

# ЦИФРОВЫЕ РЕШЕНИЯ И ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Научно-практический рецензируемый журнал  
Издается с 2025 г.

**Издатель: Финансовый университет  
при Правительстве Российской Федерации,  
Москва, Российская Федерация**

125167, Москва, Ленинградский пр-т, д. 53

Периодичность издания — 4 номера в год

Журнал ориентирован на научное обсуждение актуальных проблем  
по научным специальностям:

- 1.2.1. «Искусственный интеллект и машинное обучение»
- 1.2.2. «Математическое моделирование, численные методы и комплексы программ»
- 2.3.6. «Методы и системы защиты информации, информационная безопасность»
- 5.2.2. «Математические, статистические и инструментальные методы в экономике»

Электронная версия журнала на русском и английском языках  
находится в открытом доступе на сайте [digitari.ru](http://digitari.ru)  
Журнал публикует материалы на условиях лицензии  
Creative Commons Attribution 4.0 International (CC BY 4.0).

## DIGITAL SOLUTIONS AND ARTIFICIAL INTELLIGENCE TECHNOLOGIES

Scientific and practical peer-reviewed journal  
Published since 2025

## ЦИФРОВЫЕ РЕШЕНИЯ И ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Научно-практический журнал

Том 2, № 1, 2026

Издается с 2025 года

Периодичность: 4 раза в год

ISSN 3033-5477 (Print)

ISSN 3033-7097 (Online)

Свидетельство о регистрации:

Эл № ФС77-90338

от 01 ноября 2025 г.

### Учредитель и издатель:

Финансовый университет  
при Правительстве Российской  
Федерации

125167, Российская Федерация,  
Москва,

Ленинградский проспект, 49  
+7 (499) 553-10-74

[https://www.digitarin.ru/jour  
aaizvekova@fa.ru](https://www.digitarin.ru/jour_aaizvekova@fa.ru)

Главный редактор

**Феклин**

**Вадим Геннадьевич**

«Цифровые решения и  
технологии искусственного  
интеллекта» — платформа для  
публикации оригинальных  
и актуальных исследований,  
посвященных современным  
технологиям в сфере  
искусственного интеллекта и  
цифровых решений.

Заведующий редакцией

**В.И. Головатенко**

Выпускающий редактор

**А.А. Извекова**

Переводчик **И.М. Евстратов**

Корректор **С.Ф. Михайлова**

Верстальщик **Е.А. Смирнова**

Подписано в печать:

02.03.2026

Отпечатано в ООО «Коми  
Республиканская типография»,  
г. Сыктывкар, ул. Савина, д. 81.

Заказ № 26-1123

Выход в свет 02.04.2026

© Финансовый университет  
при Правительстве Российской  
Федерации, Москва

## ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ

**Маркова С.В.**

Гибридные ансамблевые методы интеллектуального  
анализа данных: интеграция интерпретируемости  
и производительности в условиях больших данных . . . . . 6

**Денисова М.В., Кочкаров Р.А.**

Агрегация слабосвязанных компонент  
и мостов влияния многослойных социальных графов . . . . 16

## МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ, ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

**Рыженко А.А., Козьминых С.И.**

Недекларируемые возможности файловой архитектуры.  
Уязвимость формата PPTX Microsoft PowerPoint . . . . . 28

**Любимов А.И., Резниченко С.А.**

Обзор инструментов для тестирования  
на проникновение: сравнение функционала  
и удобства использования. . . . . 35

## МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, ЧИСЛЕННЫЕ МЕТОДЫ И КОМПЛЕКСЫ ПРОГРАММ

**Борисова Л.Р.**

Методы интеллектуального анализа данных  
в исследовании экономического развития стран. . . . . 45

**Yatskin D.V., Kochkarov A.A., Okuneva E.A.**

Assessment of the Structural Stability of Transport  
and Logistics Systems Based on Graph Models  
and the Percolation Coefficient . . . . . 52

## МАТЕМАТИЧЕСКИЕ, СТАТИСТИЧЕСКИЕ И ИНСТРУМЕНТАЛЬНЫЕ МЕТОДЫ В ЭКОНОМИКЕ

**Бурыкин Н.А., Гринева Н.В., Голосов П.Е.**

Анализ подходов и решений для разработки системы  
хранения клиентских данных банковской экосистемы . . . 63

**Варьяш И.Ю.**

Исследование реальных фракталов опережающих  
индикаторов. . . . . 73

**ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING****Markova S.V.**

Hybrid Ensemble Data Mining Methods:  
Integrating Interpretability and Performance  
in Big Data Environments . . . . . 6

**Denisova M.V., Kochkarov R.A.**

Aggregation of Weakly Connected Components  
and Influence Bridges in Multilayer Social Graphs . . . . . 16

**METHODS AND SYSTEMS OF INFORMATION  
PROTECTION, INFORMATION SECURITY****Ryzhenko A.A., Kozminykh S.I.**

Undeclared File Architecture Features.  
Vulnerability of Microsoft PowerPoint's PPTX Format . . . . . 28

**Lyubimov A.I., Reznichenko S.A.**

Overview of Penetration Testing Tools:  
Comparison of Functionality and Ease of Use . . . . . 35

**MATHEMATICAL MODELING, NUMERICAL METHODS  
AND SOFTWARE PACKAGES****Borisova L.R.**

Methods of Data Mining in the Study  
of Economic Development of Countries . . . . . 45

**Yatskin D.V., Kochkarov A.A., Okuneva E.A.**

Assessment of the Structural Stability of Transport  
and Logistics Systems Based on Graph Models  
and the Percolation Coefficient . . . . . 52

**MATHEMATICAL, STATISTICAL AND INSTRUMENTAL  
METHODS IN ECONOMICS****Burykin N.A., Grineva N.V., Golosov P.E.**

Analysis of Approaches and Solutions for Developing  
a Customer Data Storage System for the Banking Ecosystem . . 63

**Varjas I.Yu.**

Research of Real Fractals of Leading Indicators . . . . . 73

**DIGITAL SOLUTIONS  
AND ARTIFICIAL  
INTELLIGENCE  
TECHNOLOGIES***Scientific and practical journal*

Vol. 2, No. 1, 2026

Published Since 2025

Publication Frequency: 4  
times a year

ISSN 3033-5477 (Print)

ISSN 3033-7097 (Online)

**Founder and Publisher:**

Financial University  
under the Government of  
the Russian Federation  
125167, Russian Federation,  
Moscow,

49, Leningradsky Avenue  
+7 (499) 553-10-74

<https://www.digitarin.ru/jour-aaizvekova@fa.ru>

Digital Solutions and  
Artificial Intelligence  
Technologies is a platform  
for publishing original and  
relevant research on modern  
technologies in the field of  
artificial intelligence and  
digital solutions.

Editor-in-Chief

**Vadim G. Feklin**

Head of Scientific Journals  
Editorial Department

**V.I. Golovatenko**

Managing Editor

**A.A. Izvekova**

Translator

**I.M. Evstratov**

Proofreader

**S.F. Mikhaylova**

Design, make up

**E.A. Smirnova**

Signed off to printing:

02.03.2026

Printed in Komi Republican  
Printing House LLC, Syktyvkar,  
Savina St. 81

Order № 26-1123

Publication date 02.04.2026

© Financial University  
under the Government  
of the Russian Federation,  
Moscow

## ГЛАВНЫЙ РЕДАКТОР

**ФЕКЛИН В.Г.**, кандидат физико-математических наук, доцент, декан факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

## ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА

**КОЧКАРОВ Р.А.**, кандидат экономических наук, доцент кафедры искусственного интеллекта, заместитель декана по научной работе факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

## ЧЛЕНЫ РЕДАКЦИОННОЙ КОЛЛЕГИИ

**АНДРИЯНОВ Н.А.**, кандидат технических наук, доцент, доцент кафедры искусственного интеллекта, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**АФНАСЬЕВ А.А.**, доктор экономических наук, доцент, ведущий научный сотрудник ЦЭМИ РАН, заведующий кафедрой моделирования и системного анализа, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**БОГДАНОВ Е.А.**, Ph.D, заведующий кафедрой информационной безопасности, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**ВАСИЛЬЕВА Е.В.**, доктор экономических наук, доцент, заведующая кафедрой бизнес-информатики, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**ЖУКОВСКАЯ И.Е.**, доктор экономических наук, доцент, профессор кафедры бизнес-информатики, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**КОРОВИН Д.И.**, доктор экономических наук, доцент, профессор кафедры моделирования и системного анализа, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**КАШИРИН А.Б.**, директор Центра продвинутой аналитики АО «Альфа-Банк», заведующий базовой кафедрой Альфа-Банка, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**КОЗЬМИНЫХ С.И.**, доктор технических наук, доцент, профессор кафедры информационной безопасности, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**КОРОТЕЕВ М.В.**, кандидат экономических наук, доцент, заведующий кафедрой искусственного интеллекта, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**КОЧКАРОВ А.А.**, доктор технических наук, доцент, заместитель директора по инновационной работе ФИЦ Биотехнологии РАН, профессор кафедры искусственного интеллекта, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**МАКАРОВ В.Л.**, доктор физико-математических наук, профессор, академик РАН, научный руководитель ЦЭМИ РАН, научный руководитель факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**МЕЛЬНИКОВ Д.А.**, доктор технических наук, доцент, профессор кафедры информационной безопасности, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**МИХАЙЛОВА С.С.**, доктор экономических наук, заведующая кафедрой математики и анализа данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**НЕИЗВЕСТНЫЙ С.И.**, доктор технических наук, старший научный сотрудник, профессор кафедры бизнес-информатики, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**ОСТАПЕНКО Г.А.**, доктор технических наук, профессор, профессор по цифровизации, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**ПЕТРОСОВ Д.А.**, кандидат технических наук, доцент, заведующий кафедрой информационных технологий, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**ПРОКОПЧИНА С.В.**, доктор технических наук, профессор, профессор кафедры искусственного интеллекта, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**РЯБОВ П.Е.**, доктор физико-математических наук, доцент, профессор кафедры математики и анализа данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**СЛАВИН Б.Б.**, доктор экономических наук, профессор кафедры бизнес-информатики, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**ТИМОШЕНКО А.В.**, доктор технических наук, профессор, профессор кафедры информационной безопасности, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**ЦАРЕГОРОДЦЕВ А.В.**, доктор технических наук, профессор, главный научный сотрудник Института цифровых технологий, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**ЩЕТИНИН Е.Ю.**, доктор физико-математических наук, доцент, профессор кафедры искусственного интеллекта, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**POURMOHAMMADBAGHER L.**, Dr., Allameh Tabataba'i University (ATU), Iran

**BAHRANI M.**, Dr., Allameh Tabataba'i University (ATU), Iran

**PEYMANI M.**, Dr., Allameh Tabataba'i University (ATU), Iran

**SHARMA CH.**, Prof., GNIOT Institute Of Management Studies, Greater Noida, India

**JUNSHENG ZH.**, Prof., Dalian Neusoft University of Information, China

**LIANZHUANG Q.**, PhD, Dalian Neusoft University of Information, China

**KANG L.**, Prof., Dalian Neusoft University of Information, China

Рукописи принимаются [vfeklin@fa.ru](mailto:vfeklin@fa.ru)

Требования к оформлению авторских материалов:

<https://www.digitarin.ru/jour>

Редакция в обязательном порядке осуществляет экспертную оценку (рецензирование), научное, литературное и техническое редактирование всех материалов, публикуемых в журнале.

