпагаров С.И., Кочкаров Р.А. О количественных оценках топологических характеристик предфрактальных графов. *Известия ТРТУ*. 2004;8(43):298–301.
12. Биккузина А.И., Жуков А.О., Никольский Ю.В., Буханец Д.И. Подход к решению задачи упорядочения альтернатив в диалоговой системе моделирования принятия решений при информационно-аналитическом обеспечении оценки и прогноза экологического состояния территорий эксплуатации крупных технических комплексов. *Новые исследования в разработке техники и технологий*. 2014;(1):33–39.
13. Гладышев А.И., Жуков А.О. Использование в автоматизированной системе контроля полномочий биометрической идентификации. *Вестник Российского нового университета. Серия: Сложные системы: модели, анализ и управление*. 2013;(4):95–98.

REFERENCES

1. Scott J. *Social Network Analysis: A Handbook*. London: SAGE Publications Ltd; 2000. 224 p.
2. Fortunato S., Castellano C. *Encyclopedia of Complexity and Systems Science*. Springer Rough Sets in Decision Making. 2009; 7753–7786.
3. Chesnokov V.O., Klyucharyov P.G. Modern methods of highlighting communities in social networks. *Nauka i Obrazovaniye: Nauchnoye izdaniye*. 2017;(4):137–152. (In Russ.).
4. Dunbar R.I.M. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*. 1992;22(6):469–493.
5. Kalashnikov N.V., Analysis of social graphs of Facebook users. *Sovremennaya matematika i kontseptsii innovatsionnogo matematicheskogo obrazovaniya*. 2018;5(1):408–413. (In Russ.).
6. Perepelitsa V.A., Kochkarov A.M., Sergienko I.V. Recognition of fractal graphs. *Cybernetics and Systems Analysis*. 1999;35(4):572–585.
7. Kochkarov A.A., Kochkarov A.M., Salpagarova L.U. Modeling the destruction of complex network systems: graph-theoretic approach. *Izvestiya YUFU. Tekhnicheskiye nauki*. 2009;5(94):234–40. (In Russ.).
8. Kochkarov A.A., Kochkarov R.A., Malinetskiy G.G. Some aspects of dynamic graph theory. *Zhurnal vychislitel'noy matematiki i matematicheskoy fiziki*. 2015;55(9):1623–1629. (In Russ.).
9. Kochkarov A.A., Kochkarov R.A. Parallel algorithm for finding the shortest path on a pre-fractal graph. *Zhurnal vychislitel'noy matematiki i matematicheskoy fiziki*. 2004;44(6):1157–1162. (In Russ.).
10. Kochkarov R.A. Multi-weighted prefractal graphs with non-deterministic weights. Applications in economics, astrophysics and network communications. Moscow: Lenand; 2017. (In Russ.).
11. Kochkarov A.A., Salpagarov S.I., Kochkarov R.A. On quantitative estimates of the topological characteristics of prefractal graphs. *Izvestiya TRTU*. 2004;8(43):298–301. (In Russ.).
12. Bikkuzina A.I., Zhukov A.O., Nikolsky Yu.V., Bukhanets D.I. An approach to solving the problem of ordering alternatives in the dialogue system for modeling decision-making with information and analytical support for assessing and predicting the ecological state of the territories of operation of large technical complexes. *Novyye issledovaniya v razrabotke tekhniki i tekhnologii*. 2014;(1):33–39. (In Russ.).
13. Gladyshev A.I., Zhukov A.O. Use in an automated control system of authority of biometric identification. *Vestnik Rossiyskogo novogo universiteta. Seriya: Slozhnyye sistemy: modeli, analiz i upravleniye*. 2013;(4):95–98. (In Russ.).

ИНФОРМАЦИЯ ОБ АВТОРАХ

Азрет Ахматович Кочкаров — кандидат физико-математических наук, доцент Департамента анализа данных, принятия решений и финансовых технологий, Финансовый университет, Москва, Россия akochkar@gmail.com



Никита Владимирович Калашников — соискатель, Финансовый университет, Москва, Россия
nikita_007_94@mail.ru

Расул Ахматович Кочкаров — кандидат экономических наук, доцент Департамента анализа данных, принятия решений и финансовых технологий, Финансовый университет, Москва, Россия
rasul_kochkarov@mail.ru

ABOUT THE AUTHORS

Azret A. Kochkarov — Cand. Sci. (Phys.-Math.), Associate Professor, Department of Data Analysis, Decision Making and Financial Technologies, Financial University, Moscow, Russia
akochkar@gmail.com

Nikita V. Kalashnikov — Postgraduate student, Financial University, Moscow, Russia
nikita_007_94@mail.ru

Rasul A. Kochkarov — Cand. Sci. (Econ.), Associate Professor, Department of Data Analysis, Decision Making and Financial Technologies, Financial University, Moscow, Russia
rasul_kochkarov@mail.ru

Заявленный вклад авторов:

Кочкаров А. А. — подготовка обзора литературы, формулировка постановки задачи, выбор инструментария анализа, выводы, подготовка текста статьи.

Калашников Н. В. — сбор данных для проведения исследования, анализ данных, компьютерное моделирование, подготовка текста статьи, подготовка списка источников.

Кочкаров Р. А. — проверка результатов моделирования, подготовка текста статьи, критический анализ, оформление, перевод на англ., подготовка списка источников.

Declared contribution of the authors:

Kochkarov A. A. — preparation of literature review, formulation of the problem statement, selection of analysis tools, conclusions, preparation of the article text.

Kalashnikov N. V. — data collection for research, data analysis, computer modelling, preparation of the article text, preparation of references.

Kochkarov R. A. — verification of simulation results, preparation of the article text, critical analysis, design, translation into English, preparation of references.

Статья поступила 15.11.2019; принята к публикации 15.12.2019.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article received on 15.11.2019; accepted for publication on 15.12.2019.

The authors read and approved the final version of the manuscript.