

Математическое моделирование
Mathematical modeling

УДК 519.63

<https://doi.org/10.32362/2500-316X-2024-12-3-78-92>

EDN WBOETG



НАУЧНАЯ СТАТЬЯ

Анализ и прогнозирование динамики настроений пользователей интернет-ресурсов на основе уравнения Фоккера – Планка

Ю.П. Перова[®],
С.А. Лесько,
А.А. Иванов

МИРЭА – Российский технологический университет, Москва, 119454 Россия

[®] Автор для переписки, e-mail: jul-np@yandex.ru

Резюме

Цели. Цель работы – вывод наблюдаемого на практике степенного закона распределения характеристик социо-динамических процессов из стационарного уравнения Фоккера – Планка и проверка возможности применения нестационарного уравнения Фоккера – Планка для описания динамики процессов в социальных системах.

Методы. При проведении исследований были использованы методы моделирования стохастических процессов, методы и модели теории графов, инструменты и технологии объектно-ориентированного программирования для разработки систем сбора данных из массмедиа-источников, методы имитационного моделирования.

Результаты. Наблюдаемое текущее состояние графа сети комментариев может быть описано с помощью вектора, элементами которого являются среднее значение коэффициента посредничества, среднее значение коэффициента кластеризации, доля пользователей в конкретном состоянии. Критическое состояние сети может быть задано базовым вектором. Зависимость от времени расстояния между базовым вектором и текущим вектором состояния образует временной ряд, значения которого можно рассматривать как «блуждающую точку», динамика перемещений которой описывается нестационарным уравнением Фоккера – Планка. Текущее состояние графа комментариев можно определить с помощью методов текстовой аналитики.

Выводы. Наблюдаемый на практике степенной закон зависимости стационарной плотности вероятности распределения новостей по числу комментариев может быть получен из решения стационарного уравнения Фоккера – Планка, а нестационарное уравнение может быть использовано для описания процессов в сложных сетевых структурах. Для описания состояний сети комментариев пользователей новостных массмедиа можно использовать векторное представление. Достижение или реализация желаемых, или нежелательных состояний всей социальной сети могут быть заданы на основе базовых векторов. Решение нестационарного уравнения Фоккера – Планка позволяет получить уравнение для плотности вероятности переходов между состояниями системы в единицу времени, которые хорошо согласуются с наблюдаемыми данными. Анализ полученной модели с использованием характеристик реального временного ряда для изменения графа комментариев читателей официальной страницы в социальной сети «ВКонтакте» информационного агентства «РИА Новости» и структурных параметров графа показывает ее адекватность.

Ключевые слова: социальные сети, моделирование социальных процессов, сетевой граф, характеристики сетевого графа, уравнение Фоккера – Планка, мониторинг, управление, нелинейная динамика, степенной закон распределения

• Поступила: 12.01.2023 • Доработана: 17.11.2023 • Принята к опубликованию: 08.04.2024

Для цитирования: Перова Ю.П., Лесько С.А., Иванов А.А. Анализ и прогнозирование динамики настроений пользователей интернет-ресурсов на основе уравнения Фоккера – Планка. *Russ. Technol. J.* 2024; 12(3):78–92. <https://doi.org/10.32362/2500-316X-2024-12-3-78-92>

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

RESEARCH ARTICLE

Analyzing and forecasting the dynamics of Internet resource user sentiments based on the Fokker–Planck equation

Julia P. Perova[®],
Sergey A. Lesko,
Andrey A. Ivanov

MIREA – Russian Technological University, Moscow, 119454 Russia

[®] Corresponding author, e-mail: jul-np@yandex.ru

Abstract

Objectives. The study aims to theoretically derive the power law observed in practice for the distribution of characteristics of sociodynamic processes from the stationary Fokker–Planck equation and apply the non-stationary Fokker–Planck equation to describe the dynamics of processes in social systems.

Methods. During the research, stochastic modeling methods were used along with methods and models derived from graph theory, as well as tools and technologies of object-oriented programming for the development of systems for collecting data from mass media sources, and simulation modeling approaches.

Results. The current state of the comment network graph can be described using a vector whose elements are the average value of the mediation coefficient, the average value of the clustering coefficient, and the proportion of users in a corresponding state. The critical state of the network can be specified by the base vector. The time dependence of the distance between the base vector and the current state vector forms a time series whose values can be considered as the “wandering point” whose movement dynamics is described by the non-stationary Fokker–Planck equation. The current state of the comment graph can be determined using text analysis methods.

Conclusions. The power law observed in practice for the dependence of the stationary probability density of news distribution by the number of comments can be obtained from solving the stationary Fokker–Planck equation, while the non-stationary equation can be used to describe processes in complex network structures. The vector representation can be used to describe the comment network states of news media users. Achieving or implementing desired or not desired states of the whole social network can be specified on the basis of base vectors. By solving the non-stationary Fokker–Planck equation, an equation is obtained for the probability density of transitions between system states per unit time, which agree well with the observed data. Analysis of the resulting model using the characteristics of the real time series to change the graph of comments of users of the RIA Novosti portal and the structural parameters of the graph demonstrates its adequacy.

Keywords: social networks, modeling of social processes, network graph, network graph characteristics, Fokker–Planck equation, monitoring, management, nonlinear dynamics, power law of distribution

• Submitted: 12.01.2023 • Revised: 17.11.2023 • Accepted: 08.04.2024

For citation: Perova J.P., Lesko S.A., Ivanov A.A. Analyzing and forecasting the dynamics of Internet resource user sentiments based on the Fokker–Planck equation. *Russ. Technol. J.* 2024;12(3):78–92. <https://doi.org/10.32362/2500-316X-2024-12-3-78-92>

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

ВВЕДЕНИЕ

Описание поведения пользователей социальных сетей и информационных ресурсов является одним из важнейших направлений математической социологии. С практической точки зрения создание моделей, описывающих динамику проявления пользовательских мнений и предпочтений, позволяет разрабатывать системы автоматизированного мониторинга общественного настроения и тенденций его изменения. Преимуществом таких систем по сравнению с традиционными методами изучения общественного мнения является их технологичность в реализации.

Следует отметить, что динамику изменения мнений и настроений пользователей сети Интернет можно в значительной степени отнести к стохастическим процессам. Присутствие человеческого фактора (множество людей с различными мнениями, предпочтениями и характером поведения) с одной стороны создает случайность изменений (в силу большого разнообразия поведенческих моделей пользователей), а с другой стороны вносит в динамику изменений элементы целенаправленности.

В [1] рассмотрена модель, описывающая пространственное и временное распространение информации в социальных сетях на основе стохастического дифференциального уравнения в частных производных. В работе была создана и исследована неавтономная диффузионная логистическая модель с граничными условиями Дирихле, которая показала, что на диффузию данных в социальных сетях сильно влияют коэффициент диффузии и внутренняя скорость роста (распространение информации или слухи можно рассматривать как своего рода вирусы, не обладающие физической формой).

В связи с этим наиболее перспективными для создания моделей динамики изменения общественного настроения на наш взгляд являются модели, которые можно создать на основе стохастических дифференциальных уравнений, например, уравнения Фоккера – Планка, которое учитывает как упорядоченные («снос»), так и случайные изменения («диффузия»). Уравнение Фоккера – Планка широко применяется для анализа и моделирования поведения временных рядов при описании процессов в сложных системах [2–5].

Следует отметить, что помимо уравнения Фоккера – Планка для моделирования на основе дифференциальных уравнений используются и другие подходы, например, уравнения Лиувилля [5, 6], уравнения диффузии [4, 7] и ряда других.

Помимо описания динамических процессов из уравнения Фоккера – Планка можно получить и стационарные решения, которые могут описывать состояние какой-либо системы в стационарном состоянии, когда, например, ее эволюция уже закончилась и изменения не происходят.

Для моделирования социальных процессов применяются не только модели на основе дифференциальных уравнений в частных производных, но, например, модели на основе теоретико-игровых подходов и принятие управленческих решений на их основе [8].

Динамика процессов в сетевых структурах неразрывно связана с их топологией, поэтому необходимо учитывать и их структурные характеристики. Например, в [9] представлена методология анализа тематик в социальной сети, которая включает сбор, обработку и классификацию информации, а также измерение времени между публикациями. Эти данные затем используются для создания шкалы времени.

Таким образом, создается граф, который позволяет отследить изменения популярности конкретных тем, обсуждаемых в социальных сетях. Этот граф также позволяет увидеть связанные события на основе настроений, присутствующих в обществе, а также определить периоды активных дискуссий по конкретным темам. А в [10] на основе графовой модели Кронекера проводилось изучение обнаружения сообществ на графах, обнаружения перекрывающихся сообществ, обнаружения сообществ в неполных сетях с отсутствующими связями и в полных сетях.

Кроме того, приобретает популярность для исследования структуры связей социальных сетей использование природоподобных алгоритмов. В частности, в [11] для оптимизации процесса решения проблем прогнозирования связей и обнаружения сообществ в анализе социальных сетей используются роевые алгоритмы. С увеличением размеров сетей поиск сходства между их узлами становится весьма трудоемким процессом.