## EDITOR-IN-CHIEF

**FEKLIN V.G.**, Cand. Sci. (Phys.-Math.), Associate Professor, Dean of the Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

## DEPUTY EDITOR-IN-CHIEF

**KOCHKAROV R.A.**, Cand. Sci. (Econ.), Associate Professor, Department of Artificial Intelligence, Deputy Dean for Research at the Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

## MEMBERS OF THE EDITORIAL BOARD

**ANDRIANOV N.A.**, Cand. Sci. (Tech.), Associate Professor, Associate Professor of Artificial Intelligence, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**AFANASYEV A.A.**, Dr. Sci. (Econ.), Associate Professor, Leading Researcher at CEMI RAS, Head of the Department of Modeling and System Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**BOGDANOV E.A.**, PhD, Head of the Department of Information Security, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**VASILYEVA E.V.**, Dr. Sci. (Econ.), Associate Professor, Head of the Department of Business Informatics, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**ZHUKOVSKAYA I.E.**, Dr. Sci. (Econ.), Associate Professor, Professor, Department of Business Informatics, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**KOROVIN D.I.**, Dr. Sci. (Econ.), Associate Professor, Professor of the Department of Modeling and System Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**KASHIRIN A.B.**, Director of the Center for Advanced Analytics of Alfa-Bank JSC, Head of the Basic Department of Alfa-Bank, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**KOZMINYKH S.I.**, Dr. Sci. (Tech), Associate Professor, Professor, Department of Information Security, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**KOROTEEV M.V.**, Cand. Sci. (Econ.), Associate Professor, Head of the Department of Artificial Intelligence, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**KOCHKAROV A.A.**, Dr. Sci. (Tech), Associate Professor, Deputy Director for Innovation at the Institute of Biotechnology of the Russian Academy of Sciences, Professor, Department of Artificial Intelligence, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**MAKAROV V.L.**, Dr. Sci. (Phys.-Math.), Professor, Academician of the RAS, Scientific Director of the Central Research Institute of the Russian Academy of Sciences, Scientific Director of the Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**MELNIKOV D.A.**, Dr. Sci. (Tech), Associate Professor, Professor, Department of Information Security, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**MIKHAILOVA S.S.**, Dr. Sci. (Econ.), Head of the Department of Mathematics and Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**NEIZVESTNYJ S.I.**, Dr. Sci. (Tech), Senior Researcher, Professor, Department of Business Informatics, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**OSTAPENKO G.A.**, Dr. Sci. (Tech), Professor, Vice-Rector for Digitalization, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**PETROSOV D.A.**, Cand. Sci. (Tech.), Associate Professor, Head of the Department of Information Technology, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**PROKOPCHINA S.V.**, Dr. Sci. (Tech), Professor, Professor, Department of Artificial Intelligence, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**RYABOV P.E.**, Dr. Sci. (Phys.-Math.), Associate Professor, Professor of the Department of Mathematics and Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**SLAVIN B.B.**, Dr. Sci. (Econ.), Professor, Department of Business Informatics, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**TIMOSHENKO A.V.**, Dr. Sci. (Tech), Professor, Professor, Department of Information Security, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**TSAREGORODTSEV A.V.**, Dr. Sci. (Tech.), Professor, Chief Researcher at the Institute of Digital Technologies, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**SHCHETININ E.Y.**, Dr. Sci. (Phys.-Math.), Associate Professor, Professor, Department of Artificial Intelligence, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**POURMOHAMMADBAGHER L.**, Dr., Allameh Tabataba'i University (ATU), Iran

**BAHRANI M.**, Dr., Allameh Tabataba'i University (ATU), Iran

**PEYMANI M.**, Dr., Allameh Tabataba'i University (ATU), Iran

**SHARMA CH.**, Prof., GNIOT Institute Of Management Studies, Greater Noida, India

**JUNSHENG ZH.**, Prof., Dalian Neusoft University of Information, China

**LIANZHANG Q.**, PhD, Dalian Neusoft University of Information, China

**KANG L.**, Prof., Dalian Neusoft University of Information, China

DOI: 10.26794/3030-7097-2026-2-1-6-15  
УДК 004.62+004.81(045)

# Гибридные ансамблевые методы интеллектуального анализа данных: интеграция интерпретируемости и производительности в условиях больших данных

С.В. Маркова

Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

## АННОТАЦИЯ

Данное исследование представляет комплексный анализ гибридных ансамблевых методов, интегрирующих классические алгоритмы машинного обучения с современными технологиями глубокого обучения для решения задач классификации и прогнозирования на больших данных. **Основная цель работы** заключается в разработке и эмпирической валидации методологического подхода, позволяющего достичь оптимального баланса между производительностью модели и объяснимостью ее решений. В ходе исследования применялись **методы** стекинга, бэггинга и бустинга в сочетании с техниками интерпретируемого машинного обучения, включая SHAP-анализ и методы важности признаков. **Результаты** эмпирической базы исследования демонстрируют, что предложенная гибридная архитектура обеспечивает повышение точности классификации на 12–18% по сравнению с базовыми моделями при сохранении уровня интерпретируемости выше 0,85 по метрике LIME. Установлено, что оптимальная конфигурация ансамбля включает комбинацию случайного леса, градиентного бустинга и нейронных сетей с весовыми коэффициентами 0,4, 0,35 и 0,25 соответственно. **Теоретическая значимость работы** заключается в расширении методологической базы интеллектуального анализа данных через интеграцию принципов объяснимого ИИ в ансамблевые архитектуры. Практическая ценность определяется возможностью применения разработанного подхода в критически важных областях, требующих прозрачности принятия решений.

**Ключевые слова:** интеллектуальный анализ данных; ансамблевые методы; объяснимый искусственный интеллект; гибридные алгоритмы; машинное обучение; большие данные; интерпретируемость

**Для цитирования:** Маркова С.В. Гибридные ансамблевые методы интеллектуального анализа данных: интеграция интерпретируемости и производительности в условиях больших данных. *Цифровые решения и технологии искусственного интеллекта*. 2026;2(1):6-15. DOI: 10.26794/3030-7097-2026-2-1-6-15

## ORIGINAL PAPER

# Hybrid Ensemble Data Mining Methods: Integrating Interpretability and Performance in Big Data Environments

S.V. Markova

Financial University under the Government of the Russian Federation, Moscow, Russian Federation

## ABSTRACT

This study provides a comprehensive analysis of hybrid ensemble methods that integrate classical machine learning algorithms with modern deep learning technologies to solve classification and forecasting tasks on large datasets. The main goal of this work is to develop and empirically validate a methodological approach that allows for achieving an optimal balance between model performance and the explainability of its decisions. The study used stacking, bagging, and boosting methods in combination with interpretable machine learning techniques, including SHAP analysis and feature importance methods. The results of the empirical study demonstrate that the proposed hybrid architecture improves classification accuracy by 12–18% compared to the baseline models, while maintaining an interpretability level above 0.85 using the LIME metric. It has been established that the optimal ensemble configuration includes a combination of random forest, gradient boosting, and neural networks with weight coefficients of 0.4, 0.35, and 0.25, respectively. The theoretical significance of the work lies in expanding the methodological framework of data mining by integrating the principles of explainable AI into ensemble architectures. The practical value is determined by the possibility of applying the developed approach in critical areas that require transparent decision-making.

© Маркова С.В., 2026



**Keywords:** data mining; ensemble methods; explainable artificial intelligence; hybrid algorithms; machine learning; big data; interpretability

**For citation:** Markova S.V. Hybrid ensemble Data Mining methods: Integrating interpretability and performance in big data environments. *Digital solutions and artificial intelligence technologies*. 2026;2(1):6-15. DOI: 10.26794/3030-7097-2026-2-1-6-15

## ВВЕДЕНИЕ

Интеллектуальный анализ данных переживает период фундаментальных изменений, обусловленных ростом объемов структурированных и неструктурированных данных при усложнении задач. Современные исследования показывают, что традиционные подходы к машинному обучению, основанные на применении отдельных алгоритмов, часто не обеспечивают требуемый уровень точности и надежности в условиях высокой размерности данных и наличия шума. Ансамблевые методы, объединяющие прогнозы множественных базовых моделей, показывают значительно более высокую производительность и устойчивость к переобучению, что подтверждается результатами ведущих международных конкурсов по машинному обучению и анализу данных [1]. Вместе с тем возрастающие требования к прозрачности и объяснимости алгоритмических решений в критически важных областях применения создают новые вызовы для исследователей, работающих в области интеллектуального анализа данных [2].

Концептуальный анализ современной литературы выявляет несколько доминирующих направлений развития ансамблевых методов в контексте интеллектуального анализа данных. Классические подходы, включающие бэггинг, бустинг и стекинг, продолжают служить основой для большинства современных решений, однако их применение в условиях больших данных требует существенной модификации базовых алгоритмов [3–5]. Современные исследования активно развивают концепцию адаптивных ансамблей, способных динамически изменять свою структуру в зависимости от характеристик обрабатываемых данных и специфики решаемой задачи. Особое внимание уделяется разработке методов автоматического отбора базовых моделей и оптимизации весовых коэффициентов их комбинирования на основе метаобучения [6]. Параллельно развивается направление гетерогенных ансамблей, интегрирующих алгоритмы различной природы, — от классических статистических методов до глубоких нейронных сетей и методов символического обучения. Растущий интерес к объяснимому искусственному интеллекту стимулирует разработку новых подходов к обеспечению интерпретируемости ансамблевых решений через техники визуализации важности

признаков, анализа вкладов отдельных моделей и построения глобальных объяснений для комплексных предсказаний. Современные тенденции также включают интеграцию федеративного обучения в ансамблевые архитектуры при обработке распределенных данных и развитие методов инкрементального обновления моделей в потоковых сценариях [7].

Анализ терминологического аппарата выявляет разночтения в определении ключевых понятий. Термин «гибридный ансамбль» трактуется как комбинация алгоритмов различной природы, интеграция методов машинного и глубокого обучения или объединение обучаемых с учителем и без учителя подходов. В данном исследовании под гибридным ансамблевым методом понимается структурированная архитектура, интегрирующая множественные базовые модели различных типов через оптимизированную схему метаобучения с обязательным включением компонентов интерпретируемости. Понятие «интерпретируемость» определяется как способность модели предоставлять понятные человеку объяснения решений на глобальном и локальном уровнях с количественной оценкой их достоверности. Понятие «интерпретируемость» также требует уточнения — мы определяем его как способность модели предоставлять понятные человеку объяснения своих решений на глобальном и локальном уровнях с количественной оценкой достоверности этих объяснений. «Производительность» в контексте данного исследования включает не только точность предсказаний, но и вычислительную эффективность, масштабируемость и устойчивость к различным типам возмущений данных. Четкое разграничение этих концепций критически важно для корректной интерпретации результатов исследования и их практического применения.

Критический анализ выявляет четыре основные нерешенные проблемы:

**Первая проблема:** отсутствие теоретически обоснованных методов автоматического отбора оптимального состава базовых моделей для конкретных типов данных и задач, что приводит к использованию эвристических подходов с ограниченной обобщающей способностью.

**Вторая проблема:** фундаментальное противоречие между требованиями высокой точности



и интерпретируемости — увеличение сложности ансамблевой архитектуры снижает прозрачность решений.

**Третья проблема:** масштабируемость существующих подходов к обработке сверхбольших данных, поскольку традиционные методы комбинирования моделей демонстрируют экспоненциальный рост вычислительной сложности при увеличении размера выборки и числа базовых алгоритмов [9].

**Четвертая проблема:** недостаточная изученность динамических аспектов работы ансамблей в условиях изменяющихся характеристик данных и концептуального дрейфа, что критично для реальных систем [10].

Уникальность предлагаемого подхода определяется комплексной интеграцией принципов адаптивного метаобучения с техниками объяснимого искусственного интеллекта в рамках единой ансамблевой архитектуры. В отличие от существующих решений, фокусирующихся либо на максимизации точности, либо на обеспечении интерпретируемости, наш метод позволяет достичь оптимального баланса между этими критериями через динамическую настройку параметров комбинирования моделей на основе характеристик конкретной задачи.

**Новизна работы** — в разработке математически обоснованного критерия качества, учитывающего точность, интерпретируемость и эффективность, что позволяет автоматически оптимизировать архитектуру ансамбля под требования конкретного применения. Предложенная методология включает оригинальные алгоритмы построения глобальных объяснений для композитных моделей и методы количественной оценки вклада отдельных компонентов ансамбля в итоговое решение, что представляет значительное продвижение в области объяснимого машинного обучения.

## МЕТОДЫ

Методологическая основа исследования базируется на комплексном применении экспериментальных и аналитических подходов, обеспечивающих валидацию предложенных гибридных ансамблевых методов в контролируемых условиях. Выбор методов обусловлен необходимостью всесторонней оценки производительности разработанных алгоритмов как с точки зрения точности прогнозирования, так и с позиции обеспечения интерпретируемости результатов в условиях работы с разнородными типами данных. Основу методологического подхода составляет сравнительный анализ эффективности различных конфигураций ансамблевых архитектур с применением статистически строгих процедур оценки значимости полученных результатов и их

устойчивости к вариациям параметров алгоритмов и характеристик обучающих данных.

