УДК 004.77, 519.25

ФАКТОРЫ РАСПРОСТРАНЕНИЯ ИНФОРМАЦИИ В ИНТЕРНЕТ-СООБЩЕСТВАХ

М. И. Купцов, к.ф.-м.н., доцент, доцент Рязанского государственного радиотехнического университета, Рязань, Россия;
orcid.org/0000-0002-6814-6423, e-mail: kuptsov_michail@mail.ru
В. А. Минаев, д.т.н., профессор, профессор Московского университета МВД России им. В. Я. Кикотя, Москва, Россия; профессор Рязанского государственного радиотехнического университета, Рязань, Россия;
orcid.org/ 0000-0002-5342-0864, e-mail: m1va@yandex.ru

Рассматривается задача изучения статистических характеристик распространения информации в маркетинговых, блогерских и политических сообществах в различных социальных сетях. ***Целью работы*** *является определение и изучение факторов, влияющих на социальное поведение пользователей сетей. Проявлением поведения может быть, например, изменение вовлечённости людей в Интернет-торговлю или в общественно-политические мероприятия, в том числе – деструктивного характера. Исследованы эмпирические законы распределения различных показателей сообществ в сетях. Проанализированы взаимосвязи между показателями. На основе кластерного анализа предложена классификация сообществ социальных сетей, отмечены характерные особенности кластеров. Обоснованы регрессионные модели прогнозирования изменения количества пользователей в сообществах.*

Ключевые слова: *социальная сеть, Интернет-сообщество, распространение информации, моделирование, статистическая закономерность.*

DOI: 10.21667/1995-4565-2021-76-81-89

Введение

Развитие информационных технологий позволило не только виртуально общаться при помощи социальных сетей, но и вести деловую жизнь, зарабатывать, вести активную политическую деятельность и делать многое другое, ранее осложненное расстояниями, скоростью обмена информацией, да и психологическими факторами в конце концов. Социальные сети успешно используются для организации гуманитарных, благотворительных и развлекательных акций [1, 2]. Увеличение доступности Интернет-возможностей приводит к возрастанию интереса государственных институтов, политических групп, промышленных корпораций к социальным сетям. Особенно большой прирост количества их пользователей произошёл в последние два года [3-6], что связано с обрушившейся пандемией коронавируса и, как следствие, с уменьшением возможностей традиционного общения и увеличением времени, которое люди проводят у компьютеров.

В этой связи усилился интерес к изучению компьютерных сетей для анализа взаимодействия между пользователями, прогнозирования их поведения, классификации, моделирования путей и интенсивности распространения информационных потоков [7, 8].

При этом выделяются два основных подхода к исследованиям [2, 9]: статический и динамический. На статический подход акцент делается тогда, когда топологические аспекты сетей более важны, чем временные [2].

При динамическом подходе к изучению социальных сетей [10] выделяются структурное и статистическое направления. Структурные исследования характеризуются выявлением причинно-следственных связей и основных информационных потоков. Статистические же сосредоточены на анализе стохастических закономерностей в информационных процессах социальных сетей [11].

Одними из самых распространённых при анализе социальных сетей являются методы, основанные на теории графов [8, 12-14]. Вершины графов, как правило, представляются агентами, рёбра же характеризуют взаимодействия (информационные обмены) между ними. Для использования графовых методов требуется изучение топологии социальных сетей, построение классификации агентов и выявление характеристик их взаимодействий.

Динамические графовые методы и модели терминологически зачастую отличаются от статических [15-17]. Так, при изучении эволюции социальных сетей с помощью динамических графовых методов рассматриваются изменения рёбер во времени [18, 19]. В частности, рёбра, как и вершины, могут быть активны не постоянно, а только периодически. Характеристики рёбер могут зависеть как от сетевых свойств пользователей, так и от их личных свойств [20, 21].

Достаточно распространёнными являются методы анализа социальных сетей, основанные на эпидемиологических моделях [8, 22, 23]. В таких моделях, базирующихся на системах дифференциальных уравнений, моделирование происходит в терминах «информационное заражение», «информационный иммунитет», «информационное воздействие», «информационное противодействие» [22].