Исследование процессов, происходящих в сложных системах с участием человеческого фактора, показывает, что очень часто для наблюдаемых характеристик параметров этих процессов выполняется степенной закон распределения $\rho(x) \sim x^{-\gamma}$ (где γ – характеристическая степень) [12–17], но в то же время вопрос теоретического обоснования возможности ее применения требует дальнейшего изучения. На наш взгляд, это обоснование очень важно. Выявление характера процессов, из которых возникает степенной закон, необходимо для более глубокого изучения поведения и анализа сложных социальных систем.

В связи с этим перспективным является исследование возможности применения уравнения Фоккера – Планка для разработки моделей динамики социальных процессов.

1. СБОР И ОБРАБОТКА ДАННЫХ

Для исследований были выбраны несколько новостных порталов и одно из сетевых сообществ социальной сети «ВКонтакте», посвященное обсуждению новостей информационного ресурса «РИА Новости»¹. Этот ресурс был выбран, исходя из его узнаваемости и популярности в российском обществе: он занимает 1 место среди медиаресурсов (за март 2022 г.) по версии «Brand-Analytics»², входит в топ-3 самых цитируемых информационных агентств в СМИ и социальных медиа (за март 2022 г.) и занимает 1 место³ по этим показателям в российском сегменте интернета.

Сначала с помощью специального программного приложения (парсера) был загружен интересующий диапазон новостей, начиная с 1 января 2019 г. по апрель 2022 г., с официальной страницы «РИА Новости» в социальной сети «ВКонтакте», используя разработанный парсер и программный интерфейс сети⁴. Внутри социальной сети у каждого поста есть свой уникальный адрес, где {owner_id} – уникальный идентификатор сообщества (в случае с «РИА новости» – это «-15755094»), а {post_id} – уникальный идентификатор поста (новости). У каждого поста (новости) присутствует ряд основных параметров, таких как: уникальный идентификатор поста в социальной сети; текст поста; дата и время публикации; количество просмотров

и комментариев пользователей. Комментарии имеют следующие параметры: уникальный идентификатор в сообществе социальной сети; уникальный идентификатор пользователя; текст комментария; дата и время появления; уровень иерархии комментария; связь по уровню комментирования с родительским комментарием (кто из пользователей комментировал кого из других пользователей при обсуждении новости).

Поскольку комментарии могли оставлять чат-боты, спамеры и недобросовестные пользователи, которые пишут комментарии на профессиональной основе, необходимо было ввести правила отчистки данных. К недобросовестным были отнесены те, кто написали за год более 7365 комментариев (в среднем более 20 за сутки) или писали с частотой более одного комментария в 5 мин.

При анализе полученных данных необходимо было определить, какому закону распределения подчиняется наблюдаемая плотность распределения. Для этого были рассмотрены три наиболее часто наблюдаемых закона распределения: Гаусса

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}; \text{ экспоненциальное } \rho(x) = ae^{-ax}$$

и степенное $\rho(x) \sim \beta x^{-\gamma}$. При обработке собранных данных с помощью линейаризации в соответствующих координатах было обнаружено, что наиболее близкая линейаризация наблюдается для степенного закона распределения (рис. 1). Для остальных законов линейаризация была хуже.

Прямая, проведенная на рис. 1, показывает, что линия тренда хорошо описывается выбранной нами линейной аппроксимацией $y = -0.76 - 1.48z$, где $y = \ln\{\rho(x)\}$, $z = \ln\{x\}$, $\ln\{\beta\} = -0.76$, а коэффициент корреляции равен 0.95.

Для подтверждения вывода о линейной аппроксимации можно исследовать поведение остатков и проверить гипотезу о том, что они нормально распределены со средним значением, равным нулю, и имеют однородную дисперсию. Вычисление остатков можно провести на основе реально наблюдаемых значений натурального логарифма доли комментаторов, давших данное число комментариев, и полученного нами уравнения. Рассчитанная величина математического ожидания для распределения остатков равна 0.25, а значение дисперсии равно 0.13. Проверка гипотезы о наклоне (двухвыборочный F-тест для дисперсий) показывает, что дисперсия остатков (рассчитанная относительно линии тренда) существенно меньше, чем дисперсия отклонения точек линейной регрессии от среднего значения величины наблюдаемых данных ($\sum y_i/n = \sum \ln\{\rho(x_i)\}/n$). Она равна 2.11 ($0.13 \ll 2.11$). Таким образом, из полученных данных можно сделать вывод, что распределение остатков очень близко к нормальному, и полученная регрессия является

¹ <https://vk.com/ria> (in Russ.). Дата обращения 02.09.2023. / Accessed September 02, 2023.

² <https://br-analytics.ru/mediatrends/media/?period=202203> (in Russ.). Дата обращения 02.09.2023. / Accessed September 02, 2023.

³ <https://www.mlg.ru/ratings/media/federal/11110/#internet> (in Russ.). Дата обращения 02.09.2023. / Accessed September 02, 2023.

⁴ <https://dev.vk.com/guide> (in Russ.). Дата обращения 02.09.2023. / Accessed September 02, 2023.

значимой, что подтверждает вывод о том, что натуральный логарифм доли комментаторов, написавших данные комментарии, линейно зависит от натурального логарифма числа комментариев, что подтверждает выполнение степенного закона.

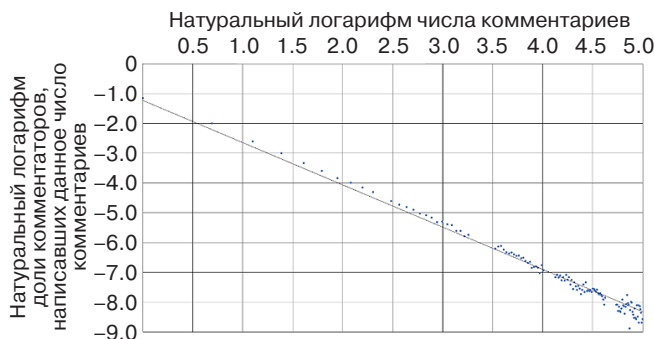


Рис. 1. Линейаризация наблюдаемых данных для степенного распределения доли комментаторов от числа сделанных ими комментариев

При проведении исследований также представляет интерес анализ динамики изменения числа комментариев к новостям, имеющим большой общественный интерес (за время просмотра такие новости набирают сотни комментариев), с течением времени.

Наблюдение этой динамики изменения показывает, что она может иметь как S-образный, так и двухступенчатый характер. В качестве демонстрации такой динамики комментирования пользователями «ВКонтакте» новостей ресурса «РИА Новости» выберем несколько новостей:

1. «Зеленский покинул Украину и переехал в Польшу, – заявил Володин» (см. рис. 2а, на котором динамика комментирования имеет S-образный характер)^{5, 6}, дата и время появления: 2022-03-04 16:13:27 UTC +03:00). Общее число комментариев составило 894. Число комментариев первого уровня (комментарий самой новости) составило 433, второго (комментарии комментариев первого уровня) – 461. Общее количество просмотров – 118764. Среднее время появления комментариев первого уровня – 73 минуты, второго уровня – 74 минуты.
2. «Скоро город будет освобожден, – заявил он» (см. рис. 2б, на котором динамика комментирования имеет двухступенчатый характер)^{7, 8}, дата и время появления: 2022-04-10 17:14:40 UTC +03:00).

⁵ https://vk.com/ria?w=wall-15755094_34243579 (in Russ.). Дата обращения 02.09.2023. / Accessed September 02, 2023.

⁶ <https://ria.ru/20220304/zelenskiy-1776545154.html> (in Russ.). Дата обращения 02.09.2023. / Accessed September 02, 2023.

⁷ https://vk.com/ria?w=wall-15755094_35202266 (in Russ.). Дата обращения 02.09.2023. / Accessed September 02, 2023.

⁸ <https://ria.ru/20220410/ukraina-1782778315.html> (in Russ.). Дата обращения 02.09.2023. / Accessed September 02, 2023.

Общее число комментариев составило 901. Число комментариев первого уровня составило 173, второго – 728. Общее количество просмотров – 173607. Среднее время появления комментариев первого уровня – 75 минуты, а второго уровня – 82 минуты.