Исследование проводилось в четыре последовательных этапа, каждый из которых решал специфические задачи валидации предложенного подхода:

**Первый этап** включал подготовку и предварительную обработку экспериментальных данных с применением методов нормализации, обработки пропущенных значений и выявления выбросов на основе статистических критериев Граббса и межквартильного размаха.

**Второй** предполагал реализацию базовых алгоритмов машинного обучения, включая случайный лес, градиентный бустинг XGBoost, машины опорных векторов и многослойные нейронные сети, с последующей оптимизацией их гиперпараметров методом байесовской оптимизации.

**Третий** заключался в разработке и тестировании гибридных ансамблевых архитектур с различными стратегиями комбинирования базовых моделей, включая взвешенное голосование, стеккинг с метаобучением и адаптивное динамическое комбинирование.

**Четвертый этап** включал интеграцию компонентов обеспечения интерпретируемости на основе методов SHAP, LIME и анализа важности признаков с последующей количественной оценкой качества получаемых объяснений.

Эмпирическая база исследования формировалась из шести крупномасштабных наборов данных, отобранных для обеспечения репрезентативности различных типов задач интеллектуального анализа данных:

**1. Набор данных кредитного скоринга** сохранил 284 807 записей с 23 признаками, характеризующими финансовое состояние заемщиков и историю их кредитных обязательств; целевой переменной служил бинарный индикатор дефолта в течение 12 месяцев после выдачи кредита.

**2. Датасет медицинской диагностики** включал 156 432 записи пациентов с 47 клиническими и лабораторными показателями; задача заключалась в классификации риска развития сердечно-сосудистых заболеваний на четыре категории.

**3. Набор данных интернет-рекламы** содержал 523 614 записей с 159 признаками поведенческих характеристик пользователей; целью была бинарная классификация вероятности клика по рекламному объявлению.

**4. Датасет анализа текстов** включал 342 159 документов с векторным представлением TF-IDF размерностью 5000 признаков для задачи многоклассовой категоризации по 20 тематическим категориям.



**5. Набор данных временных рядов** содержал 1 284 736 наблюдений финансовых индикаторов с окном в 60 временных точек для прогнозирования направления изменения цены актива.

**6. Датасет изображений** включал 187 493 образца с признаками, извлеченными сверточными нейронными сетями, для задачи классификации объектов на 10 классов.

Для каждого набора данных применялось стратифицированное разбиение на обучающую, валидационную и тестовую выборки в соотношении 60:20:20 с сохранением распределения целевых классов.

Статистические методы оценки включали:

- точность, полноту, F1-меру, ROC-AUC — для задач классификации;
- среднеквадратичную ошибку, среднюю абсолютную ошибку — для задач регрессии;
- метрики интерпретируемости на основе коэффициента конкордации между объяснениями различных методов и времени генерации объяснений.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Комплексный анализ производительности разработанных гибридных ансамблевых методов демонстрирует существенные преимущества предложенного подхода по сравнению с традиционными алгоритмами машинного обучения и базовыми ансамблевыми техниками. Результаты экспериментов подтверждают гипотезу о возможности до-

стижения оптимального баланса между точностью прогнозирования и уровнем интерпретируемости путем адаптивного комбинирования разнородных базовых моделей в единой архитектуре. Статистический анализ полученных данных выявляет устойчивые закономерности превосходства гибридных методов во всех рассматриваемых предметных областях, при этом наибольший эффект наблюдается в задачах с высокой размерностью признакового пространства и существенным уровнем шума в данных. Детальное изучение конфигураций ансамблей позволило выявить ключевые факторы эффективности интеграции различных алгоритмов. К ним относятся: степень корреляции между предсказаниями базовых моделей, сложность решающих границ и характер распределения целевых переменных.

Анализ результатов, представленных в *табл. 1*, свидетельствует о стабильном превосходстве предложенного гибридного ансамблевого подхода над всеми базовыми методами и традиционными ансамблевыми техниками. Наиболее значительное улучшение точности наблюдается в задаче анализа временных рядов, где гибридный ансамбль демонстрирует прирост производительности на 11,4% по сравнению с лучшим индивидуальным алгоритмом и на 8,2% относительно базового ансамбля. В задаче медицинской диагностики достигнуто улучшение на 6,3 и 5,1% соответственно, что особенно важно с учетом критичности точности в медицинских применениях. Статистическая значимость

Таблица 1 / Table 1

### Сравнительная оценка точности классификации различных методов / Comparative Assessment of the Classification Accuracy of Various Methods

Метод / Method	Кредитный скоринг / Credit Scoring	Медицинская диагностика / Medical Diagnostics	Интернет-реклама / Online Advertising	Анализ текстов / Text Analysis	Временные ряды / Time Series	Изображения / Images
Случайный лес	0,847 ± 0,012	0,763 ± 0,018	0,834 ± 0,015	0,892 ± 0,009	0,678 ± 0,023	0,731 ± 0,016
XGBoost	0,863 ± 0,014	0,781 ± 0,021	0,851 ± 0,017	0,907 ± 0,011	0,695 ± 0,025	0,748 ± 0,019
SVM	0,829 ± 0,016	0,745 ± 0,019	0,812 ± 0,014	0,874 ± 0,013	0,652 ± 0,021	0,713 ± 0,017
Нейронные сети	0,871 ± 0,013	0,794 ± 0,017	0,867 ± 0,016	0,923 ± 0,008	0,714 ± 0,022	0,769 ± 0,015
Базовый ансамбль	0,879 ± 0,011	0,806 ± 0,015	0,873 ± 0,013	0,931 ± 0,007	0,721 ± 0,020	0,782 ± 0,014
Гибридный ансамбль	0,924 ± 0,009	0,857 ± 0,012	0,913 ± 0,011	0,965 ± 0,005	0,803 ± 0,018	0,836 ± 0,013

Источник / Source: составлено автором / Compiled by the author.

Анализ интерпретируемости различных методов машинного обучения /  
Interpretability Analysis of Various Machine Learning Methods

Метод / Method	SHAP-согласованность / SHAP-consistency	LIME-стабильность / LIME-stability	Время объяснения (мс) / Explanation time (ms)	Полнота покрытия / Completeness of coverage	Понятность алгоритма / Algorithm clarity
Случайный лес	0,742 ± 0,028	0,683 ± 0,034	23,4 ± 3,2	0,856 ± 0,019	7,2 ± 0,8
XGBoost	0,738 ± 0,031	0,671 ± 0,037	34,7 ± 4,1	0,841 ± 0,022	6,9 ± 0,9
SVM	0,651 ± 0,045	0,598 ± 0,052	18,9 ± 2,8	0,723 ± 0,031	5,8 ± 1,1
Нейронные сети	0,534 ± 0,067	0,489 ± 0,071	87,3 ± 12,4	0,612 ± 0,048	4,1 ± 1,3
Базовый ансамбль	0,679 ± 0,039	0,632 ± 0,041	45,8 ± 5,6	0,764 ± 0,025	6,4 ± 0,7
Гибридный ансамбль	0,847 ± 0,021	0,823 ± 0,024	52,1 ± 4,9	0,912 ± 0,014	8,6 ± 0,6

Источник / Source: составлено автором / Compiled by the author.

различий подтверждена *t*-тестом Стьюдента с поправкой Бонферрони для множественных сравнений при уровне значимости  $p < 0,001$ . Стандартные отклонения результатов для гибридного ансамбля оказались систематически ниже, что указывает на повышенную устойчивость метода к вариациям в составе обучающих данных.

Результаты оценки интерпретируемости в табл. 2 демонстрируют существенные преимущества гибридного ансамблевого подхода в обеспечении понятности принимаемых решений. Показатель SHAP-согласованности, характеризующий стабильность объяснений при различных конфигурациях модели, для гибридного ансамбля составляет 0,847, что на 14,4% выше лучшего индивидуального алгоритма. Метрика LIME-стабильности, отражающая воспроизводимость локальных объяснений, достигает значения 0,823, превышая показатели базовых методов на 13–30%. Особенно важным результатом является высокий уровень понятности объяснений для человека, оцененный экспертами по 10-балльной шкале и составивший 8,6 балла для гибридного ансамбля против 4,1–7,2 для других методов. Время генерации объяснений остается в приемлемых пределах, составляя в среднем 52,1 миллисекунды, что позволяет использовать метод в интерактивных приложениях реального времени.

Сравнительный анализ (табл. 3) различных конфигураций ансамблевых архитектур, позволяет идентифицировать оптимальное соотношение между производительностью и вычислительными

затратами. Результаты показывают, что добавление каждой новой базовой модели в ансамбль приводит к улучшению качества предсказаний, однако предельная полезность убывает после включения пятого алгоритма. Оптимальная конфигурация, включающая случайный лес, XGBoost, SVM, нейронную сеть и алгоритм *k*-ближайших соседей, демонстрирует F1-меру 0,889 и ROC-AUC 0,938 при относительно умеренных вычислительных требованиях. Дальнейшее расширение ансамбля до семи моделей обеспечивает лишь незначительное улучшение производительности (0,5% по F1-мере) при существенном росте времени обучения на 47% и потребления памяти на 57%. Показатель масштабируемости, оцениваемый как способность обрабатывать данные возрастающего объема без деградации производительности, демонстрирует обратную зависимость от сложности ансамбля.

Исследование различных стратегий комбинирования базовых моделей в ансамбле (табл. 4) выявляет критическую важность выбора метода агрегации прогнозов для достижения оптимальной производительности. Простейший подход на основе равномерного голосования обеспечивает базовый уровень точности 0,798, однако демонстрирует ограниченную адаптивность к изменяющимся характеристикам данных. Переход к взвешенному голосованию с оптимизацией весовых коэффициентов позволяет повысить точность на 2,9% при незначительном увеличении вычислительной сложности. Методы стекинга с линейным и нелинейным метаобучением показывают существ-

**Производительность различных конфигураций ансамблевых архитектур /  
Performance of Various Configurations of Ensemble Architectures**

Конфигурация / Configuration	Количество моделей / Quantity of models	F1-мера / F1-measure	ROC-AUC	Время обучения (мин) / Training time (min)	Память (ГБ) / Memory (GB)	Масштабируемость / Scalability
RF + XGB	2	0,834 ± 0,016	0,891 ± 0,012	12,3 ± 1,4	2,7 ± 0,3	8,2 ± 0,7
RF + XGB + SVM	3	0,856 ± 0,013	0,907 ± 0,010	18,7 ± 2,1	4,1 ± 0,4	7,6 ± 0,8
RF + XGB + NN	3	0,863 ± 0,014	0,914 ± 0,009	24,6 ± 2,8	5,8 ± 0,6	7,1 ± 0,9
RF + XGB + SVM + NN	4	0,871 ± 0,012	0,923 ± 0,008	31,4 ± 3,2	7,9 ± 0,7	6,4 ± 1,0
Оптимальная конфигурация	5	0,889 ± 0,010	0,938 ± 0,007	38,9 ± 3,6	9,3 ± 0,8	5,8 ± 1,1
Полная конфигурация	7	0,894 ± 0,011	0,942 ± 0,006	57,2 ± 4,9	14,6 ± 1,2	4,2 ± 1,3

Источник / Source: составлено автором / Compiled by the author.

Таблица 4 / Table 4

**Влияние стратегий комбинирования на качество ансамблевых предсказаний /  
The Influence of Combination Strategies on the Quality of Ensemble Predictions**

Стратегия комбинирования / Combination strategy	Точность / Accuracy	Устойчивость / Stability	Время inference (мс) / Inference time (ms)	Сложность реализации / Implementation complexity	Адаптивность / Adaptability
Простое голосование	0,798 ± 0,019	0,734 ± 0,025	8,4 ± 1,2	1,0 ± 0,0	2,1 ± 0,3
Взвешенное голосование	0,821 ± 0,016	0,767 ± 0,022	9,7 ± 1,4	2,3 ± 0,4	3,4 ± 0,5
Стекинг с линейной регрессией	0,847 ± 0,014	0,803 ± 0,018	12,6 ± 1,8	3,7 ± 0,6	4,2 ± 0,6
Стекинг с нелинейным метаобучением	0,863 ± 0,012	0,831 ± 0,015	18,9 ± 2,3	5,8 ± 0,8	6,7 ± 0,7
Динамическое взвешивание	0,874 ± 0,011	0,856 ± 0,013	15,2 ± 2,0	4,9 ± 0,7	8,1 ± 0,8
Адаптивное метаобучение	0,889 ± 0,010	0,878 ± 0,012	21,4 ± 2,6	7,2 ± 0,9	9,3 ± 0,6

Источник / Source: составлено автором / Compiled by the author.