При их реализации приходится преодолевать значительные сложности, связанные с формализацией многих качественных параметров. Поскольку обмен информацией для части агентов происходит периодически, то возникает существование периодических решений и устойчивых интегральных многообразий соответствующих систем дифференциальных уравнений [24].» надо «Поскольку обмен информацией для части агентов происходит периодически, то возникает проблема существования периодических решений и устойчивых интегральных многообразий соответствующих систем дифференциальных уравнений [24]. В частности, при моделировании распространения информации в социальных сетях могут выделяться социальные группы, устойчивые к негативным (позитивным) информационным воздействиям. Более подробно различные методы и модели анализа социальных сетей представлены в [8].

Таким образом, анализ, моделирование и прогнозирование поведения информационных потоков в социальных сетях, изучение их стохастических закономерностей является весьма актуальной задачей.

В данной работе изучаются вероятностные закономерности и специфические характеристики распространения информации в компьютерных социальных сетях в течение промежутков времени от суток до года. Обсуждаются результаты, которые продолжают и развивают статьи [11, 22, 24, 25].

Теоретическая часть

При постановке задач исследования выделены четыре типа сообществ и личных аккаунтов в социальных сетях, условно названных «индивидуальные», «блогерские», «маркетинговые» и «политические». К «индивидуальным» отнесены те личные аккаунты, в которых более 80 % публикаций отражают лишь процессы общения и представления себя в Интернет-сообществе. К «блогерским» отнесены личные аккаунты и сообщества, в которых более 80 % публикаций служат для увеличения просмотров контента с целью получения рекламных контрактов. В «маркетинговые» вошли те, которые более чем в 80 % публикаций нацелены на увеличение продаж товаров и услуг. Наконец, к «политическим» отнесены личные аккаунты и сообщества, в которых более 80 % публикаций имеют агитационно-пропагандистский характер и направлены на привлечение сторонников политической группы или отдельного политика. Такие публикации используются, как правило, для получения электорального преимущества при проведении выборов в органы власти различных уровней. Существуют и сообщества «смешанного» типа, например, средства массовой информации, имеющие свои страницы в соцсетях. В настоящем исследовании сообщества и личные аккаунты «смешанного» и «индивидуального» типов не рассматривались.

Каждый пользователь социальной сети отнесен к одному из четырёх непересекающихся классов: «посетитель», «подписчик», «подписчик-посетитель», «активный сторонник». К «активным сторонникам» отнесем тех пользователей социальной сети, которые являются потребителями товаров или услуг маркетингового сообщества, а также участвуют в реальных акциях, связанных с агитационно-пропагандистской деятельностью политических сообществ, в том числе в протестных акциях.

«Посетитель» – это пользователь социальной сети, который в конкретный день, когда проводилось исследование, посетил Интернет-страницу сообщества, но при этом не является его «активным сторонником».

«Подписчик» – это пользователь, у которого оформлена подписка на новости сообщества, но который в указанный день не посещал его Интернет-страницу и не является активным последователем данного сообщества.

«Подписчик-посетитель» – это пользователь, у которого оформлена подписка на новости сообщества, и он посетил Интернет-страницу сообщества, но при этом не является «активным сторонником» данного сообщества.

Еще раз подчеркнем, что любой пользователь социальной сети, посетивший в день исследования страницу сообщества, может быть отнесён только к одному классу. Пересечение этих классов – пустое множество.

Отметим, что основной целью политических и маркетинговых сообществ является увеличение количества именно «активных сторонников», тогда как «блогерские сообщества» такой цели не ставят в принципе.

Предлагаемая классификация дает возможность найти и изучить факторы, влияющие на количество «активных сторонников» сообществ в социальных сетях. Для этого решалась задача определения эмпирических законов распределения различных характеристик популярности сообществ в социальных сетях, а также их взаимосвязей.