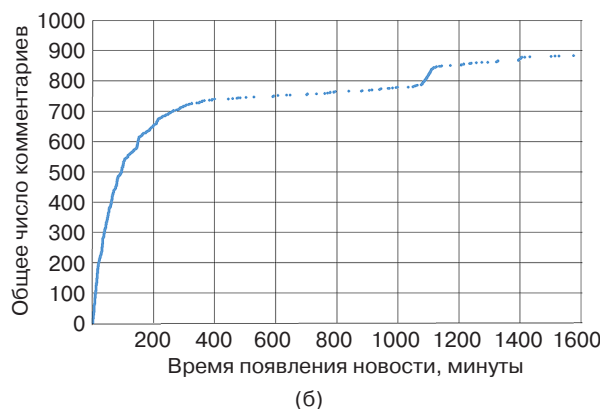
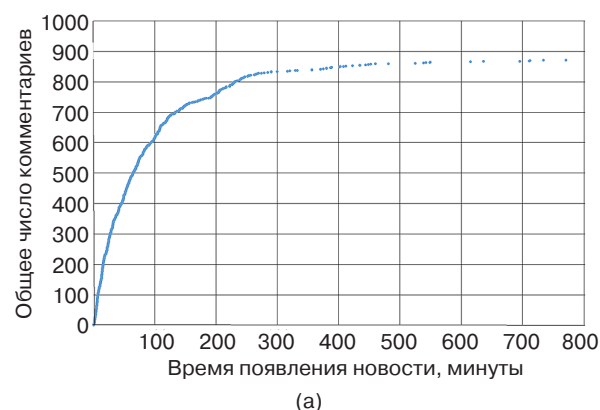


Рис. 2. Наблюдаемая динамика изменения числа комментариев к новостям

На наш взгляд, это может быть связано как с различием в среднем времени появления комментариев второго уровня (интервал времени между появлением комментария первого уровня и комментарием к данным комментариям), так и с соотношением между числом комментариев первого и второго уровней. Если для первой новости интервалы среднего времени появления комментариев первого и второго уровней практически совпадают, то для второй наблюдается небольшое увеличение интервала среднего времени появления комментариев второго уровня (происходит запаздывание по времени). Кроме того, для второй новости их число существенно превосходит число комментариев первого уровня.

Можно сформулировать следующую задачу дальнейшего теоретического исследования: какая природа процессов комментирования новостей и блогов, и какие особенности этих сложных социальных систем приводят к тому, что зависимость плотности вероятности распределения комментариев

от их числа описывается с помощью степенного закона, а динамика имеет во многих случаях сложный двухступенчатый характер?

2. РЕШЕНИЕ СТАЦИОНАРНОГО УРАВНЕНИЯ ФОККЕРА – ПЛАНКА

В общем виде уравнение Фоккера – Планка имеет вид:

$$\frac{\partial \rho(x, t)}{\partial t} = -\frac{\partial}{\partial x}[\mu(x)\rho(x, t)] + \frac{1}{2} \cdot \frac{\partial^2}{\partial x^2}[D(x)\rho(x, t)], \quad (1)$$

где $\rho(x, t)$ – зависящая от времени t плотность вероятности распределения по состояниям x (в нашем случае состояние x – это число комментариев, наблюдаемое в момент времени t); $D(x)$ – зависящий от состояния x коэффициент, определяющий случайное изменение состояния x («диффузия»); $\mu(x)$ – зависящий от состояния x коэффициент, определяющий целенаправленное изменение состояния x («снос»). Применительно к нашей модели $D(x)$ можно трактовать как действия пользователя, вызванные спонтанным импульсом, возникшим при прочтении новости или комментариев к ней других пользователей, когда описываемое в новости или блоге событие не является существенно важным, но пользователь готов потратить время на комментарий или ответить другому комментатору (у пользователя возникло спонтанное желание отреагировать на данную новость). Коэффициент $\mu(x)$ можно интерпретировать как целенаправленные действия, вызванные желанием отреагировать на существенно важные для пользователя новость или блог, а также дать комментарий на комментарий другого пользователя, если он затронул важную с точки зрения данного пользователя тему (пользователь постоянно интересуется данной темой).

Для построения модели необходимо сделать предположения о зависимости $D(x)$ и $\mu(x)$ от состояния x и рассмотреть два условия. Во-первых, учтем размерность членов, входящих в уравнение (1), а во-вторых, можно сделать предположение, что с ростом состояния x (ростом числа возможных комментариев, значимости новости или блога) величины $D(x)$ и $\mu(x)$ также должны увеличиваться. Физические соображения подсказывают, что все члены уравнения (1) должны иметь одинаковую размерность, которую имеет $\rho(x, t)$. И первое, и второе условие будет выполнено, если зависимости $D(x)$ и $\mu(x)$ от состояния x будут иметь вид: $\mu(x) = \mu_0 x$ и $D(x) = D_0 x^2$. Коэффициенты μ_0 и D_0 имеют размерность $1/t$.

Решение стационарного уравнения Фоккера – Планка:

$$-\frac{d}{dx}[\mu(x)\rho(x)] + \frac{1}{2} \cdot \frac{d^2}{dx^2}[D(x)\rho(x)] = 0$$

при следующих допущениях: $\mu(x) = \mu_0 x$ и $D(x) = D_0 x^2$ имеет вид $\rho(x) = [\gamma - 1]x^{-\gamma}$, что соответствует наблюдаемому на практике степенному закону распределения. По полученным из анализа наблюдаемых данных результатам $\gamma = 1.48$, а $\gamma - 1 = 0.48$, натуральный логарифм $(\gamma - 1)$ равен -0.73 , что с достаточно высокой точностью равно $\ln\{\beta\} = -0.76$ (см. полученное уравнение линеаризации: $y = -0.76 - 1.48z$, где $y = \ln\{\rho(x)\}$, $z = \ln\{x\}$, $\ln\{\beta\} = \ln\{\gamma - 1\} - 0.76$, а коэффициент корреляции равен 0.95). В целом это указывает на адекватность разработанной модели. Заметим, что члены стационарного уравнения для $\rho(x)$ также должны иметь одинаковую размерность. В решение стационарного уравнения входит выражение $2\left[1 - \frac{\mu_0}{D_0}\right] = \gamma$, а оно от t не зависит.

Вывод степенного закона может быть описан следующим образом. Подставляем в стационарное уравнение Фоккера – Планка выражения для $\mu(x)$ и $D(x)$.

Подстановка $D(x)$ и $\mu(x)$ в уравнение (1) дает:

$$-\mu_0 x \frac{d\rho(x)}{dx} - \mu_0 \rho(x) + \frac{1}{2} D_0 x^2 \frac{d^2 \rho(x)}{dx^2} + 2D_0 x \frac{d\rho(x)}{dx} + D_0 x \rho(x) = 0.$$

Обозначим $2\left[1 - \frac{\mu_0}{D_0}\right] = \gamma$, тогда

$$x^2 \frac{d^2 \rho(x)}{dx^2} + [2 + \gamma]x \frac{d\rho(x)}{dx} + \gamma \rho(x) = 0.$$

Решение полученного уравнения можно искать в виде: $\rho(x) = \sum_k C_k x^{q_k}$, где C_k – постоянные коэффициенты при соответствующих корнях характеристического уравнения, которое имеет вид: $q(q-1) + [2 + \gamma]q + \gamma = 0$. Данное уравнение имеет два корня: $q_1 = -1$ и $q_2 = -\gamma$. Таким образом для $\rho(x)$ получаем: $\rho(x) = C_1 x^{-1} + C_2 x^{-\gamma}$. Постоянные коэффициенты C_1 и C_2 находятся с использованием условия нормировки функции $\rho(x)$:

$$\int_1^{\infty} \rho(x) dx = C_1 \ln x \Big|_1^{\infty} + C_2 \frac{x^{1-\gamma}}{1-\gamma} \Big|_1^{\infty} \equiv 1.$$

Интеграл вычисляется от 1 до ∞ , поскольку могут быть пользователи, которые дали очень большое число комментариев к новостям, но не может быть комментаторов, написавших менее одного комментария. Учитывая, что при $x \rightarrow \infty \ln x \Big|_{\infty} = \infty$, $C_1 = 0$ и соответственно $C_2 = \gamma - 1$, окончательно получаем: $\rho(x) = [\gamma - 1]x^{-\gamma}$.