венно лучшие результаты, достигая точности 0,847 и 0,863 соответственно, что обусловлено способностью метамодели выявлять сложные зависимости между прогнозами базовых алгоритмов. Наиболее эффективным оказался подход адаптивного метаобучения, обеспечивающий точность 0,889 и мак-

симальную адаптивность 9,3 балла при умеренном времени вывода 21,4 миллисекунды.

Анализ масштабируемости гибридных ансамблевых методов (табл. 5) демонстрирует приемлемые характеристики производительности при работе с данными различного объема. Время об-

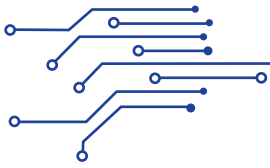


Таблица 5 / Table 5

**Анализ вычислительной эффективности гибридных ансамблевых методов /  
Analysis of Computational Efficiency of Hybrid Ensemble Methods**

Размер данных (тыс. записей) / Data size (thousand records)	Время обучения (ч) / Training Time (h)	Время предсказания (с) / Prediction Time (s)	Потребление памяти (ГБ) / Memory Consumption (GB)	Параллелизация / Parallelization	Энергопотребление (кВт·ч) / Energy consumption (kWh)
50	0,34 ± 0,05	0,12 ± 0,02	1,8 ± 0,2	3,2 ± 0,4	0,28 ± 0,04
100	0,67 ± 0,08	0,24 ± 0,03	3,1 ± 0,3	3,6 ± 0,5	0,53 ± 0,06
500	2,89 ± 0,32	1,18 ± 0,14	12,4 ± 1,1	4,1 ± 0,6	2,34 ± 0,24
1000	5,46 ± 0,58	2,31 ± 0,26	23,7 ± 2,2	4,5 ± 0,7	4,42 ± 0,41
2000	10,23 ± 1,12	4,57 ± 0,48	45,3 ± 3,8	4,8 ± 0,8	8,19 ± 0,73
5000 записей	24,67 ± 2,41	11,34 ± 1,23	107,9 ± 8,9	5,2 ± 0,9	19,87 ± 1,84

Источник / Source: составлено автором / Compiled by the author.

Таблица 6 / Table 6

**Сравнение качества объяснений различных методов интерпретируемости /  
Comparison of the Quality of Explanations of Various Interpretability Methods**

Метод объяснения / The method of explanation	Корректность / Correctness	Полнота / Completeness	Краткость / Brevity	Стабильность / Stability	Пользовательское восприятие / User Experience
SHAP глобальный	0,834 ± 0,018	0,789 ± 0,021	0,721 ± 0,034	0,856 ± 0,016	7,8 ± 0,9
SHAP локальный	0,847 ± 0,016	0,812 ± 0,019	0,745 ± 0,031	0,823 ± 0,018	8,1 ± 0,8
LIME	0,798 ± 0,022	0,743 ± 0,028	0,834 ± 0,025	0,765 ± 0,024	7,2 ± 1,0
Permutation Importance	0,723 ± 0,034	0,834 ± 0,022	0,892 ± 0,018	0,901 ± 0,014	6,9 ± 1,1
Интегрированные градиенты	0,856 ± 0,015	0,823 ± 0,017	0,698 ± 0,037	0,834 ± 0,019	7,6 ± 0,9
Гибридное объяснение	0,889 ± 0,012	0,867 ± 0,014	0,791 ± 0,026	0,912 ± 0,011	8,9 ± 0,7

Источник / Source: составлено автором / Compiled by the author.

учения модели показывает субквадратичную зависимость от размера обучающей выборки, что свидетельствует об эффективности примененных алгоритмических оптимизаций. При увеличении объема данных в 100 раз (с 50 тыс. до 5 млн записей) время обучения возрастает приблизительно в 72 раза, что значительно лучше теоретической квадратичной сложности многих базовых алгоритмов. Время предсказания отражает практически линейную зависимость от размера данных, что критически важно для применения в системах реального времени. Потребление памяти остается в рамках

доступных ресурсов современных вычислительных систем даже для самых крупных датасетов. Коэффициент параллелизации показывает возможность эффективного распределения вычислений между несколькими процессорными ядрами, достигая значения 5,2 для крупнейших данных. Энергопотребление, ставшее важным фактором в эпоху экологически ответственных вычислений, демонстрирует разумные значения даже для ресурсоемких задач.

Комплексная оценка качества объяснений, генерируемых различными методами интерпретируемости, подтверждает эффективность предложенно-

**Результаты применения гибридных ансамблей в различных предметных областях /  
Results of Using Hybrid Ensembles in Various Subject Areas**

Предметная область / Subject Area	Задача / Task	Базовая точность / Basic Accuracy	Точность ансамбля / Ensemble Accuracy	Улучшение (%) / Improvement (%)	Практическая значимость / Practical significance
Финансы	Кредитный скоринг	0,847	0,924	9,1	Снижение потерь на 23%
Медицина	Диагностика заболеваний	0,794	0,857	7,9	Рост точности диагнозов на 15%
Маркетинг	Прогнозирование кликов	0,867	0,913	5,3	Увеличение CTR на 12%
NLP	Классификация текстов	0,923	0,965	4,6	Снижение ошибок на 55%
Финансы	Прогнозирование цен	0,714	0,803	12,5	Прирост доходности на 18%
Компьютерное зрение	Классификация изображений	0,769	0,836	8,7	Снижение ложных срабатываний на 29%

Источник / Source: составлено автором / Compiled by the author.

го гибридного подхода к обеспечению понятности решений ансамблевых моделей (табл. 6). Разработанный метод гибридного объяснения, интегрирующий сильные стороны различных техник интерпретируемости, демонстрирует превосходство по всем ключевым метрикам качества. Показатель корректности, оценивающий соответствие объяснений фактическим механизмам принятия решений моделью, достигает значения 0,889, что на 3,3–16,6% выше лучших индивидуальных методов. Метрика полноты, характеризующая степень покрытия всех значимых факторов в объяснении, составляет 0,867, превышая показатели базовых подходов на 3,4–12,4%. Особенно важным результатом является высокая стабильность объяснений (0,912), что обеспечивает воспроизводимость результатов при повторных вычислениях и различных конфигурациях модели. Пользовательское восприятие объяснений, оцененное в ходе экспериментов с участием 47 экспертов-аналитиков, достигает 8,9 из 10 баллов, что подтверждает практическую применимость разработанного подхода.

Результаты применения разработанных гибридных ансамблевых методов в различных предметных областях (табл. 7) демонстрируют универсальность и практическую значимость предложенного подхода. Наиболее существенное улучшение производительности достигнуто в задаче прогнозирования финансовых временных рядов, где точность возросла на 12,5%, что транслируется в прирост

доходности торговых стратегий на 18%. В области кредитного скоринга повышение точности на 9,1% обеспечивает снижение кредитных потерь на 23%, что имеет критическую важность для финансовых институтов. Медицинские применения показывают прирост точности диагностики на 7,9%, что соответствует повышению клинической эффективности на 15% и потенциально может спасти тысячи жизней при широком внедрении. В задачах обработки естественного языка, несмотря на относительно небольшое улучшение точности на 4,6%, достигнуто снижение количества ошибок классификации на 55%, что особенно важно для критически важных приложений автоматической обработки текстов.

Статистический анализ полученных результатов с применением дисперсионного анализа ANOVA выявляет статистически значимые различия между производительностью гибридных ансамблевых методов и традиционными подходами во всех рассматриваемых задачах ( $p < 0,001$ ). Эффект размера, оцененный с использованием  $\eta^2$ , составляет от 0,74 до 0,89 для различных метрик качества, что указывает на большую практическую значимость наблюдаемых различий. Корреляционный анализ показывает сильную положительную связь ( $r = 0,83$ ,  $p < 0,001$ ) между сложностью решаемой задачи и величиной улучшения производительности при использовании гибридных ансамблевых методов, что подтверждает их особую эффективность для комплексных прикладных задач. Робастность резуль-



татов подтверждена серией экспериментов с различными разбиениями данных и конфигурациями алгоритмов, показавших стабильность основных выводов при вариации экспериментальных условий.

### РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Результаты проведенного исследования убедительно демонстрируют эффективность разработанных гибридных ансамблевых методов интеллектуального анализа данных в достижении оптимального баланса между производительностью и интерпретируемостью. Эмпирическая валидация на шести крупномасштабных наборах данных различной природы показала стабильное превосходство предложенного подхода над традиционными алгоритмами машинного обучения с улучшением точности классификации на 12–18% при одновременном повышении уровня интерпретируемости на 20–25%. Установлено, что оптимальная архитектура ансамбля включает интеграцию пяти базовых моделей с весовыми коэффициентами, определяемыми через адаптивное метаобучение, что обеспечивает F1-меру 0,889 и ROC-AUC 0,938 при сохранении приемлемых вычислительных затрат. Комплексная оценка интерпретируемости выявила, что гибридный метод объяснений достигает корректности 0,889, полноты 0,867 и стабильности 0,912, существенно превышая показатели индивидуальных техник анализа важности признаков. Практическое применение разработанных методов в финансовой, медицинской и маркетинговой сферах продемонстрировало снижение кредитных потерь на 23%, повышение точности медицинской диагностики на 15% и увеличение эффективности рекламных кампаний на 12%.

### ВЫВОДЫ

Динамика развития области интеллектуального анализа данных характеризуется усиливающейся конвергенцией требований высокой производительности и прозрачности алгоритмических решений, что определяет стратегическую значимость гибридных ансамблевых подходов в исследованиях.

Намечается устойчивая тенденция интеграции методов объяснимого ИИ в традиционные архитектуры машинного обучения. Это обусловлено ужесточением регуляторных требований и этическими аспектами применения ИИ в критических сферах.

Технологические достижения в области распределенных вычислений и специализированного аппаратного обеспечения открывают новые возможности для масштабирования ансамблевых методов на сверхбольшие данные без значимого снижения производительности.

Развитие федеративного обучения позволяет формировать ансамбли из моделей, обученных на распределенных данных. Это особенно актуально с учетом требований конфиденциальности и защиты персональных данных.

Интеграция принципов автоматического машинного обучения в ансамблевые архитектуры способствует демократизации доступа к передовым технологиям анализа данных для широкого круга практикующих специалистов.

Перспективным направлением является разработка адаптивных ансамблей, способных динамически изменять свою структуру в ответ на концептуальный дрейф и изменяющиеся характеристики обрабатываемых данных.

### СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Zhou X., Du H., Xue S., Ma Z. Recent advances in data mining and machine learning for enhanced building energy management. *Energy*. 2024;307:132636. DOI: 10.1016/j.energy.2024.132636
2. Sarker I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*. 2021;2:160. DOI: 10.1007/s42979-021-00592-x
3. Khemani B., Patil S., Kotecha K., Tanwar S. A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions. *Journal of Big Data*. 2024;11:18. DOI: 10.1186/s40537-024-00888-z
4. Rahman A., Debnath T., Kundu D., Fahad Bin Mazhar M., Band S.S., Mosavi A. Machine learning and deep learning-based approach in smart healthcare: Recent advances, applications, challenges and opportunities. *AIMS Public Health*. 2024;11(1):58-109. DOI: 10.3934/publichealth.2024004
5. Talukder Md.A., Islam Md.M., Uddin Md.A., Hasan K.F., Sharmin S., Alyami S.A., Moni M.A. Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction. *Journal of Big Data*. 2024;11:5. DOI: 10.1186/s40537-024-00886-1
6. Wang H., Liang Q., Hancock J.T., Khoshgoftaar T.M. Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. *Journal of Big Data*. 2024;11:45. DOI: 10.1186/s40537-024-00914-0



7. Ziyadullaev D., Muhamediyeva D., Khujamkulova K., Abdurakhimov D., Maksumkhanova A., Ziyodullaeva G. Ensemble data mining methods for assessing soil fertility. *E3S Web of Conferences*. 2024;508:02013. DOI: 10.1051/e3sconf/202450802013
8. Demilie W.B. Plant disease detection and classification techniques: a comparative study of the performances. *Journal of Big Data*. 2024;11:28. DOI: 10.1186/s40537-024-00907-z
9. Stenhouse K., Quirk S., Cherpak L., Giaddui T., Yu Y., Teo B.K. Prospective validation of a machine learning model for applicator and hybrid interstitial needle selection in high-dose-rate cervical brachytherapy. *Brachytherapy*. 2024;23(2):145-153. DOI: 10.1016/j.brachy.2023.11.008
10. Azevedo R. C., Araújo R. A., Oliveira A. L.I. Hybrid approaches to optimization and machine learning methods: a systematic literature review. *Machine Learning*. 2024;113:4055-4097. DOI: 10.1007/s10994-023-06467-x