Для сравнения эмпирических распределений вероятностей с теоретическими использовался критерий согласия Колмогорова – Смирнова. Для определения статистически значимых факторов, оказывающих влияние на те или иные переменные, использовались инструменты построения линейной и нелинейной парных регрессий, а также множественной линейной регрессии. Проверка статистической значимости коэффициентов регрессий проводилась по критерию Стьюдента, а значимость уравнения в целом – по критерию Фишера. Для доказательства статистической значимости различий выделенных групп сообществ использовался иерархический кластерный анализ, реализованный в программном пакете SPSS. Мерой расстояния служила евклидова метрика, а в качестве метода кластеризации – метод Варда. Для выявления переменных, определяющих статистически значимые различия между кластерами, использовался критерий Пирсона.

Экспериментальные исследования

Проведён статистический анализ эмпирических распределений посещений страниц сообществ, охвата подписчиков, количества отметок «нравится», репостов, комментариев, изменения количества подписчиков 24 сообществ и личных аккаунтов (10 – «блогерских», 10 – «политических» и 5 – «маркетинговых») в сетях ВКонтакте и «Instagram». Эмпирические распределения сравнивались с тремя теоретическими: нормальным, равномерным и экспоненциальным. Выборки формировались на временных периодах 50 и более подряд следующих друг за другом дат. Рассматривались два периода: от 50 до 100 суток («короткий временной период», КВП) и свыше 100 суток («длинный временной период», ДВП).

Анализ показал, что на КВП в 58 % случаев эмпирические распределения близки к нормальным. Их доля значительно выше, чем доли соответствия равномерному и экспоненциальному распределениям (8 % и 3 %, соответственно).

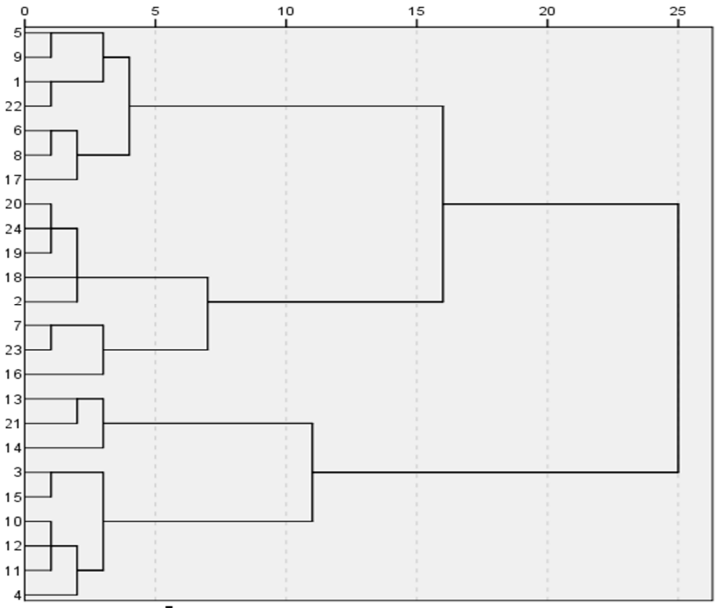
На ДВП для рассматриваемых показателей наиболее часты распределения, близкие к нормальным для количества посещений, количества репостов и охвата подписчиков. Изуче-

ние статистических взаимосвязей также показало, что на ДВП в 64 % случаев между изучаемыми переменными наблюдаются значимые парные и множественные регрессионные уравнения (от линейных до логарифмических) с коэффициентами детерминации до 0,73. При этом чаще других статистически значимые регрессии наблюдаются для зависимой переменной «относительные посещения» (71 %). Для других зависимых переменных количество статистически значимых регрессий несколько ниже и находится в интервале от 29 % до 67 %.

В то же время для КВП уравнения регрессий оказались статистически значимыми в 52 % случаев с коэффициентами детерминации до 0,86. Для зависимой переменной «относительные посещения» процент статистически значимых регрессий оказался наиболее высоким – 59 %.

По КВП имелось значительно большее количество статистических данных. В качестве переменных для кластеризации выбраны пять наиболее характерных для большинства изучаемых сообществ: «доля значимых нормальных распределений количества посещений», «доля значимых нормальных распределений количества подписчиков», «доля значимых нормальных распределений количества отметок «нравится», «доля значимых парных линейных регрессий для зависимой переменной «относительное количество посещений»», «доля значимых парных линейных регрессий для зависимой переменной «относительное количество подписчиков»».