3. РЕШЕНИЕ НЕСТАЦИОНАРНОГО УРАВНЕНИЯ ФОККЕРА – ПЛАНКА И АНАЛИЗ МОДЕЛИ

Приведем краткий вывод решения нестационарного уравнения Фоккера – Планка. Используя метод преобразований Лапласа для уравнения Фоккера – Планка, можно записать:

$$\begin{aligned} s\overline{G(s, x)} - \rho(x, 0) = \\ = -\frac{d}{dx} \left[\mu(x) \overline{G(s, x)} \right] + \frac{1}{2} \cdot \frac{d^2}{dx^2} \left[D(x) \overline{G(s, x)} \right]. \end{aligned} \quad (2)$$

Подставив в уравнение (2) соответствующие производные и зависимости $\mu(x)$ и $D(x)$, получим:

$$\begin{aligned} x^2 \frac{d^2 \overline{G(s, x)}}{dx^2} + 2 \left[2 - \frac{\mu_0}{D_0} \right] x \frac{d \overline{G(s, x)}}{dx} + \\ + 2 \left[1 - \frac{\mu_0 + s}{D_0} \right] \overline{G(s, x)} = -\frac{\delta(x - x_0)}{D_0}. \end{aligned} \quad (3)$$

Решение уравнения (3) будем искать в виде: $\overline{G(s, x)} = \sum_k C_k x^q$, где C_k – коэффициенты для корней характеристического уравнения, которое имеет вид:

$$q(q-1) + 2 \left[2 - \frac{\mu_0}{D_0} \right] q + 2 \left[1 - \frac{\mu_0 + s}{D_0} \right] = 0.$$

Решение нестационарного уравнения Фоккера – Планка (3) при сделанных $\mu(x)$ и $D(x)$ допущениях имеет вид:

$$\begin{aligned} \rho(x, t) = \int \frac{\left[\frac{[\ln x]^2}{D_0 t} + \left[\frac{1}{2} - \frac{\mu_0}{D_0} \right] \ln x - 1 \right]}{\sqrt{2\pi D_0 t^3}} \times \\ \times e^{-\left[\frac{[\ln x]^2}{2D_0 t} + \left[\frac{3}{2} \frac{\mu_0}{D_0} \right] \ln x + \left[\frac{1}{2} \frac{\mu_0}{D_0} \right]^2 \frac{D_0 t}{2} \right]} dt. \end{aligned} \quad (4)$$

Вероятность того, что число комментариев к моменту времени t достигнет некоторого числа L , можно найти по формуле: $P(L, t) = 1 - \int_0^L \rho(x, t) dx$. Зависимость числа комментариев от времени t будет описываться уравнением: $N(t) = P(L, t)L$.

Для анализа полученного решения проведем имитационное моделирование. В качестве примера выберем $L = 100$ и три набора значений μ_0 и D_0 : $\mu_0 = 0.45$ и $D_0 = 0.50$ условных единиц ($\mu_0 < D_0$ – кривая 1 на рис. 3а), $\mu_0 = 0.50$ и $D_0 = 0.50$ условных единиц ($\mu_0 = D_0$ – кривая 2 на рис. 3)

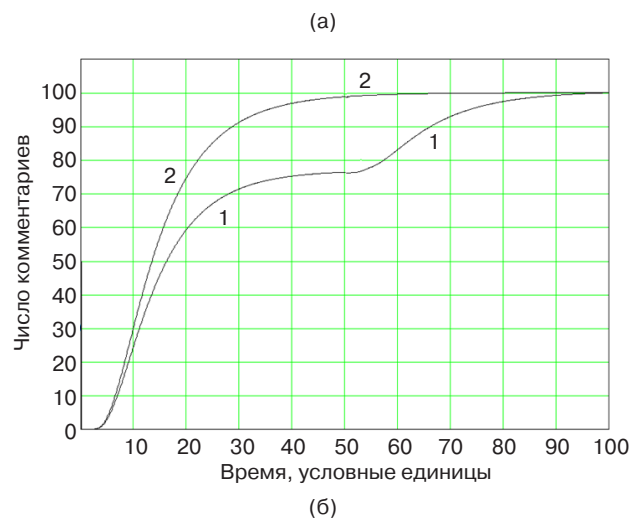
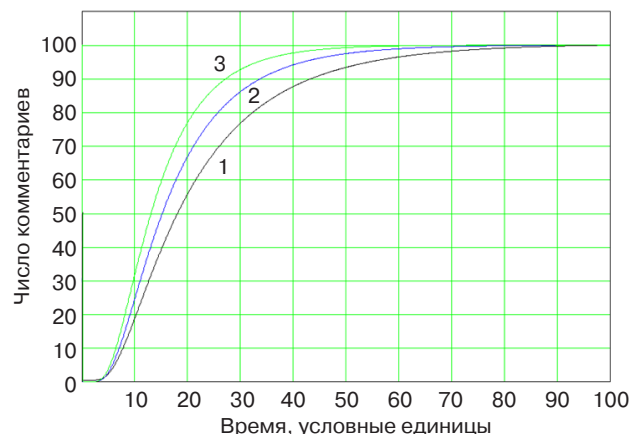


Рис. 3. Динамика изменения с течением времени числа комментариев к новости в имитационной модели на основе уравнения Фоккера – Планка:

(а) 1 процесс, (б) 2 параллельных процесса

и $\mu_0 = 0.55$ и $D_0 = 0.50$ условных единиц ($\mu_0 > D_0$ – кривая 3 на рис. 3а). Расчеты показывают, что с ростом μ_0 относительно D_0 скорость роста кривых для числа комментариев $N(t)$ при выбранных значениях параметров модели μ_0 , D_0 и L увеличивается (рис. 3а).

Двухступенчатую кривую можно получить, если использовать функцию плотности распределения, учитывающую время задержки τ :

$$\begin{aligned} \rho(x, t - \tau) = \int \frac{\left[\frac{[\ln x]^2}{D_0(t - \tau)} + \left[\frac{1}{2} - \frac{\mu_0}{D_0} \right] \ln x - 1 \right]}{\sqrt{2\pi D_0(t - \tau)^3}} \times \\ \times e^{-\left[\frac{[\ln x]^2}{2D_0(t - \tau)} + \left[\frac{3}{2} \frac{\mu_0}{D_0} \right] \ln x + \left[\frac{1}{2} \frac{\mu_0}{D_0} \right]^2 \frac{D_0(t - \tau)}{2} \right]} dt. \end{aligned} \quad (5)$$

Подтверждение соответствия теоретической модели и наблюдаемых данных (рис. 2 и 3б) можно получить, если предположить, что могут одновременно

протекать два процесса с различными коэффициентами μ_0 и D_0 . Причем сумма парциальных долей процессов должна быть равна 1, т.е. $P_{\text{общ}}(L, t) = \alpha_1 P_1(L, t) + \alpha_2 P_2(L, t)$. В качестве примера моделирования выберем для процесса комментирования самой новости или блога следующие параметры модели: $\mu_{0,1} = 0.55$, $D_{0,1} = 0.50$, а для комментирования комментариев: $\mu_{0,2} = 0.50$, $D_{0,2} = 0.50$, $\tau = 50$ условных единиц, $\alpha_1 = 0.75$, $\alpha_2 = 0.25$ и $L = 100$ ($\mu_{0,1} > \mu_{0,2}$ было выбрано, исходя из предположения, что комментирование новости является более первичным процессом для пользователей, чем комментирование комментариев), где $\alpha_1 + \alpha_2 = 1$. Отметим, что это просто пример имитационного моделирования, чтобы показать свойства модели.

Как было указано ранее, плато на кривой 1 на рис. 3б может быть связано с существенным различием в среднем времени появления комментариев второго уровня (интервал времени между появлением комментария первого уровня и комментарием к данному комментарию), что может приводить к реализации двухступенчатой динамики в появлении комментариев.

На рис. 3б представлены результаты моделирования динамики изменения с течением времени числа комментариев $N(t)$ с учетом того, что два процесса могут протекать параллельно. Как видно из полученных результатов, наблюдается хорошее совпадение реальных данных и теоретических расчетов.

4. МОДЕЛИРОВАНИЕ ДИНАМИКИ НАСТРОЕНИЙ ПОЛЬЗОВАТЕЛЕЙ ИНТЕРНЕТ-МАССМЕДИА НА ОСНОВЕ УРАВНЕНИЯ ФОККЕРА – ПЛАНКА И ИЗМЕНЕНИЯ ПАРАМЕТРОВ ГРАФОВ СЕТЕЙ ИХ КОММЕНТАРИЕВ

Для описания состояний сети комментариев предлагается использовать векторное представление. Элементами векторов являются допустимые значения параметров сети (плотность, средняя величина коэффициента посредничества, среднее значение коэффициента кластеризации, эластичность и другие), а также такие характеристики как доля пользователей, которые могут быть отнесены на основе анализа текстов комментариев к одной из групп: 1) лоялист (безусловно поддерживает действия правительства и органов власти); 2) оппозиционер; 3) «тролль» (пользователь, использующий ресурс для того, чтобы устроить скандал); 4) неопределившийся или нейтральный. Достижение или реализация желаемых или нежелательных состояний всей социальной сети в целом могут быть на основе базовых векторов.

Изменение во времени величины расстояния между базовым вектором и вектором текущего состояния, можно рассматривать как «блуждания точки» на отрезке $[L_{\min}, L_{\max}]$ или как случайный (или

почти случайный) временной ряд. Некоторое заданное значение этого расстояния (состояние, при достижении которого необходимо принимать управленческие решения) можно рассматривать как ловушку или точку допустимого порога реализации, куда может с течением времени попасть «блуждающая точка». Это позволяет построить вероятностные социодинамические модели для прогноза динамики общественных настроений.

При традиционном описании поведения «блуждающей точки», как правило, используется диффузионная модель. Однако в данном случае она не может считаться надежной. Как правило, временные ряды, описывающие процессы в сложных системах (например, финансовые показатели фондовых и сырьевых бирж), не являются стационарными, что обусловлено разными причинами, в т.ч. наличием человеческого фактора. Их выборочные функции распределения имеют зависящее от времени математическое ожидание, что противоречит простой диффузионной модели и показывает нестационарность временных рядов.