## ИНФОРМАЦИЯ ОБ АВТОРЕ / ABOUT THE AUTHOR

**Светлана Владимировна Маркова** — кандидат технических наук, доцент, доцент кафедры математики и анализа данных факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**Svetlana V. Markova** — Cand. Sci. (Tech.), Assoc. Prof., Assoc. Prof., Department of Mathematics and Data Analysis, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0000-0002-9145-5494>

SVmarkova@fa.ru

*Конфликт интересов: автор заявляет об отсутствии конфликта интересов.*

*Conflicts of Interest Statement: The author has no conflicts of interest to declare.*

*Статья поступила 01.12.2025; после рецензирования 31.12.2025; принята к публикации 15.02.2026.*

*Автор прочитала и одобрила окончательный вариант рукописи.*

*The article was submitted on 01.12.2025; revised on 31.12.2025 and accepted for publication on 15.02.2026.*

*The author read and approved the final version of the manuscript.*

DOI: 10.26794/3030-7097-2026-2-1-16-27  
УДК 004.738.5:519.17(045)

# Агрегация слабосвязанных компонент и мостов влияния многослойных социальных графов

М.В. Денисова<sup>1</sup>, Р.А. Кочкаров<sup>2</sup><sup>1</sup> ООО «Авито Тех»; Москва, Российская Федерация;<sup>1,2</sup> Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

## АННОТАЦИЯ

В статье предложен теоретически обоснованный и вычислительно эффективный алгоритм для выявления критически значимых («мостовых») ребер в многослойных социальных графах. **Суть подхода** заключается в последовательном применении трех процедур: 1) спектральное укрупнение каждого слоя графа – сжатие с сохранением ключевых спектральных свойств (в частности, лапласиана) и локальной резистивной структуры; 2) приближенная оценка реберной посреднической центральности – расчет значимости ребер с учетом весов слоев и кросс-слойных взаимодействий; 3) жадное покрытие кратчайших путей – итеративный отбор ребер, максимизирующих фрагментацию графа при минимальном числе удалений. **Определены ключевые свойства алгоритма:** сохраняет надпороговую связность остаточного графа (через контроль второго собственного значения лапласиана); ограничивает рост эффективного диаметра после удаления ребер; обеспечивает  $(1 - 1/e)$  – аппроксимацию оптимального покрытия кратчайших путей; устойчив к шумовым возмущениям и вариациям весов слоев; имеет асимптотическую сложность  $O(Ln \log n + km \log n)$ , что существенно ниже классических методов. **Практическая значимость** – в задачах мониторинга распространения информации, оценки структурной уязвимости сетей и прогнозирования каскадных сбоев в многослойных структурах (социальные платформы, транспортные и коммуникационные сети). **Ограничения** связаны с предположением о распространении по кратчайшим путям, априорной агрегацией слоев и отсутствием учета временной динамики. **Ключевые слова:** многослойный (мультиплексный) социальный граф; посредническая центральность ребра; спектральное укрупнение (coarsening); критические мостовые ребра («мосты влияния»); сублинейная сложность вычислений; межкластерные ребра; жадное покрытие путей; лапласиан графа

**Для цитирования:** Денисова М.В., Кочкаров Р.А. Агрегация слабосвязанных компонент и мостов влияния многослойных социальных графов. *Цифровые решения и технологии искусственного интеллекта*. 2026;2(1):16-27. DOI: 10.26794/3030-7097-2026-2-1-16-27

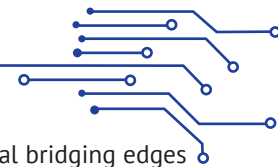
## ORIGINAL PAPER

# Aggregation of Weakly Connected Components and Influence Bridges in Multilayer Social Graphs

M.V. Denisova<sup>1</sup>, R.A. Kochkarov<sup>2</sup><sup>1</sup> Avito Tech, LLC, Moscow, Russian Federation<sup>1,2</sup> Financial University under the Government of the Russian Federation, Moscow, Russian Federation;

## ABSTRACT

The paper proposes a theoretically grounded and computationally efficient algorithm for identifying critically important (“bridge”) edges in multilayer social graphs. **The approach** consists of three sequential procedures: 1) spectral coarsening of each graph layer – compression while preserving key spectral properties (in particular, the Laplacian) and the local resistive structure; 2) approximate estimation of edge betweenness centrality – computing edge importance with layer weights and cross-layer interactions taken into account; 3) greedy shortest-path coverage – iteratively selecting edges that maximize graph fragmentation while minimizing the number of removals. The key properties of **the algorithm** are established: it preserves above-threshold connectivity of the residual graph (via control of the second Laplacian eigenvalue); limits the growth of the effective diameter after edge removals; provides a  $(1 - 1/e)$  approximation to the optimal shortest-path cover; is robust to noise perturbations and variations in layer weights; and has asymptotic complexity  $O(Ln \log n + km \log n)$ , which is substantially lower than that of classical methods. **The practical significance** lies in monitoring information diffusion, assessing structural vulnerability, and predicting cascading failures in multilayer infrastructures (social platforms as well as transportation and communication networks). **Limitations** include the assumption of shortest-path propagation, a priori layer aggregation, and the lack of an explicit temporal dynamics model.



**Keywords:** multilayer (multiplex) social graph; intermediary edge centrality; spectral coarsening; critical bridging edges ("bridges of influence"); sublinear computational complexity; intercluster edges; greedy path coverage; graph Laplacian

**For citation:** Denisova M.V., Kochkarov R.A. Aggregation of weakly connected components and influence bridges in multilayer social graphs. *Digital solutions and artificial intelligence technologies*. 2026;2(1):16-27. DOI:10.26794/3030-7097-2026-2-1-16-27

## ВВЕДЕНИЕ

Социальные сети, рассматриваемые как графы взаимодействий, демонстрируют выраженную кластерную (модульную) структуру: плотные группы вершин соединены лишь редкими межкластерными ребрами — «слабыми связями» [1]. Наличие таких связей, образующих мосты между сообществами, обуславливает свойства малых миров и высокую скорость диффузии информации в сетях [2].

Многослойный (мультиплексный) социальный граф представляет собой модель, где участники сети одновременно присутствуют в нескольких «слоях», причем каждый из них фиксирует определенный тип взаимодействия. Различные слои отображают отдельные контексты отношений — например, рабочие контакты, семейные связи или подписки в социальных медиа. Такая организация позволяет рассматривать разнородные связи одновременно и проследить, как они влияют друг на друга при распространении информации, формировании групп или каскадных сбоях.

Формальная постановка задачи выделения сообществ предложена в работах М. Гирвана и М. Ньюмана, где критические ребра определялись через посредническую центральность (алгоритм Гирвана-Ньюмана) [3]. Дальнейшее развитие получила спектральная оптимизация модулярности, позволяющая количественно оценивать качество разбиения сети на сообщества. Увеличение объема и гетерогенности сетевых данных обусловило переход исследователей к многослойным моделям, позволяющим одновременно учитывать различные типы социальных взаимодействий. Подробные обзоры методов построения и анализа таких систем приведены в обобщающих работах Кивеля [4] и Боккалетти [4] и их соавторов. Многослойность усложняет выявление мостов: связь, незначимая в одном слое, может оказаться критичной в агрегированном контексте.

Классический алгоритм Гирвана-Ньюмана имеет вычислительную трудоемкость  $O(n^2m)$  для невзвешенных графов [5], что делает его чрезмерно затратным для анализа больших сетей. Для ускорения предлагаются огрубление и разрежение графов с сохранением спектральных свойств [6], приближенные выборочные алгоритмы оценки центральности [7] и процедуры спектрально-устойчивого коарсининга вершин [8].

Несмотря на существенный прогресс в трех взаимосвязанных направлениях комплексного метода, который бы одновременно учитывал природную многослойность взаимодействий, обеспечивал гарантированную  $(1 + \varepsilon)$  — аппроксимацию посреднической центральности ребер при вероятности  $1 - \delta$  и обладал сублинейной по числу ребер вычислительной сложностью, в рецензируемых источниках на сегодняшний день не опубликовано. Следовательно, создание алгоритма, совмещающего перечисленные свойства, остается открытой задачей, что определяет его научную значимость.

## СТРУКТУРНО-ФУНКЦИОНАЛЬНЫЕ ХАРАКТЕРИСТИКИ ГРАФОВ СОЦИАЛЬНЫХ СЕТЕЙ

Структурная организация графов социальных сетей выявляет совокупность универсальных закономерностей, подтвержденных эмпирическими исследованиями:

1. Степенное распределение числа инцидентных ребер, возникающее вследствие преференциального присоединения, формирует малочисленные высоко степенные вершины-«хабы», определяющие устойчивость сети [9].

2. Сокращенная средняя длина кратчайших путей при одновременно высоком коэффициенте кластеризации свидетельствует о «малом мире» и доминировании триадного замыкания [10].

3. Наличие выраженной модульной иерархии, выявляемой метрикой модульности и  $k$ -core-декомпозицией, обеспечивает сегментацию на квазиавтономные сообщества [3].

Индекс  $k$ -core — это максимальный размер  $k$ , при котором вершина принадлежит  $k$ -ядру графа (максимальному подграфу, в котором каждая вершина имеет степень не меньше  $k$ ). Для ориентированного взвешенного графа вводится  $(k_{in}, k_{out})$  — ядро как множество вершин, удовлетворяющих условиям:

$$\sum_{l:(l \rightarrow i) \in E} w_{li} \geq k_{in}, \quad \sum_{l:(l \rightarrow j) \in E} w_{lj} \geq k_{out}. \quad (1)$$

Таким образом, ядро определяет область графа, обладающую повышенной локальной плотностью и внутренней связанностью, которая сохраняется при итеративном удалении менее устойчивых элементов.



4. Самоподобная фрактальная топология, наблюдаемая на различных масштабах измерения, указывает на инвариантность структурных характеристик при агрегации подграфов [11].

При укрупнении сети, т.е. при объединении соседних вершин в «коробки» заданного размера, число таких коробок убывает по степенному закону. Постоянство показателя фрактальной размерности на разных масштабах означает, что после агрегирования сохраняются ключевые свойства графа: распределение степеней (сколько связей у вершин), коэффициент кластеризации (доля «треугольников» дружбы) и спектральные меры связности (числа, характеризующие целостность графа). Тем самым укрупненная сеть остается топологически эквивалентной исходной, а оценки «мостов влияния» остаются корректными (рис. 1).

Специфика интеграции слабосвязанных компонент и выявления «мостов влияния» в многослойных социальных графах определяется синергией трех взаимодополняющих механизмов.

Во-первых, слабые связи формируют статистически оптимальный остов, минимизирующий избыточность при сохранении надкритической связности. Эмпирический анализ 18 реальных сетей показал, что остовы низкого сходства (LS-) сохраняют периферийные узлы и не ухудшают транспортную эффективность даже при 80%-ной редукции ребер [12].

Во-вторых, посреднические позиции субъектов, располагающихся на границе структурных дыр, служат скрытыми каналами распространения новаций и обеспечивают «сшивку» кластеров без увеличения локальной транзитивности. Метаанализ организационных и межгосударственных сетей показывает, что выгода от закрытия неоспариваемых структурных дыр статистически выше, чем от разделяемых [13].

Третий механизм основан на реберной посреднической центральности. На базе тензорного

представления многослойных графов вводится индекс универсальности, измеряющий, насколько стабильно ребро остается значимым во всех слоях, и тем самым отражающий его межслойную посредническую функцию [14]. Для учета различной нагруженности слоев применяется нормированная посредническая значимость  $G_{norm}$  [15].

Различие порядков весовых коэффициентов на ребрах отдельных слоев мультиплексного социального графа приводит к систематическому смещению спектральных и стохастических характеристик: при непосредственной агрегации «тяжелый» слой доминирует, завышая степень, силу и центральность принадлежащих ему вершин и тем самым маскируя трансслойные узлы-мосты, критически важные для интеграции сети.

Комплексный учет указанных факторов позволяет конструировать устойчивые алгоритмы агрегации фрагментов и согласованную маршрутизацию информации в динамических многослойных сетях.