Результаты кластеризации показаны на рисунке.



Дендрограмма кластеризации по методу Варда
Dendrogram of clustering by the Ward method

При изъятии из дендрограммы после детального анализа объектов, очевидно чужеродных, не интерпретируемых в рамках исследуемых сообществ, выявлена следующая картина: первый кластер содержит 5 блогерских; второй кластер – 6 политических; третий кластер 5 – маркетинговых сообществ.

Таким образом, выделенные однородные Интернет-группы (блогерские, политические и маркетинговые) имеют характерные особенности в стохастических эмпирических закономерностях.

При анализе переменных, характеризующих полученные кластеры, выявлены существенные расхождения в количестве соответствий эмпирических распределений теоретическому нормальному для количества комментариев и отметок «нравится» на КВП при переходе от одного типа сообщества к другому (блогерские, политические и маркетинговые) со значениями асимптотической значимости менее 0,01.

Попарное сравнение количества соответствий теоретическому нормальному распределению позволило выявить существенные отличия по обоим указанным показателям для марке-

тинговых сообществ (асимптотические значимости менее 0,01). Блогерские и политические сообщества по количеству соответствий эмпирических распределений теоретическим различаются лишь для переменной «количество посещений» (значимость 0,05).

При сравнении количеств статистически значимых регрессий в сообществах различного типа (блогерские, политические и маркетинговые) показаны отличия в 67 % случаев. Различия относились к трём регрессиям «количество посещений от количества отметок «нравится»», «количество подписчиков от количества отметок «нравится»» и «количество подписчиков от количества комментариев» (уровни значимости менее 0,04).

При попарном сравнении количества значимых регрессий для различных типов сообществ выявлены различия для маркетинговых типов сообществ в 46 % случаев, для политических – в 29 % и для блогерских – в 21 %. Различия по количеству значимых регрессий для пары типов «маркетинговые-блогерские» составил 33 % от рассмотренных случаев, для пары «маркетинговые-политические» – 58 %, для пары «блогерские-политические» – 13 %.

На наш взгляд, выявленные различия означают, что статистические взаимосвязи и эмпирические распределения формируются в соответствии с различающимися механизмами для разных сообществ.

Для прогнозирования количества «активных сторонников» маркетинговых групп получены уравнения линейных парных и множественных регрессий между количеством заказов и другими переменными. Для зависимой переменной «количество посещений» 47 % построенных регрессий оказались статистически значимыми. Для остальных зависимых переменных доли значимых линейных парных регрессий составили от 0 до 29 % (в среднем – 12 %).

Такой результат позволил предположить, что количество «активных сторонников» маркетинговых групп тесно связано с количеством просмотров контента. Для проверки этой гипотезы построены уравнения линейных множественных регрессий и уравнения парных регрессий 11 видов с зависимой переменной – «количество заказов за неделю». Все независимые переменные также суммировались за неделю (например, рассматривались переменные «количество посещений за неделю», «количество публикаций за неделю» и т.п.). Причём регрессии строились как без временных сдвигов, так и со сдвигами на один или два временных периода (неделю). Данный подход позволил построить по одной-две статистически значимых регрессии во всех маркетинговых сообществах. В 80 % случаев – это регрессии без временных сдвигов, в 60 % случаев – со сдвигом на один временной период. При этом для 80 % маркетинговых сообществ удалось получить только парные регрессии, причём для наибольшего коэффициента детерминации во всех случаях зависимая переменная одна и та же – «количество посещений за неделю». Для 20 % маркетинговых сообществ с удовлетворительной интерпретацией получены статистически значимые линейные множественные регрессии, имеющие по две независимые переменные: «количество посещений за неделю» и «суммарное количество подписчиков за неделю» или «количество посещений за неделю» и «суммарное количество комментариев за неделю».