В этой связи предполагается рассмотреть более сложные модели поведения «блуждающей точки», например, на основе уравнений Фоккера – Планка.

На рис. 4 представлена структура графа, полученного при обработке комментариев к новости. При цветовой визуализации полученных данных узлы графа, в зависимости от их состояния (отнесения к одному из 4 типов), могут быть размечены разными цветами (оппозиционеры – зелено-песочного цвета, «тролли» – красного, лоялисты – синего, неопределившиеся – зеленого цвета). Связи на рисунке показывают взаимное комментирование пользователей друг другом. Таким образом, по цвету узлов можно судить о их состоянии, а по ребрам графа – о взаимодействии.

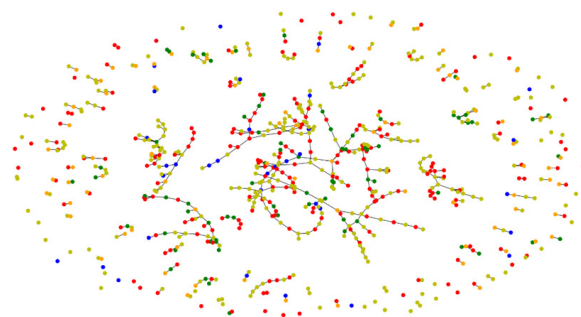


Рис. 4. Структура графа комментариев к рассматриваемой новости

На рис. 4 видно, что в данной структуре имеется много несвязанных одиночных вершин. Однако можно заметить и присутствие связанной компоненты графа, которая отдельно представлена на рис. 5. Замкнутые овалы на этом рисунке (это ядро графа на рис. 4) показывают комментирование пользователями самих себя.

Рассмотрим элементы вектора состояния сети, которые мы будем использовать в нашей модели.

1. Доля узлов, имеющих определенное состояние (например, тех, кто настроен отрицательно по отношению к какому-либо событию общественной жизни).
2. Коэффициент кластеризации – мера плотности связей данной вершины с соседними. Отношение реального количества связей, которые соединяют ближайших соседей данного узла i , к максимально возможному (такому, при котором все ближайшие соседи данного узла были бы соединены непосредственно друг с другом) называется коэффициентом кластеризации узла, его величина лежит на отрезке $[0, 1]$. Чем больше его величина, тем более значимым является данный узел в процессе обмена информацией.
3. Степень посредничества показывает отношение числа кратчайших путей между всеми парами узлов сети, проходящих через данный узел, к общему числу всех кратчайших путей в сети. Ее величина лежит на отрезке $[0, 1]$. Чем больше ее величина, тем значительнее роль данной вершины в обмене информацией.

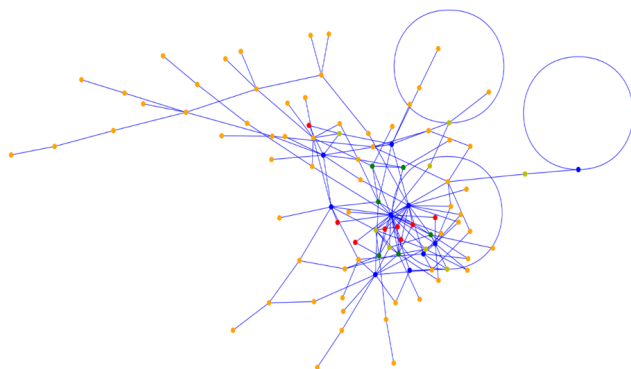


Рис. 5. Граф взаимосвязей пользователей по комментариям

Определим величины элементов базового вектора состояния сети (обозначим его θ). Они задают пороговые значения, переход через которые является с точки зрения управления состояниями нежелательным. Учитывая, что в любом сообществе всегда существует от 0.10 до 0.15 доли участников, всегда несогласных по любому вопросу, примем долю тех, кто настроен отрицательно по отношению к рассматриваемому событию, равной 0.12. Желаемое среднее значение (по всем узлам) коэффициента кластеризации такой сети тоже примем небольшим, например, равным 0.05, и значение средней степени посредничества узлов в такой сети тоже равным 0.05. Таким образом, базовый вектор будет иметь вид: $\theta = (0.12; 0.05; 0.05)$.

Заметим, что число параметров, с помощью которых можно описывать состояние сети, может быть больше. Были выбраны только те, которые, на наш взгляд, являются наиболее значимыми. Кроме того, выбранные параметры являются нормированными (лежат на отрезке $[0, 1]$), поэтому в одинаковой мере влияют на расчет метрики расстояний.

В предлагаемом нами подходе различные графы комментирования новостей на выбранном ресурсе по определенным тематикам в течении суток могут быть объединены в единую структуру через связи между узлами, которые относятся к пользователям. Таким образом можно выделить большой граф, который будет описывать активность пользователей данного сетевого информационного ресурса в течение суток. Далее можно определить элементы вектора текущего состояния, который описывает его характеристики.

Изменения компонентов этого вектора за каждые сутки в течение определенного времени будут образовывать временной ряд.

Для описания изменения значения расстояния между величиной текущего вектора состояния и заданного базового вектора с течением времени рассмотрим решение нестационарного уравнения Фоккера – Планка, что может позволить построить вероятностные социодинамические модели для прогнозирования динамики общественных настроений.

Сформулируем краевую задачу, решение которой будет описывать процесс изменения значения расстояния между величиной текущего вектора состояния графа сети комментариев и заданного базового вектора с течением времени.

Первое граничное условие. При выборе первого граничного условия будем исходить из следующих соображений: состояние $x = L_{\min}$ (левая граница отрезка возможных состояний) определяет состояние, переход через которое необходимо избежать (область, находящаяся левее данного состояния на отрезке, является нежелательной). Вероятность обнаружить такое состояние системы может быть не равна 0, а плотность вероятности, которая определяет поток в состоянии $x = L_{\min}$, нужно взять равной 0, т.к. состояния не должны выходить за эту границу (здесь реализуется условие отражения). Таким образом:

$$\rho(x, t)|_{x=L_{\min}} = 0. \quad (6a)$$

Второе граничное условие. Ограничим область возможных состояний справа некоторой величиной $x = L_{\max}$ (метрика, используемая при расчетах, не может быть больше величины вектора,

элементы которого имеют максимальные значения в пространстве выбранных координат). Вероятность обнаружить такое состояние с течением времени будет отлична от нуля. Однако плотность вероятности, определяющую поток в состоянии $x = L_{\max}$, необходимо положить равной нулю (расстояние между текущим и базовым вектором состояний ограничено максимальными значениями возможных координат в используемом векторном пространстве (реализуется условие отражения от границы)):

$$\rho(x, t)|_{x=L_{\max}} = 0. \quad (66)$$

Для формулировки краевой задачи необходимо задать начальное условие. Так как в момент времени $t = 0$ состояние системы (расстояние между базовым вектором и вектором текущего состояния) может быть равно некоторому значению x_0 , то начальное условие можно задать в виде:

$$\rho(x, t = 0) = \begin{cases} \int \delta(x - x_0) dx = 1, & x = x_0, \\ 0, & x \neq x_0. \end{cases} \quad (6b)$$

Кратко описание решения данной краевой задачи имеет следующий вид. Используя метод преобразований Лапласа для уравнения Фоккера – Планка, можно записать:

$$\begin{aligned} s\overline{G(s, x)} - \rho(x, 0) = \\ = -\frac{d}{dx} \left[\mu(x) s \overline{G(s, x)} \right] + \frac{1}{2} \cdot \frac{d^2}{dx^2} \left[D(x) s \overline{G(s, x)} \right]. \end{aligned} \quad (7)$$

Подставив в уравнение (7) соответствующие производные и зависимости $\mu(x)$ и $D(x)$, получим:

$$\begin{aligned} x^2 \frac{d^2 \overline{G(s, x)}}{dx^2} + 2 \left[2 - \frac{\mu_0}{D_0} \right] x \frac{d \overline{G(s, x)}}{dx} + \\ + 2 \left[1 - \frac{\mu_0 + s}{D_0} \right] \overline{G(s, x)} = -\frac{\delta(x - x_0)}{D_0}. \end{aligned} \quad (8)$$

Решение данного уравнения ищем в виде: $\overline{G(s, x)} = \sum_k C_k x^q$, где C_k – коэффициенты для корней q характеристического уравнения:

$$q(q-1) + 2 \left[2 - \frac{\mu_0}{D_0} \right] q + 2 \left[1 - \frac{\mu_0 + s}{D_0} \right] = 0.$$

Решение будем искать в виде двух функций: $\overline{G_1(s, x)}$ на отрезке $[L, x_0]$ и $\overline{G_2(s, x)}$ на отрезке $[x_0, 1]$ с использованием условия сшивки в точке $x = x_0$ функций $\overline{G_1(s, x)}$ и $\overline{G_2(s, x)}$. Наличие в уравнении δ -функции приводит к тому, что решение, оставаясь непрерывным в точке $x = x_0$, испытывает в ней разрыв производной.