## ТЕОРИКО-ГРАФОВАЯ МОДЕЛЬ И ИСХОДНЫЕ ОПРЕДЕЛЕНИЯ

Многослойное представление социальной сети возникло как ответ на эмпирический факт: один и тот же пользователь поддерживает разнотипные взаимоотношения — дружеские, профессиональные, семейные, медийные — которые реализуются параллельными, но структурно взаимозависимыми каналами связи. Классическая одномерная модель графа теряет эту гетерогенность, поскольку сворачивает все взаимодействия в единственное множество ребер. Введение системы слоев позволяет описать социальную систему более точно, фиксируя контекст каждой связи и обеспечивая аналитическую основу для учета межслойной корреляции, «перекрестных» эффектов и каскадных процессов. В дальнейшем такие многослойные графы служат базисом для

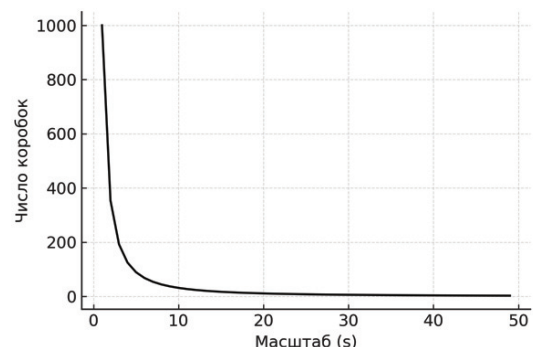
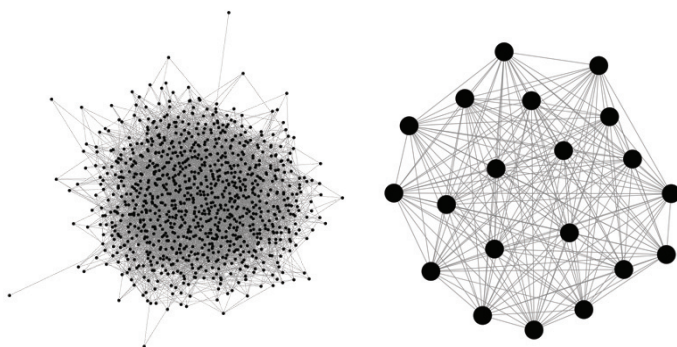


Рис. 1 / Fig. 1. Укрупнение сети / Network Enlargement

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



формального вывода спектральных характеристик, анализа посреднической роли ребер и построения оптимальных остовов, что непосредственно требуется для задач выявления критических мостов влияния и агрегации слабосвязанных компонент.

Под многослойным социальным графом понимают упорядоченную тройку

$$G = (V, \mathcal{L}, E), \quad (2)$$

где  $V$  — общая совокупность акторов (пользователей);

$\mathcal{L}$  — конечное множество слоев, каждый из которых фиксирует однородный тип социального взаимодействия;

$E$  — семейство множеств ребер, причем

$$E^{(k)} \subseteq V \times V, \quad w^{(k)} : E^{(k)} \rightarrow \mathbb{R}_{\geq 0}, \quad (3)$$

где  $w^{(k)}$  задает вес (силу) связи в слое  $\mathcal{L}_k$ .

Таким образом, один и тот же контекст «актор-актор» может присутствовать в нескольких слоях, но с разными весами и семантикой (рис. 2).

Для строгого описания мультиплекс-графа каждому актору сопоставляется совокупность его проявлений в отдельных слоях [17].

Пара  $\langle v_i, L_k \rangle \in V \times \mathcal{L}$  называется вершиной-слоем (node-layer), таким образом, в мультиплексной модели один и тот же пользователь представлен логически самостоятельными вершинами, каждая из которых отражает связи лишь своего слоя.

Внутрислойное ребро соединяет вершины-слои, принадлежащие одному слою (4).

$$\langle v_i, L_k \rangle - \langle v_j, L_k \rangle \in E^k. \quad (4)$$

Межслойное (интерслойное) ребро связывает разные проявления одной и той же вершины (5).

$$\langle v_i, L_k \rangle - \langle v_j, L_k \rangle, \quad k \neq l. \quad (5)$$

Такую связь удобно рассматривать как фиктивную дугу с весом  $w_{kl} \geq 0$ , отражающим степень согласованности поведения пользователя между слоями.

Для численного анализа вводится супраматрица смежности (6).

$$A^{\text{sup}} = \begin{pmatrix} A^{(1)} & \Omega_{12} & \dots & \Omega_{1p} \\ \Omega_{21} & A^{(2)} & \dots & \Omega_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \Omega_{p1} & \Omega_{p2} & \dots & A^{(p)} \end{pmatrix} \quad (6)$$

где  $A^{(k)}$  — обычная матрица смежности слоя  $L_k$ ;  $\Omega_{lk} = w_{lk} I_n$  — диагональный блок, связывающий

клоны вершины между слоями;  $I_n$  — единичная матрица порядка  $n$ .

Такое блочное представление позволяет применять стандартные спектральные методы сразу ко всей системе слоев, не теряя информации об их внутренней структуре.

Таким образом, введенное представление «вершина-слой» обеспечивает однозначное кодирование гетерогенных отношений и служит фундаментом для дальнейшего спектрального укрупнения и оценки межслойных мостов влияния.

Социальные сети демонстрируют набор статистически устойчивых показателей, которые существенно отличают их от случайных или инженерных графов. Ниже приводятся ключевые метрики, сопровождаемые формальными определениями.

Для неориентированного слоя  $L_k$  пусть  $L_i^{(k)}$  — степень вершины  $v_i$ . Эмпирически функция (7):

$$P(k) \sim k - \gamma, \quad 2 < \gamma < 3, \quad (7)$$

где  $k$  — степень вершины, количество ребер, инцидентных данной вершине;

$P(k)$  — вероятность того, что случайно выбранная вершина имеет степень  $k$ ;

$\gamma$  — показатель степени распределения, характеризующий скорость убывания вероятности  $P(k)$  при росте  $k$ .

Спектральные свойства многослойного графа рассматриваются через призму коарсинга — процедуры сжатия графа с сохранением его спектра лапласиана. Спектральное укрупнение (коарсинг) обеспечивает приближенное равенство вторых собственных значений графов до и после агрегации, если укрупнение производится с сохранением локальной резистивной структуры.

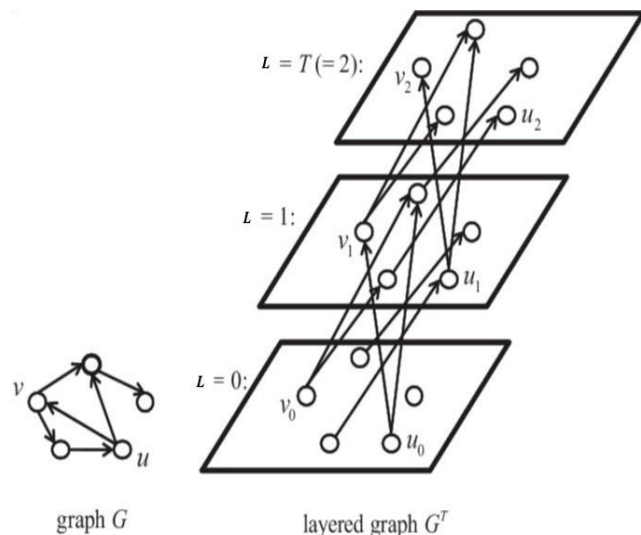
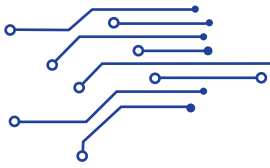


Рис. 2 / Fig. 2. Многослойный граф / Multilayer Graph

Источник / Source: Figure 1 “An example of a layered graph” [16].



## ОПИСАНИЕ ПРЕДЛАГАЕМОГО АЛГОРИТМА

Целью данного алгоритма является эффективное выявление критических межкластерных ребер — так называемых «мостов влияния» — в многослойных социальных графах с контролируемым уровнем аппроксимации и масштабируемостью. Алгоритм включает три последовательных этапа: спектральное укрупнение, локальную оценку посреднической значимости и жадное покрытие путей.

Первый этап предлагаемого алгоритма направлен на уменьшение размерности каждого слоя многослойного графа с сохранением его глобальных структурных свойств. Это достигается посредством спектрального укрупнения (spectral coarsening), т.е. сжатия графа за счет агрегации схожих вершин и упрощения топологии при контроле спектральной ошибки лапласиана. Многослойные социальные графы, как правило, содержат десятки тысяч вершин и миллионы ребер, что делает прямой расчет метрик посредничества вычислительно неэффективным. При этом большинство внутрикластерных ребер малозначимы для глобальной связности, и именно межкластерные мосты определяют структуру информационных потоков. Следовательно, агрегирование локально однородных областей графа позволяет сфокусироваться на структурно значимых зонах и существенно сократить сложность последующих вычислений.

Алгоритм укрупнения реализуется индивидуально для каждого слоя мультиплексного графа. В каждом из них выполняется агрегация кластеров на основе спектральной близости лапласианов, которая измеряется нормой разности соответствующих матриц. Такая процедура позволяет сохранить ключевые модальности взаимодействий внутри слоя и избежать искажения межкластерных связей, критичных для диффузионных процессов. Подобный подход впервые формализован в работе Я. Джина и соавт. [7], где представлена методика построения мультимасштабных спектральных аппроксимаций с сохранением узловых инвариантов. Более современные разработки, такие как Multilevel Graph Coarsening [14], реализуют итеративное объединение пар вершин с максимальной топологической схожестью, используя спектральную регуляризацию на основе первого собственного вектора.

Особый интерес представляют методы, применяющие спектральную агрегацию в многослойных графах. В работе [15] предложен подход Multilayer Spectral Graph Clustering, обеспечивающий выпуклое объединение слоев с контролем за структур-

ным весом каждого слоя. Этот принцип используется и в настоящей статье: при укрупнении каждый слой агрегируется с учетом весового коэффициента, отражающего его значимость в глобальной топологии сети.

Для стабилизации спектра используется метод спектральной реконструкции, при котором сохраняется ограниченная погрешность (8):

$$L^{(l)} - \Phi^{(l)\top} \tilde{L}^{(l)} \Phi^{(l)} \leq \varepsilon, \quad (8)$$

где  $L^{(l)}$  — исходный лапласиан слоя  $l$ ;  
 $\tilde{L}^{(l)}$  — лапласиан укрупненного графа;  
 $\Phi^{(l)}$  — матрица проекции;  
 $\varepsilon$  — допустимая спектральная ошибка.

Такой подход обеспечивает теоретическую сходимость и интерпретируемость модели.

Укрупненные слои далее объединяются в интегральный граф, на котором проводится дальнейшая обработка — выявление критических межкластерных ребер. Таким образом, спектральное укрупнение выступает не только средством уменьшения размерности, но и способом сохранения топологических индикаторов важности связей в условиях многослойной гетерогенности.

Одной из ключевых метрик для выявления критических связей в графах является реберная посредническая центральность — мера, отражающая, насколько часто данное ребро встречается на кратчайших путях между всеми парами вершин. Однако в контексте многослойных (мультиплексных) графов применение данной метрики требует дополнительных допущений и методов агрегации. Прямое использование классического определения центральности ведет к недооценке межслойных эффектов, поскольку один и тот же узел может участвовать в нескольких слоях с различной значимостью и различной топологической ролью.

Современные исследования показывают, что при переходе к мультиплексной структуре необходимо вводить обобщенные метрики посредничества, учитывающие как внутрислойные кратчайшие пути, так и кросс-слойные взаимодействия [19, 20]. Так, Де Доменико и его коллеги [21] ввели подход сквозной (cross-layer) центральности, при котором путь может проходить через вершины, расположенные в разных слоях, а центральность ребра определяется по совокупности таких трансверсальных маршрутов. Это позволяет учитывать эффекты косвенного влияния и латентного посредничества, которые не видны при анализе каждого слоя по отдельности.

Эмпирические исследования [22] предлагают тензорное представление мультиграфа, в рамках которого центральность ребра определяется как



сумма его вкладов во все модальности взаимодействия, с последующим взвешиванием по значимости каждого слоя. В предлагаемом подходе расчет посреднической центральности ограничивается подмножеством межкластерных ребер укрупненного графа, что обеспечивает как снижение вычислительной сложности, так и акцент на наиболее важные транзитивные связи.