Анализ маркетинговых сообществ позволяет сделать вывод о том, что основным фактором, влияющим на увеличение заказов, является количество просмотров контента. Влияние этого фактора доказано для всех маркетинговых сообществ, рассмотренных в исследовании. Вторым по значению фактором оказалось количество подписчиков, влияние его доказано для 20 % маркетинговых сообществ. Остальные изученные факторы либо не оказывали никакого влияния на изменение количества заказов, либо это влияние было со сравнительно небольшим коэффициентом детерминации.

Для исследования количества «активных сторонников» политических групп изучены массовые политические акции на длительных временных периодах, проходивших в России и Беларуси. Первый период – с 13.07.2019 по 30.09.2019, который отмечен протестными акциями, связанными с выборами в Московскую городскую думу (Россия). Второй – период с 09.08.2020 по 08.11.2020, когда прошли протестные акции, связанные с выборами президента в Беларуси.

Для получения статистически достоверных сведений о количестве участников протестных акций изучено 88 источников информации (64 – Минск, 24 – Москва), включающих как официальные данные министерств внутренних дел России и Беларуси, так и данные, приведённые в различных СМИ как своими корреспондентами, так и общественными организациями. В качестве приближенной оценки количества участников акций на конкретную дату использовалось математическое ожидание, вычисленное по всем источникам, доступным на эту дату.

Затем с помощью поиска в социальных сетях «ВКонтакте», «Instagram», «Telegram» и «Twitter» выделены 49 аккаунтов, которые характеризовались относительно большим количеством просмотров публикаций, связанных с протестными акциями. Статистические данные именно по этим аккаунтам были выгружены и в дальнейшем использовались в изучении факторов, влияющих на количество «активных сторонников» политических сообществ. Для выявления факторов, влияющих на количество «активных сторонников» политических сообществ, выгружалась статистика по количеству подписчиков, посещений, комментариев, репостов, публикаций и отметок «нравится».

Сначала рассмотрим результаты изучения политических акций в Москве (Россия) в 2019 году. Построены различные регрессии количества участников политических акций в зависимости от переменных: относительное количество отметок «нравится», репостов, комментариев и посещений. При этом указанные переменные выбирались как по отдельным аккаунтам, так и строились обобщённые (суммарные) показатели по всем или части аккаунтов. Анализировались данные для социальных групп за три, два, один день до акции и в день проведения акции. Ни одна из построенных регрессий не оказалась значимой. Это относится и к различным типам парных регрессий, и к линейным множественным регрессиям.

Затем введена фиктивная переменная, имеющая только два значения: «0» – акция является неразрешённой, а «1» – разрешённой. С фиктивной переменной получены две значимые регрессии: для независимой переменной «сумма просмотров за сутки до акции» десяти рассмотренных групп с коэффициентом детерминации 0,8 и для независимой переменной «сумма количеств отметок «нравится» за сутки до акции» всех десяти групп с коэффициентом детерминации 0,73. Заметим, что для политических групп на КВП количество просмотров и отметок «нравится» в 85 % случаев связано значимой линейной регрессией.

Далее рассмотрим результаты изучения политических акций в Минске (Беларусь) в 2020 году. Данные выгружены по 13 сообществам из «Instagram», 7 – из «Telegram», 2 – из «Twitter» и 1 – из «ВКонтакте». Как и в предыдущем случае, построены уравнения парных и множественных регрессий со всеми выгруженными независимыми переменными и, кроме того, вновь добавлена фиктивная переменная «день недели проведения акции» со значениями «0» и «1» («1» – воскресенье, «0» – не воскресенье). Статистически значимыми оказались две линейные регрессии: парная с зависимой переменной «суммарное количество просмотров в «Telegram» и «ВКонтакте»» с коэффициентом детерминации 0,71 и множественная с двумя зависимыми переменными, характеризующими количество просмотров в Telegram-каналах, с коэффициентом детерминации 0,85.

Резюмируя изложенное, можно констатировать, что на количество «активных сторонников» маркетинговых и политических сообществ чаще других оказывает влияние количество просмотров соответствующего контента. К факторам, влияющим на увеличение количества «активных сторонников», следует отнести количество отметок «нравится» и количество подписчиков. Следует также отметить, что «количество просмотров» является интегральным отображением многих факторов, влияние которых сказывается опосредованно через сложные механизмы.