При $L_{\min} \leq x \leq x_0$

$$\overline{G_1(s, x)} = C_1 x^{-[1+\alpha]+\beta\sqrt{k+s}} + C_2 x^{-[1+\alpha]-\beta\sqrt{k+s}}.$$

При $x_0 \leq x \leq L_{\max}$

$$\overline{G_2(s, x)} = C_3 x^{-[1+\alpha]+\beta\sqrt{k+s}} + C_4 x^{-[1+\alpha]-\beta\sqrt{k+s}}.$$

После соответствующих преобразований получим:

- при $L_{\min} \leq x \leq x_0$

$$\begin{aligned} \overline{G_1(s, x)} = \frac{x_0^\alpha x^{-[1+\alpha] \text{sh} \left\{ \left[\beta \ln \left(\frac{L_{\max}}{x_0} \right) \right] \sqrt{k+s} \right\}}}{\beta D_0 \sqrt{k+s}} \times \\ \times \frac{\text{sh} \left\{ \left[\beta \ln \left(\frac{x}{L_{\min}} \right) \right] \sqrt{k+s} \right\}}{\text{sh} \left\{ \left[\beta \ln \left(\frac{L_{\max}}{L_{\min}} \right) \right] \sqrt{k+s} \right\}}; \end{aligned}$$

- при $x_0 \leq x \leq L_{\max}$

$$\begin{aligned} \overline{G_2(s, x)} = \frac{x_0^\alpha x^{-[1+\alpha] \text{sh} \left\{ \left[\beta \ln \left(\frac{x_0}{L_{\min}} \right) \right] \sqrt{k+s} \right\}}}{\beta D_0 \sqrt{k+s}} \times \\ \times \frac{\text{sh} \left\{ \left[\beta \ln \left(\frac{x}{L_{\max}} \right) \right] \sqrt{k+s} \right\}}{\text{sh} \left\{ \left[\beta \ln \left(\frac{L_{\max}}{L_{\min}} \right) \right] \sqrt{k+s} \right\}}. \end{aligned}$$

Сделаем обратные преобразования Лапласа:

$$\overline{G_1(s, x)} = \frac{A(s)}{B(s)} = \sum_{n=1}^M \frac{A(s_n)}{B(s_n)} \cdot e^{s_n t},$$

где s_n – простые, отличные от нуля корни $B(s)$,

$$B(s) = \sqrt{k+s} \cdot \text{sh} \left\{ [\beta \ln L] \sqrt{k+s} \right\},$$

$$A(s) = \text{sh} \left\{ [\beta \ln x_0] \sqrt{k+s} \right\} \cdot \text{sh} \left\{ \left[\beta \ln \frac{x}{L} \right] \sqrt{k+s} \right\}.$$

После соответствующих преобразований получим при $L_{\min} \leq x \leq x_0$:

$$\rho_1(x, t) = 2 \cdot \frac{x_0^\alpha x^{-[1+\alpha]} e^{-kt}}{D_0 \beta^2 \ln \left(\frac{L_{\max}}{L_{\min}} \right)} \cdot \sum_{n=1}^M \frac{\sin \left\{ \pi n \frac{\ln \left(\frac{L_{\max}}{x_0} \right)}{\ln \left(\frac{L_{\max}}{L_{\min}} \right)} \right\} \cdot \sin \left\{ \pi n \frac{\ln \left(\frac{x}{L_{\min}} \right)}{\ln \left(\frac{L_{\max}}{L_{\min}} \right)} \right\}}{\cos \pi n} \cdot e^{-\frac{\pi^2 n^2 t}{\left[\beta \ln \left(\frac{L_{\max}}{L_{\min}} \right) \right]^2}}.$$

Аналогично при $x_0 \leq x \leq L_{\max}$

$$\rho_2(x, t) = -4 \cdot \frac{x_0^\alpha x^{-[1+\alpha]} e^{-kt}}{D_0 \beta^2 \ln L} \cdot \sum_{n=1}^M \frac{\sin \left\{ \pi n \frac{\ln \left(\frac{x_0}{L} \right)}{\ln L} \right\} \cdot \sin \left\{ \pi n \frac{\ln x}{\ln L} \right\}}{\cos \pi n} \cdot e^{-\frac{\pi^2 n^2 t}{[\beta \ln L]^2}}.$$

С учетом $\alpha = \frac{1 - 2 \cdot \frac{\mu_0}{D_0}}{2} = \frac{1}{2} - \frac{\mu_0}{D_0}$, $\beta^2 = \frac{2}{D_0}$ и $k = \frac{\alpha^2}{\beta^2}$ получим:

- при $L_{\min} \leq x \leq x_0$:

$$\rho_1(x, t) = -2 \cdot \frac{x_0^\alpha x^{-[1+\alpha]} \cdot e^{-\frac{D_0 \alpha^2}{2} t}}{\ln \left(\frac{L_{\max}}{L_{\min}} \right)} \cdot \sum_{n=1}^M \frac{\sin \left\{ \pi n \frac{\ln \left(\frac{L_{\max}}{x_0} \right)}{\ln \left(\frac{L_{\max}}{L_{\min}} \right)} \right\} \cdot \sin \left\{ \pi n \frac{\ln \left(\frac{x}{L_{\min}} \right)}{\ln \left(\frac{L_{\max}}{L_{\min}} \right)} \right\}}{\cos \pi n} \cdot e^{-\frac{\pi^2 n^2 D_0 t}{2 \left[\ln \left(\frac{L_{\max}}{L_{\min}} \right) \right]^2}}; \quad (9a)$$

- при $x_0 \leq x \leq L_{\max}$:

$$\rho_2(x, t) = 2 \cdot \frac{x_0^\alpha x^{-[1+\alpha]} \cdot e^{-\frac{D_0 \alpha^2}{2} t}}{\ln \left(\frac{L_{\max}}{L_{\min}} \right)} \cdot \sum_{n=1}^M \frac{\sin \left\{ \pi n \frac{\ln \left(\frac{x_0}{L_{\min}} \right)}{\ln \left(\frac{L_{\max}}{L_{\min}} \right)} \right\} \cdot \sin \left\{ \pi n \frac{\ln \left(\frac{x}{L_{\max}} \right)}{\ln \left(\frac{L_{\max}}{L_{\min}} \right)} \right\}}{\cos \pi n} \cdot e^{-\frac{\pi^2 n^2 D_0 t}{2 \left[\ln \left(\frac{L_{\max}}{L_{\min}} \right) \right]^2}}, \quad \alpha = \frac{1}{2} - \frac{\mu_0}{D_0}. \quad (9b)$$

Вероятность того, что к моменту времени t состояние системы будет находиться на отрезке от $L_{\min}$ до $L_{\max}$, т.е. пороговое состояние (θ) не будет достигнуто, можно вычислить следующим образом:

$$P(\theta, t) = \int_{L_{\min}}^{x_0} \rho_2(x, t) dx + \int_{x_0}^{L_{\max}} \rho_1(x, t) dx. \quad (10)$$

Вероятность $Q(\theta, t)$ того, что пороговое состояние θ к моменту времени t окажется достигнутым или превзойденным, вычисляется по формуле:

$$Q(\theta, t) = 1 - P(\theta, t). \quad (11)$$

Определение границ отрезка возможных состояний от $L_{\min}$ до $L_{\max}$ обсудим позже.

Используя средства сентимент-анализа, все комментарии пользователей в течение суток могут быть размечены на позитивные и негативные, и можно таким образом получить временной ряд. На рис. 6, представлен временной ряд суточной активности пользователей (имеющих негативное отношение) официальной страницы «РИА Новости» в социальной сети «ВКонтакте» по комментированию новостей с января 2013 г. по декабрь 2020 г. Активность определялась как отношение общего числа уникальных комментариев, оставленных пользователями ко всем новостям в течение суток, к общему

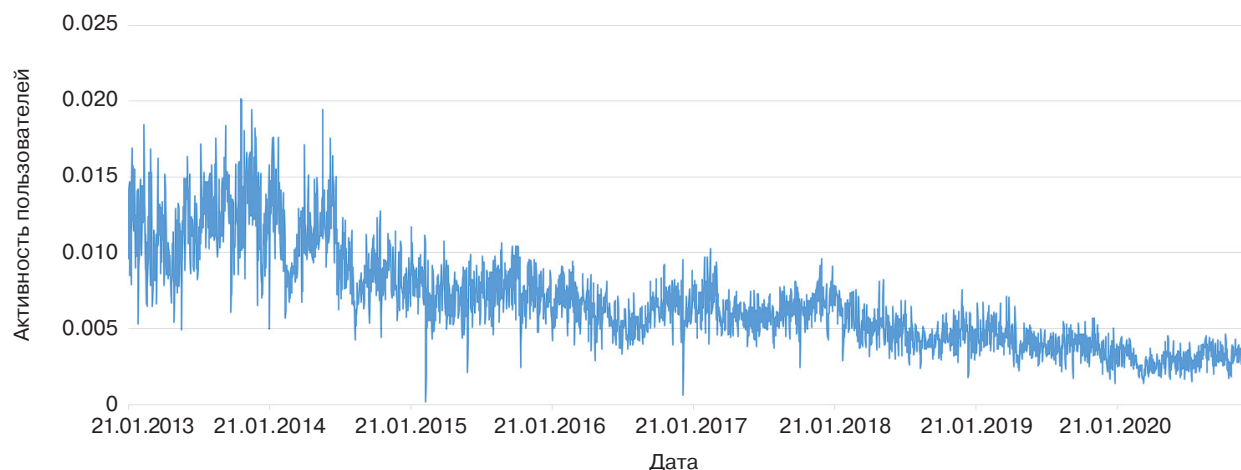


Рис. 6. Временной ряд активности пользователей портала «РИА Новости» по комментированию новостей с января 2013 г. по декабрь 2020 г.