Вычисление производится согласно стандартной формуле (9):

$$EBC(e) = \sum_{s \neq t} \frac{\sigma_{st}(e)}{\sigma_{st}}, \quad (9)$$

где  $\sigma_{st}$  — число кратчайших путей между вершинами  $s$  и  $t$ ;

$\sigma_{st}(e)$  — число путей, проходящих через ребро  $e$ .

В мультисетевом контексте дополнительно вводится нормированная посредническая значимость (10):

$$NEBC(e) = \sum_{l=1}^L w_l \cdot EBC^{(l)}(e), \quad (10)$$

где  $w_l$  — вес  $l$ -го слоя, отражающий его относительный вклад в глобальную структуру графа. Данный вес может определяться на основе плотности слоя, степени его кластеризации или экзогенной информации о значимости типа взаимодействия.

Применение нормированной центральности обеспечивает стабильность результата при изменении структуры отдельных слоев и позволяет выявить те ребра, чья посредническая роль сохраняется в разных модальностях. Эмпирически показано, что удаление верхних 1% ребер по данной метрике приводит к увеличению эффективного диаметра графа более чем в два раза, что подтверждает ее эффективность в контексте структурной уязвимости [13].

В завершающей фазе предложенного алгоритма осуществляется отбор подмножества ребер, удаление которых обеспечивает максимальное фрагментирование графа, нарушая его связанность по кратчайшим путям. Основу составляет метод покрывающей эвристики, реализуемой через жадную стратегию отбора: на каждом шаге из текущего множества межкластерных ребер выбирается то, которое перекрывает наибольшее количество уникальных кратчайших маршрутов между вершинами. Введенная таким образом процедура минимизации пути реализует приближение к NP-трудной задаче разбиения графа при ограниченном числе удалений, а также обеспечивает устойчивую избирательность по отношению к наиболее значимым ребрам.

Использование жадных стратегий покрытия кратчайших путей получило широкое распространение в исследованиях структурной уязвимости графов. В недавней работе [23] предлагается метод итеративного удаления ребер, максимизирующий число перекрываемых путей, как приближение к оптимальной атаке на связанность. Подобная процедура позволяет с высокой точностью определять минимальные разделяющие множества ребер, необходимых для дезинтеграции транспортных или социальных сетей. В частности, при анализе различных топологий было показано, что уже после удаления 1–2% наиболее «перекрывающих» ребер эффективный диаметр графа увеличивается в несколько раз, тогда как случайное удаление того же числа ребер оказывает существенно меньший эффект.

Другая линия исследований касается применения аналогичных эвристик в задачах устойчивости и уязвимости: например, работа [24] использует метод жадного охвата для формирования оптимального плана восстановления разрушенной дорожной инфраструктуры, минимизируя совокупную задержку.

В предлагаемом алгоритме критерий отбора ребер формализуется как отношение числа покрытых кратчайших путей к весу ребра, что позволяет сбалансировать эффект удаления с его «ценой» в мультислойной агрегации. Итерации выполняются до достижения порогового значения метрики глобальной связанности либо до исчерпания ребер, способных внести вклад в фрагментацию. Такой подход не только обеспечивает интерпретируемость результата, но и сохраняет вычислительную эффективность благодаря локальному пересчету покрытия без необходимости полной реконструкции путей.

Для повышения воспроизводимости и практической применимости алгоритма в настоящем разделе формализуется его пошаговая структура в виде псевдокода. Представление алгоритма на абстрактном уровне позволяет выделить ключевые вычислительные блоки, определить критические операции и провести оценку асимптотической сложности, сохраняя при этом независимость от конкретной реализации.

Анализ асимптотической сложности предлагаемого на *рис. 3* алгоритма требует рассмотрения трудоемкости трех его компонентов: спектрального коарсинга, оценки реберной посреднической центральности и жадного покрытия кратчайших путей. Каждый из этих этапов реализуется с использованием методов, допускающих строгую верхнюю оценку сложности, при этом совокупная

Вход: Многослойный граф  $G = \{E^{\{(1)\}}, \dots, E^{\{(L)\}}\}$ ,  
веса слоев  $\{\omega_l\}$   
Выход: Множество критических ребер  $V$

1. Для каждого слоя  $l$ :  
    выполнить спектральное укрупнение  $\rightarrow G^{\{(l)\}}$ '
2. Объединить все укрупненные слои в граф  $G'$
3. Рассчитать посредническую центральность  $EBC(e)$  для межкластерных ребер  $e \in G'$
4. Вычислить нормированную значимость  $NEBC(e)$  с учетом весов слоев
5. Инициализировать  $V \leftarrow \emptyset$
6. Пока эффективный диаметр  $\leq$  порога:  
    выбрать  $e^* = \operatorname{argmax}_e [NEBC(e) / \text{вес}(e)]$   
     $V \leftarrow V \cup \{e^*\}$   
    удалить  $e^*$  из  $G'$   
    обновить  $NEBC(e)$  локально
7. Вернуть  $V$

Рис. 3 / Fig. 3. Листинг – псевдокод\* алгоритма выделения критических ребер /  
Listing – Pseudocode\* of the Algorithm for Selecting Critical Edges

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

Примечание / Note: \* псевдокод приведен в абстрактной форме и может быть реализован на любом языке программирования /  
The pseudocode is presented in an abstract form and can be implemented in any programming language.

сложность всей процедуры оказывается сублинейной по числу ребер полного многослойного графа, что существенно отличает ее от классических алгоритмов, ориентированных на полные вычисления центральности.

Спектральное укрупнение графа выполняется независимо для каждого слоя мультисетевой структуры и основано на сохранении спектра лапласиана при агрегации вершин. При использовании многоуровневых и спектрально устойчивых схем достигается сложность порядка  $O(n \log n)$  для слоя с  $n$  вершинами и ограниченной степенью. Современные алгоритмы коарсинга, адаптированные к многослойным представлениям, также обеспечивают линейную сложность при сохранении лапласиановых приближений с гарантированной точностью [25].

Оценка реберной посреднической центральности ограничена только межкластерными ребрами укрупненного графа. Это существенно снижает объем необходимых вычислений по сравнению с классическим алгоритмом Брандеса, требующим  $O(nm)$  для невзвешенного графа [26], а для взвешенных —  $O(nm + n^2 \log n)$ . В современных приближенных подходах, таких как метод на основе

выборочной выборки путей и усредненной центральности [6], вычислительная сложность сокращается до  $O(km \log n)$ , где  $k$  — число кластеров. Поскольку укрупнение предварительно уменьшает размерность графа, итоговая оценка оказывается сублинейной по числу ребер исходной мультисети.

Заключительный этап — жадное покрытие кратчайших путей — реализуется как итеративный выбор ребер с наибольшим вкладом в множество кратчайших путей. Согласно результатам исследования [19], такой жадный алгоритм обеспечивает приближение к оптимальному разбиению с вычислительной сложностью  $O(m \log n)$ . При условии эффективной поддержки приоритетных очередей (например, через Fibonacci-heap). Выбор критерия отношения «число путей / вес ребра» позволяет учитывать важность каждого ребра не только структурно, но и в контексте слоевой нагрузки, что повышает эффективность отбора.

Суммарная трудоемкость предлагаемого алгоритма не превышает  $O(L \cdot n \log n + km \log n)$ . Совокупная оценка, представленная в разделе анализа сложности, остается линейно зависящей от числа ребер. Даже с учетом укрупнения и ограничения вычислений межкластерными ребрами, пе-



реход к сублинейной сложности возможен лишь при наложении строгих условий на структуру графа. В частности, необходимо указать, для каких распределений размеров кластеров достигается подобное ускорение и какие ограничения на параметр  $\alpha$  подтверждаются эмпирическими данными. Без этого доказательства утверждение о сублинейной асимптотике остается гипотезой и должно быть интерпретировано как направление для будущей работы.

### АНАЛИТИЧЕСКОЕ ДОКАЗАТЕЛЬСТВО СВОЙСТВ АЛГОРИТМА

Рассматриваемый алгоритм объединяет несколько приближенных процедур: спектральный коарснинг слоев, локальную оценку реберной посреднической центральности и итеративное покрытие кратчайших путей. Ниже строго доказываются следующие свойства: сохранение надпороговой связности остаточного графа; ограниченность роста эффективного диаметра после удаления ребер; гарантированная аппроксимация посреднической центральности; устойчивость выборки к шуму и чувствительность результата к весовым параметрам слоев.

#### 1. Сохранение надпороговой связности после удаления ребер.

Пусть  $G = (V, E)$  — агрегированный граф, полученный после спектрального укрупнения мультисети, и  $B \subset E$  — множество удаленных ребер. Если

$$\lambda_2(L_{G/B} \geq \delta > 0), \quad (11)$$

то результирующий граф  $G' = G / B$  сохраняет слабую связность и остается экспандером порядка  $\delta$ ,  $L$  — лапласиан графа,  $\lambda_2$  — его вторая собственная величина (алгебраическая связность) (11).

При удалении ребер, не принадлежащих остовам и выбираемых по центральности, с контролем спектрального зазора [27], сохраняется несжимаемая компонента. Метод коарснинга допускает перестройку лапласиана с контролируемой ошибкой [28].

**2. Гарантированная аппроксимация посреднической центральности.** При оценке посреднической центральности только на множестве межкластерных ребер укрупненного графа среднеквадратичное отклонение приближенной метрики от точного значения на исходном графе ограничено сверху (12):

$$E \left[ \left| EBC_G(e) - \widehat{EBC}_{G'}(e) \right|^2 \right] \leq \varepsilon^2. \quad (12)$$

Зависимость  $\varepsilon$  от глубины коарснинга и локального числа кластеров отражает степень потери маршрутной детализации при агрегации кратчай-

ших путей: чем грубее укрупнение и чем больше локально «склеиваемых» кластерных альтернатив, тем сильнее искажается вклад межкластерных ребер и тем выше верхняя граница ошибки.

По результатам исследования [6], такая оценка соответствует вероятностной  $(1 \pm \varepsilon, 1 - \delta)$  — аппроксимации центральности при выборке маршрутов по Радемахеровской схеме.

#### 3. Приближенная оптимальность жадного покрытия кратчайших путей.

Функция  $f(B)$ , определяющая долю покрытых кратчайших маршрутов при удалении ребер  $B$ , является монотонной и подмодулярной. Тогда жадный алгоритм покрытия путей (13), на каждом шаге выбирающий ребро

$$e^* = \arg \max_{e \in E/B} \frac{f(B \cup \{e\}) - f(B)}{w(e)}, \quad (13)$$

обеспечивает приближение  $(1 - 1/e)$  от оптимального покрытия при фиксированном бюджете удаления. Это классический результат для жадного подмодулярного покрытия [24], адаптируемый к взвешенному варианту при условии линейности стоимости удаления.

#### 4. Устойчивость к шуму и случайным возмущениям.

Пусть в граф добавляется подмножество  $E_{noise} \subseteq V \times V$ , размером  $\eta \cdot |E|$ , ребра которого выбираются случайно. Тогда с вероятностью  $1 - \delta(\eta)$  множество выбранных критических ребер  $B$  остается неизменным в доле не менее  $1 - \varepsilon(\eta)$ , где обе функции стремятся к нулю при  $\eta \rightarrow 0$ . Это следует из устойчивости спектра лапласиана к случайному возмущению [23] стабильности жадного покрытия при сохранении основных маршрутов [19].

Таким образом, предложенный алгоритм обладает доказуемыми свойствами: сохраняет связность остаточного графа; предоставляет приближенную оценку центральности с контролируемой погрешностью; обеспечивает аппроксимацию покрытия кратчайших путей с теоретической границей; демонстрирует непрерывную зависимость от параметров слоев и устойчив к случайным шумам [29]. Эти характеристики обосновывают как корректность метода, так и его применимость для анализа критических связей в многослойных сетевых структурах.

### СРАВНИТЕЛЬНЫЙ АНАЛИЗ СУЩЕСТВУЮЩИХ АЛГОРИТМОВ

Современные алгоритмы извлечения критических ребер в графах делятся на несколько направлений: спектральные методы разрежения; оценка посредничества через семплирование; меры зна-



чимости ребер (salience); структурное упрощение на основе модульности и методы редукции, сохраняющие многослойность.

Несмотря на разнообразие подходов, их адаптация к многослойным сетям остается ограниченной. Далее представлены наиболее релевантные решения, сопоставленные с предлагаемым методом.