Заключение

При изучении факторов, влияющих на количество «активных сторонников» сообществ социальных сетей, получен важный результат, заключающийся в построении статистически

значимых регрессий, позволяющих планировать, прогнозировать и моделировать активность маркетинговых и политических сообществ. При этом, несмотря на выявленные различия в статистических закономерностях в различных сообществах (блогерские, политические и маркетинговые), общих особенностей распространения информации в них посредством социальных сетей больше, чем различий. Однако для более адекватного моделирования выявленные различия также необходимо учитывать. В частности, при моделировании распространения информации в маркетинговых сообществах временной единичный период должен быть равен одной неделе, а для блогерских и политических – одному дню.

Выявление регрессионных связей между статистическими показателями сообществ в социальных сетях, а также нормальность целого ряда распределений дают возможность эффективного применения имитационных моделей распространения информации в социальных сетях на основе «проигрывания» различного сочетания двух параметров – математического ожидания и дисперсии.

Кроме того, перспективно сочетание статистического анализа структуры сообществ социальных сетей с имитационным моделированием динамики переходов между составляющими их контингентами для учёта латентных (ненаблюдаемых) состояний в реальных социальных группах и в соответствующих моделях. Целью построения таких моделей выступает обнаружение устойчивых статистических зависимостей для длинных и коротких временных периодов, которые позволят определять направления развития сообществ в социальных сетях.

Библиографический список

1. Вопросы медиалогии: монография / В. В. Смеюха; науч. ред. Н. И. Бусленко. Ростов н/Д: ФГБОУ ВПО РГУПС, 2013. 324 с.

2. **Никонов Ю. В., Чураков В. С.** Моделирование временных социальных сетей онлайн (обзор иностранных исследований) // Альманах современной науки и образования. 2012. № 9. С. 150-156.

3. Q1 2019 Earnings Report. https://s22.q4cdn.com/826641620/files/doc_financials/2019/q1/Q1-2019-Slide-Presentation.pdf. Режим доступа 26.12.2020.

4. 400 миллионов пользователей, 20 000 стикеров, улучшенные опросы и €400K для авторов образовательных тестов. https://telegram.org/blog/400-million/ru. Режим доступа 26.12.2020.

5. Выручка ВКонтакте продолжает расти, несмотря на мировой кризис из-за пандемии. https://vk.com/press/q2-2020-results. Режим доступа 26.12.2020.

6. Welcome to IGTV, our New Video App. https://about.instagram.com/blog/announcements/welcome-to-igtv. Режим доступа 26.12.2020.

7. **Батура Т. В.** Методы анализа компьютерных социальных сетей // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2012. Том 10. № 4. С. 13-28.

8. **Li M., Wang X., Gao K., Zhang S.** A survey on information diffusion in online social net-works: Models and methods // Information (Switzerland). 2017. № 8. С. 118.

9. **Чураков А. Н.** Анализ социальных сетей // СоцИс. 2001. № 1. С. 109-121.

10. **Попова Е. П., Леоненко В. Н.** Прогнозирование реакции пользователей в социальных сетях методами машинного обучения // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 20. № 1. С. 118-124.

11. **Kuptsov M. I., Minaev V. A., Yablochnikov S. L., Dzobelova V. B., Sharopatova A. V.** Some statistical features of the information exchange in social networks // Conference Proceedings: 2020 Systems of Signal Synchronization, Generating and Processing in Telecommunications. IEEE, 2020. С. 1-4.

12. **Bagrow J., Wang D., Barabasi A.-L.** Collective Response of Human Populations to Large-Scale Emergencies // LoS. 2011. Т. 6. № 3. С. 1-8.

13. **Barthelemy M.** Spatial Networks // Physics Reports. 2011. № 499. С. 1-101.

14. **Kolar M., Song L., Ahmed A., Xing E.** Estimating Time-Varying Networks // Annals of Applied Statistics. 2010. № 4. С. 94-123.