количеству уникальных просмотров пользователями в течение суток всех опубликованных новостей.

Необходимо отметить, что посетители в целом очень неохотно оставляют комментарии. Общее число тех, кто их делает, не превышает 1–2% от всех просмотров портала.

Можно построить объединенный за сутки граф комментариев, определить его характеристики и вектор состояния (он изменяется от суток к суткам).

Обсудим выбор границ отрезка возможных состояний от $L_{\min}$ до $L_{\max}$. Для графа комментариев пользователей портала «РИА Новости», элементы вектора текущего состояния в момент времени 15 мая 2019 г. (который принимаем за $t = 0$) будут равны $X(t) = (0.0035; 0.07; 0.12)$: доля негативных комментариев равна 0.0035; среднее значение коэффициента кластеризации равно 0.07; значение средней степени посредничества узлов равно 0.12.

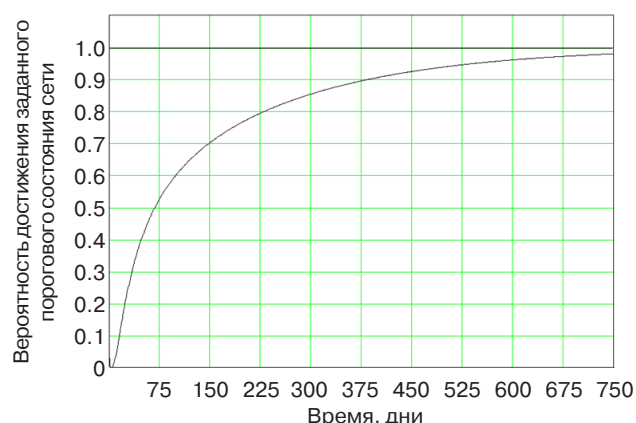


Рис. 7. Зависимость вероятности достижения заданного порогового состояния сети от времени для рассмотренного примера в момент времени 15 мая 2019 г.

Базовый вектор желаемого состояния зададим в виде $\theta = (0.0025; 0.05; 0.05)$: доля негативных комментариев равна 0.0025; среднее значение коэффициента кластеризации равно 0.05; значение средней степени посредничества узлов равно 0.05. Значение доли негативных комментариев, равное 0.0025, выбрано для проверки модели, т.к. известно, когда это значение будет действительно наблюдаться на графике временного ряда. Сравнив результаты моделирования прогноза и реальное значение, можно оценить адекватность модели. Расстояние между заданным базовым вектором желаемого состояния $\theta = (0.0025; 0.05; 0.05)$ и вектором текущего состояния $X(t)$ в момент времени $t = 0$ будет равно $x_0 = 0.073$. Проведя анализ динамики временного ряда изменения состояния сети за несколько предыдущих дней и используя уравнения модели, можно решить обратную задачу и определить величины параметров модели μ_0 и D_0 . В данном случае $\mu_0 = 0.0002$ и $D_0 = 0.0009$.

Правая граница отрезка возможных состояний $L_{\max}$ может быть задана как расстояние между базовым вектором (θ) и вектором максимально возможных значений параметров сети $X(t) = (1; 1; 1)$. В рассматриваемом случае $L_{\max} = 1.70$. Левую границу для надежности рассмотрения можно определить, например, как половину длины заданного базового вектора (в данном случае $|\theta| = 0.071$), таким образом, $L_{\min}$ будет равно 0.035.

Как показывают полученные результаты, если не оказывать на сеть пользователей никакого воздействия, то в сложившихся условиях требуемое состояние может быть достигнуто с вероятностью 0.8 на 225 день и с вероятностью 0.9 на 375 день (рис. 7), что примерно и произошло в действительности, судя по наблюдаемым данным.

5. АЛГОРИТМ ПРОГНОЗИРОВАНИЯ ДОСТИЖЕНИЯ ЗАДАННОГО СОСТОЯНИЯ ГРАФА СЕТИ КОММЕНТАРИЕВ ПОЛЬЗОВАТЕЛЕЙ НОВОСТНЫХ МАССМЕДИА

Проведенные исследования и разработанная модель динамики изменения настроений пользователей интернет-массмедиа на основе уравнения Фоккера – Планка и изменения параметров графов сетей их комментариев позволяют создать алгоритм их прогнозирования, который состоит из следующих шагов:

- а) Сбор с привязкой к дате и времени текстовых комментариев и метаданных пользователей по определенной тематике с новостных сетевых массмедиа-ресурсов.
- б) Обработка данных с использованием средств текстовой аналитики и сентимент-анализ, получение графа комментариев пользователей по определенной тематике и расчет его характеристик (плотность сети, средняя величина коэффициента посредничества, среднее значение коэффициента кластеризации, эластичность, доля пользователей с тем или иным настроением).
- в) Задание значения элементов базового вектора, который будет определять достижение желаемого или нежелательного состояния (θ), и формирование на основе обработанных данных и заданного вектора временного ряда изменения графа комментариев пользователей по определенной тематике с течением времени.
- г) Задание длительности шага τ (час, день, неделя и т.д.). По значениям временного ряда на нескольких шагах для заданного τ определение с помощью численных расчетов с использованием наблюдаемых данных и уравнений (5) и (6) параметров модели μ_0 и D_0 .
- д) Принятие последнего среднего значения метрики расстояний между базовым вектором и вектором текущего состояния сети за начальное состояние x_0 и с использованием полученных значений μ_0 и D_0 , а также уравнений (5) и (6). Проведение расчетов и получение зависимости от времени вероятности достижения желаемого или нежелательного состояния. Далее можно задать величину вероятности (например, 0.95) и оценить время достижения заданного уровня вероятности (сделать прогноз по времени).

ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

Результаты, полученные в работе:

1. Наблюдаемое на практике стационарное распределение новостей по числу комментариев к ним соответствует степенному закону: $\rho(x) = [\gamma - 1]x^{-\gamma}$, где $\rho(x)$ – доля новостей в общем их числе,

имеющая x комментариев, а γ – показатель степени.

2. Динамика изменения с течением времени числа комментариев к новости или блогу может иметь как S-образный вид, так и двухступенчатый, что может быть связано с существенным различием в среднем времени появления комментариев второго уровня (интервал времени между появлением комментария первого уровня и комментарием к данному комментарию), т.е. величиной средней задержки.
3. Наблюдаемый на практике степенной закон зависимости стационарной плотности вероятности распределения новостей по числу комментариев (состояниям x) может быть получен из решения стационарного уравнения Фоккера – Планка, если при его выводе сделать ряд допущений. В частности, предположить, что коэффициент $\mu(x)$, отвечающий в уравнении Фоккера – Планка за целенаправленное изменение состояния системы x (x – текущее число комментариев к новости), линейно зависит от состояния x , а коэффициент $D(x)$, отвечающий за случайное изменение, зависит от x квадратично. Все это позволяет предположить, что уравнение Фоккера – Планка может быть использовано для описания процессов в сложных сетевых структурах.
4. Решение нестационарного уравнения Фоккера – Планка при допущениях о линейной зависимости $\mu(x)$ от состояния x и квадратичной зависимости $D(x)$ от состояния x позволяет получить уравнение для плотности вероятности переходов между состояниями системы в единицу времени, которое хорошо согласуется с наблюдаемыми данными с учетом влияния времени задержки между появлением комментария первого уровня и комментарием к данному комментарию.
5. Разработанные на основе уравнения Фоккера – Планка модели хорошо согласуются с наблюдаемыми данными, что позволяет создать алгоритмы мониторинга и прогнозирования эволюции общественного мнения пользователей новостных информационных ресурсов.
6. Параметры графов сетей комментариев могут быть получены при использовании как готовых инструментов и библиотек языка Python, применяемых для анализа сложных сетей, так и с помощью дополнительно разработанного программного обеспечения.
7. Анализ полученной модели с использованием характеристик реального временного ряда для изменения графа комментариев пользователей портала «РИА Новости» и структурных параметров графа показывает ее адекватность и непротиворечивость. Как показывают полученные

результаты, если не оказывать на сеть пользователей никакого воздействия, то в сложившихся условиях требуемое состояние по числу негативно настроенных комментаторов может быть достигнуто с вероятностью 0.8 на 225 день, и с вероятностью 0.9 на 375 день, что примерно и произошло в действительности, судя по наблюдаемым данным.