15. **Barrat A., Barthelemy M., Vespignani A.** Dynamical Processes on Complex Networks. Cambridge University Press, 2008.

16. **Kuhn F., Oshman R.** Dynamic Networks: Models and Algorithms // ACM SIGACT News. 2011. № 42. С. 82-96.

17. **Newman M.** Networks: an introduction. Oxford University Press, 2010.

18. **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. C. 556-559.

19. **Borgwardt K. M., Kriegel H.-P., Wackersreuther P.** Pattern Mining in Frequent Dynamic Subgraphs // Proceedings of the IEEE International Conference on Data Mining. 2006. C. 818-822.

20. **Kumar R., Novak J., Raghavan P., Tomkins A.** Structure and Evolution of Blogspace // Commun. ACM. 2004. T. 47. № 12. C. 35-39.

21. **Vega-Oliveros D.A., Berton L., Vazquez F., Rodrigues F.A.** The impact of social curiosity on information spreading on networks // Proc. 9th IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2017. C. 459-466.

22. **Minaev V., Vaiz E., Kuptsov M., Yablochnikov S., Vidov S.** Modeling of information impacts on elements of onboard system // Conference Proceedings: 2018 Systems of Signals Generating and Processing in the Field of on-Board Communications. IEEE, 2018. C. 1-5.

23. **Rodrigues H. S.** Application of SIR epidemiological model: new trends // International Journal of Applied Mathematics and Informatics. 2016. № 10. C. 92-97.

24. **Yablochnikov S. L., Yablochnikova I. O., Vidov S. V., Kuptsov M. I., Olisaeva A. V.** The aspects of modeling information processes realized in complex telecommunication systems // Conference Proceedings: 2018 Wave Electronics and its Application in Information and Telecommunication Systems. IEEE, 2018. C. 1-5.

25. **Yablochnikov S., Yablochnikova I., Minaev V., Kuptsov M., Vidov S., Shved V.** System-dynamic modeling of information influences and co-operations // CEUR Workshop Proceedings. T. 2104. 2018. C. 222-232.

UDC 004.77, 519.25

FACTORS DETERMINING PROPAGATION OF INFORMATION IN SOCIAL NETWORKS

M. I. Kuptsov, Ph.D (in physics and mathematics), associate professor, RSREU, Ryazan, Russia;
orcid.org/0000-0002-6814-6423, e-mail: kuptsov_michail@mail.ru
V. A. Minaev, Dr. in technical sciences, full professor, Moscow University of the Ministry of Internal Affairs of Russia named after V. Ya. Kikotya, Moscow, Russia, RSREU, Ryazan, Russia;
orcid.org/ 0000-0002-5342-0864, e-mail: m1va@yandex.ru

The problem of studying the statistical characteristics of information propagation in marketing, blogging and political communities in various social networks is considered. ***The aim of the work*** *is to identify and study the factors that affect social behavior of network users. The manifestation of behavior can be, for example, the change in people involvement into online commerce or socio-political events including those of destructive nature. The empirical laws for the distribution of communities various indicators in networks are studied. The relationships between the indicators are analyzed. Based on cluster analysis, a classification of social network communities is proposed, and characteristic features of clusters are noted. Regression models for predicting changes in the number of users in communities are justified.*

Key words: *social network, Internet community, propagation of information, modeling, statistical regularity.*

DOI: 10.21667/1995-4565-2021-76-81-89

References

1. Questions of mediology: monograph / V. V. Smeyukha; scientific. ed. N. I. Buslenko. Rostov n / D: FGBOU VPO RGUPS, 2013, 324 p. (in Russian).

2. **Nikonov Yu. V., Churakov V. S.** Modeling Temporary Social Networks Online (Review of Foreign Studies) // *Almanac of Contemporary Science and Education*. 2012, no. 9, pp. 150-156 (in Russian).

3. Q1 2019 Earnings Report. https://s22.q4cdn.com/826641620/files/doc_financials/2019/q1/Q1-2019-Slide-Presentation.pdf. Access mode 26.12.2020.

4. 400 million users, 20,000 stickers, improved polls and € 400K for educational test authors. https://telegram.org/blog/400-million/ru. Access mode 26.12.2020.