8. Созданная на основе уравнения Фоккера – Планка модель позволяет разработать алгоритм прогнозирования достижения заданного состояния графа сети комментариев пользователей новостных массмедиа, что является практически очень важным.

В заключении отметим, что сложный характер динамики процессов в сложных социальных системах можно описывать не только на основе моделей, созданных на основе уравнения Фоккера – Планка. Например, в работах [18–21] представлены разработанные авторами модели описания стохастической динамики изменения состояний в сложных социальных системах, учитывающие процессы самоорганизации и наличие памяти. Для создания модели были рассмотрены графические схемы вероятностей переходов между возможными состояниями описываемых систем с учетом предыдущих

состояний, что позволяет учитывать память и описывать не только марковские, но и немарковские процессы.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке Российского научного фонда, грант № 22-21-00109 «Разработка моделей прогнозирования динамики социальных настроений на основе анализа временных рядов текстового контента социальных сетей с использованием уравнений Фоккера – Планка и нелинейной диффузии».

ACKNOWLEDGMENTS

The study is financially supported by the Russian Science Foundation, grant No. 22-21-00109 “Development of models for forecasting the dynamics of social sentiments based on analyzing time series of text content of social networks using the Fokker–Planck equations and nonlinear diffusion.”

Вклад авторов. Все авторы в равной степени внесли свой вклад в исследовательскую работу.

Authors' contribution. All authors equally contributed to the research work.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Du B., Lian X., Cheng X. Partial differential equation modeling with Dirichlet boundary conditions on social networks. *Bound. Value Probl.* 2018;2018(1):50. <https://doi.org/10.1186/s13661-018-0964-4>
2. Lux T. Inference for systems of stochastic differential equations from discretely sampled data: a numerical maximum likelihood approach. *Ann. Finance.* 2013;9(2):217–248. <http://doi.org/10.1007/s10436-012-0219-9>
3. Hurn A., Jeisman J., Lindsay K. Teaching an Old Dog New Tricks: Improved Estimation of the Parameters of Stochastic Differential Equations by Numerical Solution of the Fokker–Planck Equation. In: Dungey M., Bardsley P. (Eds.). *Proceedings of the Australasian Meeting of the Econometric Society.* 2006. The Australian National University, Australia. P. 1–36.
4. Elliott R.J., Siu T.K., Chan L. A PDE approach for risk measures for derivatives with regime switching. *Ann. Finance.* 2007;4(1):55–74. <http://dx.doi.org/10.1007/s10436-006-0068-5>
5. Орлов Ю.Н., Федоров С.Л. Генерация нестационарных траекторий временного ряда на основе уравнения Фоккера–Планка. *ТРУДЫ МФТИ.* 2016;8(2):126–133. [Orlov Y.N., Fedorov S.L. Nonstationary time series trajectories generation on the basis of the Fokker–Planck equation. *TRUDY MFTI = Proceedings of MIPT.* 2016;8(2):126–133 (in Russ.).]
6. Chen Y., Cosimano T.F., Himonas A.A., Kelly P. An Analytic Approach for Stochastic Differential Utility for Endowment and Production Economies. *Comput. Econ.* 2013;44(4):397–443. <http://doi.org/10.1007/s10614-013-9397-4>
7. Savku E., Weber G.-W. Stochastic differential games for optimal investment problems in a Markov regime-switching jump-diffusion market. *Ann. Oper. Res.* 2020;132(6):1171–1196. <https://doi.org/10.1007/s10479-020-03768-5>
8. Красников К.Е. Математическое моделирование некоторых социальных процессов с помощью теоретико-игровых подходов и принятие на их основе управленческих решений. *Russ. Technol. J.* 2021;9(5):67–83. <https://doi.org/10.32362/2500-316X-2021-9-5-67-83> [Krasnikov K.E. Mathematical modeling of some social processes using game-theoretic approaches and making managerial decisions based on them. *Russ. Technol. J.* 2021;9(5):67–83 (in Russ.). <https://doi.org/10.32362/2500-316X-2021-9-5-67-83>]
9. Kirn S.L., Hinders M.K. Dynamic wavelet fingerprint for differentiation of tweet storm types. *Soc. Netw. Anal. Min.* 2020;10(1):4. <https://doi.org/10.1007/s13278-019-0617-3>
10. Hoffmann T., Peel L., Lambiotte R., Jones N.S. Community detection in networks without observing edges. *Sci. Adv.* 2020;6(4):1478. <https://doi.org/10.1126/sciadv.aav1478>

11. Pulipati S., Somula R., Parvathala B.R. Nature inspired link prediction and community detection algorithms for social networks: a survey. *Int. J. Syst. Assur. Eng. Manag.* 2021. <https://doi.org/10.1007/s13198-021-01125-8>
12. Dorogovtsev S.N., Mendes J.F.F. Evolution of networks. *Adv. Phys.* 2002;51(4):1079–1187. <http://doi.org/10.1080/00018730110112519>
13. Newman M.E.J. The structure and function of complex networks. *SIAM Rev.* 2003;45(2):167–256. <https://doi.org/10.1137/S003614450342480>
14. Dorogovtsev S.N., Mendes J.F.F., Samukhin A.N. Generic scale of the scale-free growing networks. *Phys. Rev. E.* 2001;63(6):062101. <https://doi.org/10.1103/PhysRevE.63.062101>
15. Golder S., Wilkinson D., Huberman B. Rhythms of Social Interaction: Messaging Within a Massive Online Network. In: Steinfield C., Pentland B.T., Ackerman M., Contractor N. (Eds.). *Communities and Technologies*. 2007. https://doi.org/10.1007/978-1-84628-905-7_3
16. Kumar R., Novak J., Tomkins A. Structure and evolution of online social networks. In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and data Mining, KDD '06*. 2006. P. 611–617. <https://doi.org/10.1145/1150402.1150476>
17. Mislove A., Marcon M., Gummadi K.P., Druschel P., Bhattacharjee B. Measurement and analysis of online social networks. In: *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, IMC'07*. 2007. P. 29–42. <https://doi.org/10.1145/1298306.1298311>
18. Zhukov D., Khvatova T., Zaltsman A. Stochastic Dynamics of Influence Expansion in Social Networks and Managing Users' Transitions from One State to Another. In: *Proceedings of the 11th European Conference on Information Systems Management (ECISM 2017)*. 2017. P. 322–329.
19. Zhukov D., Khvatova T., Millar C., Zaltsman A. Modelling the stochastic dynamics of transitions between states in social systems incorporating self-organization and memory. *Technol. Forecast. Soc. Change*. 2020;158:120134. <https://doi.org/10.1016/j.techfore.2020.120134>
20. Zhukov D., Khvatova T., Istratov L. A stochastic dynamics model for shaping stock indexes using self-organization processes, memory and oscillations. In: *Proceedings of the European Conference on the Impact of Artificial Intelligence and Robotics (ECIAIR 2019)*. 2019. P. 390–401.
21. Zhukov D.O., Zaltsman A.D., Khvatova T.Yu. Forecasting Changes in States in Social Networks and Sentiment Security Using the Principles of Percolation Theory and Stochastic Dynamics. In: *Proceedings of the 2019 IEEE International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS)*. 2019. Article number 8928295. P. 149–153. <https://doi.org/10.1109/ITQMIS.2019.8928295>

Об авторах

Перова Юлия Петровна, старший преподаватель, кафедра телекоммуникаций, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: jul-np@yandex.ru. Scopus Author ID 57431908700, <https://orcid.org/0000-0003-4028-2842>

Лесько Сергей Александрович, д.т.н., доцент, профессор кафедры «Информационно-аналитические системы кибербезопасности», Институт кибербезопасности и цифровых технологий, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: sergey@testor.ru. Scopus Author ID 57189664364, <https://orcid.org/0000-0002-6641-1609>

Иванов Андрей Андреевич, магистрант, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: heliosgoodgame@gmail.com. <https://orcid.org/0009-0002-7199-2871>

About the authors

Julia P. Perova, Senior Lecturer, Department of Telecommunications, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: jul-np@yandex.ru. Scopus Author ID 57431908700, <https://orcid.org/0000-0003-4028-2842>

Sergey A. Lesko, Dr. Sci. (Eng.), Docent, Professor of the Cybersecurity Information and Analytical Systems Department, Institute of Cybersecurity and Digital Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: sergey@testor.ru. Scopus Author ID 57189664364, <https://orcid.org/0000-0002-6641-1609>

Andrey A. Ivanov, Student, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: heliosgoodgame@gmail.com. <https://orcid.org/0009-0002-7199-2871>