5. VKontakte's revenue continues to grow, despite the global crisis due to the pandemic. https://vk.com/press/q2-2020-results. Access mode 26.12.2020.

6. Welcome to IGTV, our New Video App. https://about.instagram.com/blog/announcements/welcome-to-igtv. Access mode 26.12.2020.

7. **Batura T. V.** Methods of analysis of computer social networks. *Bulletin of the Novosibirsk State University. Series: Information Technolog*y. 2012, vol. 10, no. 4, pp. 13-28 (in Russian).

8. **Li M., Wang X., Gao K., Zhang S.** A survey on information diffusion in online social net-works: Models and methods. *Information (Switzerland)*. 2017, no. 8, p. 118.

9. **Churakov A. N.** Analysis of social networks. *SotsIs*. 2001, no. 1, pp. 109-121(in Russian).

10. **Popova E. P., Leonenko V. N.** Prediction of user reactions in social networks by machine learning methods. *Scientific and technical bulletin of information technologies, mechanics and optics*. 2020, vol. 20, no. 1, pp. 118-124 (in Russian).

11. **Kuptsov M. I., Minaev V. A., Yablochnikov S. L., Dzobelova V. B., Sharopatova A. V.** Some statistical features of the information exchange in social networks. *Conference Proceedings: 2020 Systems of Signal Synchronization, Generating and Processing in Telecommunications*. *IEEE*, 2020, pp. 1-4.

12. **Bagrow J., Wang D., Barabasi A.-L.** Collective Response of Human Populations to Large-Scale Emergencies. *LoS*. 2011, vol. 6, no. 3, pp. 1-8.

13. **Barthelemy M.** Spatial Networks. *Physics Reports*. 2011, no. 499, pp. 1-101.

14. **Kolar M., Song L., Ahmed A., Xing E.** Estimating Time-Varying Networks. *Annals of Applied Statistics*. 2010, no. 4, pp. 94-123.

15. **Barrat A., Barthelemy M., Vespignani A.** Dynamical Processes on Complex Networks. *Cambridge University Press*, 2008.

16. **Kuhn F., Oshman R.** Dynamic Networks: Models and Algorithms. *ACM SIGACT News*. 2011, no. 42, pp. 82-96.

17. **Newman M.** Networks: an introduction. *Oxford University Press*, 2010.

18. **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.

19. **Borgwardt K. M., Kriegel H.-P., Wackersreuther P.** Pattern Mining in Frequent Dynamic Subgraphs. *Proceedings of the IEEE International Conference on Data Mining*. 2006, pp. 818-822.

20. **Kumar R., Novak J., Raghavan P., Tomkins A.** Structure and Evolution of Blogspace. *Commun. ACM*. 2004, vol. 47, no. 12, pp. 35-39.

21. **Vega-Oliveros D. A., Berton L., Vazquez F., Rodrigues F. A.** The impact of social curiosity on information spreading on networks. *Proc. 9th IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 2017, pp. 459-466.

22. **Minaev V., Vaiz E., Kuptsov M., Yablochnikov S., Vidov S.** Modeling of information impacts on elements of onboard system. *Conference Proceedings: 2018 Systems of Signals Generating and Processing in the Field of on-Board Communications. IEEE*, 2018, pp. 1-5.

23. **Rodrigues H. S.** Application of SIR epidemiological model: new trends. *International Journal of Applied Mathematics and Informatics*. 2016, no. 10, pp. 92-97.

24. **Yablochnikov S. L., Yablochnikova I. O., Vidov S. V., Kuptsov M. I., Olisaeva A. V.** The aspects of modeling information processes realized in complex telecommunication systems. *Conference Proceedings: 2018 Wave Electronics and its Application in Information and Telecommunication Systems. IEEE,* 2018, pp. 1-5.

25. **Yablochnikov S., Yablochnikova I., Minaev V., Kuptsov M., Vidov S., Shved V.** System-dynamic modeling of information influences and co-operations. *CEUR Workshop Proceedings*, vol. 2104. 2018. pp. 222-232.