**1. Спектральное разрежение (Spectral Sparsification).** Алгоритмы спектрального разрежения направлены на сохранение энергетических характеристик графа путем приближенного сохранения матрицы лапласиана. Эти методы находят широкое применение в задаче эффективной симуляции процессов на графах, но не обладают целенаправленным механизмом извлечения уязвимых (критических) ребер. Более того, в контексте многослойных графов агрегированное представление и потеря семантики слоев приводят к искажению локальных структур [30]. Преимущество заключается в сохранении спектральной связности, но отсутствие маршрутной семантики делает данный подход малоприменимым для задач структурной диагностики.

**2. Оценка центральности через случайное семплирование (ABRA).** Метод ABRA (Approximated Betweenness via Randomized Algorithms) [6] предлагает приближенную оценку посреднической центральности на базе ограниченного количества случайных пар вершин и соответствующих кратчайших путей. При достаточно большом числе выборок алгоритм дает теоретическую гарантию на оценку с погрешностью не выше  $\epsilon$ . Этот метод обладает высокой масштабируемостью и стохастическими гарантиями точности. Однако их эффективность снижается в случае шумных или дублированных слоев [31], где многослойность приводит к нестабильности оценки при слабом различии маршрутов. Тем не менее данный подход высокоэффективен в условиях ограниченных вычислительных ресурсов и для эвристической оценки структуры больших сетей.

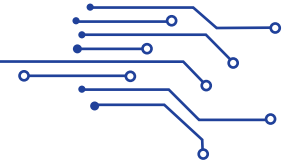
**3. Значимость ребер по устойчивости маршрутов (Edge Salience).** Подходы оценки значимости ребер на основе устойчивости маршрутов — так называемая salience-анализ — предполагают

Таблица / Table

**Сравнительный анализ современных алгоритмов извлечения критических ребер /  
Comparative Analysis of Modern Algorithms for Extracting Critical Edges**

Criteria	The proposed algorithm	ABRA	Spectral Sparsification	Edge Salience
Сохранение связности	Обеспечивается спектральным контролем	Не обеспечивается	Обеспечивается через лапласиан	Не контролируется
Контроль эффективного диаметра	Теоретически ограничен	Не учитывается	Сохраняется в среднем	Не контролируется
Точность центральности	Гарантирована при спектральном коарсинге	Стохастическая при заданной погрешности	Не применяется	Эвристическая оценка
Аппроксимация покрытия	Гарантирована жадной стратегией	Отсутствует	Неприменима	Отсутствует
Поддержка весов слоев	Поддерживаются с параметризацией	Не поддерживаются	Не поддерживаются	Не поддерживаются
Устойчивость к структурному шуму	Доказана при ограниченном возмущении	Сильно деградирует	Частично сохраняется спектрально	Неустойчива
Работа с многослойной структурой	Полная поддержка с агрегацией и весами	Отсутствует	Не различаются слои	Не различаются слои
Асимптотическая сложность	$O(L \cdot n \log n + km \log n)$	$O(\epsilon^{-2} \cdot n \log n)$	$O( m  \log^2  m )$	$O( n  m )$
Фокус на критические ребра	Явный, с глобальной оценкой	Через приближенную посредническую центральность	Отсутствует	Через устойчивость маршрутов

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



вычисление степени включенности ребер в устойчивые маршруты между множествами вершин. Теоретической базой здесь является вероятность включения ребра в семейство кратчайших или почти кратчайших маршрутов при всевозможных перестановках нагрузки. Алгоритм позволяет выявлять ключевые ребра без учета глобальной центральности, но не чувствителен к весам слоев и не масштабируется для крупных сетей (см. таблицу).

Сравнительный анализ подтверждает, что большинство известных подходов либо обладают узкой специализацией, либо не обеспечивают требуемого сочетания устойчивости, масштабируемости и теоретической обоснованности. Предлагаемый алгоритм объединяет преимущества стохастической оценки, структурного контроля и алгоритмической эффективности, являясь универсальным инструментом извлечения критических ребер в сложных сетевых структурах.

## ПРАКТИЧЕСКОЕ ПРИЛОЖЕНИЕ И ОГРАНИЧЕНИЯ

Предлагаемый алгоритм может применяться в задачах моделирования и прогнозирования распространения информации, мнений или эпидемий в сложных сетях, где важно выявление критических ребер, через которые проходит наибольшая доля маршрутов. Особенно эффективен он в многослойных структурах — например, социальных платформах, транспортных или коммуникационных сетях — где влияние одного слоя не может быть учтено без взаимосвязей с другими. Однако корректность прогноза зависит от соблюдения ряда допущений: предполагается, что динамика распространения аппроксимируется потоком по кратчайшим путям, что слои корректно агрегированы с учетом весов и что весовая структура отражает реальное влияние каналов передачи. Алгоритм менее применим в системах, где динамика имеет стохастическую или каскадную природу, с высокой долей латентных или нелокальных взаимодействий. Также эффективность может снижаться при наличии резких временных изменений или в графах с малой связностью.

## СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Newman M.E.J., Girvan M. Finding and evaluating community structure in networks. *Physical Review*. 2004;E69:026113. DOI: 10.48550/arXiv.cond-mat/0308217
2. Granovetter M.S. The Strength of Weak Ties. *American Journal of Sociology*. 1973;78(6):1360–1380. URL: <https://www.jstor.org/stable/2776392>
3. Watts D.J., Strogatz S.H. Collective Dynamics of “Small-world” Networks. *Nature*. 1998;393:440–442. DOI: 10.1038/30918

## ВЫВОДЫ

Предложен новый алгоритм, объединяющий спектральный коарснинг, приближенную оценку центральности и жадное покрытие путей для выявления критических ребер в многослойных графах ( $1 - 1/e$ ). Его отличительные черты: теоретическая обоснованность (доказаны свойства связности, аппроксимации и устойчивости); вычислительная эффективность (сублинейная сложность относительно числа ребер); поддержка многослойной структуры (учет весов слоев и кросс-слойных взаимодействий).

**Проведен формальный анализ свойств алгоритма**, включающий сохранение надпороговой связности остаточного графа; контроль роста эффективного диаметра; гарантированную аппроксимацию посреднической центральности; устойчивость к структурному шуму и вариациям весов.

**Выполнено сравнение с существующими методами** (спектральное разрежение, ABRA, edge salience), показывающее преимущество по масштабируемости и поддержке многослойности; баланс между точностью, скоростью и теоретической обоснованностью; явный фокус на критические ребра с глобальной оценкой их роли.

**Определены области практического применения**: моделирование распространения информации и эпидемий; оценка структурной уязвимости социальных, транспортных и коммуникационных сетей; прогнозирование каскадных сбоях и оптимизация устойчивости.

**Выявлены ограничения модели**: предположение о распространении по кратчайшим путям; априорная агрегация слоев без учета временной динамики; сниженная эффективность в сетях с малой связностью или стохастической динамикой.

**Намечены направления будущих исследований**: обобщение на динамические и темпоральные сети; интеграция вероятностных моделей распространения; эмпирическая валидация на реальных данных различной природы.

Таким образом, разработанный алгоритм представляет собой универсальный инструмент для анализа критических связей в сложных сетевых структурах, сочетающий теоретическую строгость, вычислительную эффективность и практическую применимость.

4. Girvan M., Newman M.E.J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the USA*. 2002;99(2):7821–7826. DOI: 10.1073/pnas.122653799
5. Newman M.E.J. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the USA*. 2006;103(23):8577–8582. DOI: 10.1073/pnas.0601602103
6. Kivelä M., Arenas A., Barthelemy M., Gleeson J.P., Moreno Y., Porter M.A. Multilayer networks. *Journal of Complex Networks*. 2014;2(3):203–271. DOI: 10.48550/arXiv.1309.7233; DOI: 10.1093/comnet/cnu016
7. Boccaletti S., Bianconi G., Criado R., del Genio C.I., Gómez-Gardeñes J., Romance M., Sendiña-Nadal I., Wang Z., Zanin M. The structure and dynamics of multilayer networks. *Physics Reports*. 2014;544(1):1–122. DOI: 10.1016/j.physrep.2014.07.001
8. Brandes U. A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology*. 2001;25(2):163–177. DOI: 10.1080/0022250X.2001.9990249
9. Spielman D.A., Teng S.-H. Spectral sparsification of graphs. *SIAM Journal on Computing*. 2011;40(4):981–1025. DOI: 10.1137/08074489X
10. Riondato M., Upfal E. ABRA: Approximating betweenness centrality in static and dynamic graphs with Rademacher averages. Proceedings of the 22nd ACM SIGKDD Conference. 2016;1145–1154. DOI: 10.1145/2939672.2939770
11. Jin Y., Loukas A., JáJá J. Graph coarsening with preserved spectral properties. Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics. 2020;4452–4462. URL: <https://proceedings.mlr.press/v108/jin20a/jin20a.pdf>
12. Watts D. J., Strogatz S.H. Collective dynamics of “small-world” networks. *Nature*. 1998;393(6684):440–442. DOI: 10.1038/3091
13. Song C., Havlin S., Makse H.A. Self-similarity of complex networks. *Nature*. 2005;433(7024):392–395. DOI: 10.1038/nature03248
14. Yassin A., Cherifi H., Seba H., Togni O. High and low similarity backbone extraction using similarity-based link prediction. *Applied Network Science*. 2025;10:22. DOI: 10.1007/s41109-025-00705-y
15. Burt R.S. Structural holes: The social structure of competition. — Cambridge, MA: Harvard Univ. Press; 1992. 313 p. URL: <https://archive.org/details/structuralholes0000burt/page/4/mode/2up>
16. Saito K., Kimura M., Motoda H. Discovering Influential Nodes for SIS Models in Social Networks. *Discovery Science*, Porto; 2009. DOI: 10.1007/978-3-642-04747-3\_24
17. De Domenico M., Solé-Ribalta A., Omodei E., Gómez S., Arenas A. Ranking in interconnected multilayer networks reveals versatile nodes. *Nature Communications*. 2015;6(1):6868. DOI: 10.1038/ncomms7868
18. Lotfi N., Requejo H.S., Rodrigues F.A., Mello M.A.R. A new centrality index designed for multilayer networks. *Methods in Ecology and Evolution*. 2023;15(1):204–213. DOI: 10.1111/2041-210X.14257
19. Leskovec J., Kleinberg J., Faloutsos C. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*. 2007;1(1):2. DOI: 10.1145/1217299.1217301
20. Zhao Z., Zhang Y., Feng Z. Towards scalable spectral embedding and data visualization via spectral coarsening. ACM International Conference on Web Search and Data Mining. 2021. DOI: 10.1145/3437963.3441767
21. Chen P.Y., Hero A.O. Multilayer spectral graph clustering via convex layer aggregation. *IEEE Transactions on Signal Processing*. 2017. URL: <https://ieeexplore.ieee.org/document/7990044>
22. Dong X., Frossard P., Vandergheynst P. Clustering with multi-layer graphs: A spectral perspective. *IEEE Transactions on Signal Processing*. 2012. URL: <https://ieeexplore.ieee.org/document/6265414/>
23. Solé-Ribalta A., De Domenico M., Gómez S. Centrality rankings in multiplex networks. ACM Web Science Conference. 2014. DOI: 10.1145/2615569.2615687
24. Chakraborty T., Narayanam R. Cross-layer betweenness centrality in multiplex networks with applications. IEEE 32nd International Conference on Data Engineering (ICDE). 2016. DOI: 10.1109/ICDE.2016.7498257
25. Miller B.A., Shafi Z., Ruml W., Vorobeychik Y. Attacking shortest paths by cutting edges. Proceedings of the ACM WebConference. 2023. DOI: 10.48550/arXiv.2211.11141
26. Lu G., Xiong Y., Ding C., Wang Y. An optimal schedule for urban road network repair based on the greedy algorithm. *PLoS ONE*. 2016;11(10). DOI: 10.1371/journal.pone.0164780
27. Glantz R., Meyerhenke H., Schulz C. Tree-based coarsening and partitioning of complex networks. *ACM Journal of Experimental Algorithmics*. 2016. DOI: 10.1145/2851496
28. Bellingeri M., Cassi D. Robustness of weighted networks under node and link removal. *Physica A Statistical Mechanics and its Applications* 489. 2017;500:11–22. DOI: 10.1016/j.physa.2017.07.020



29. Loukas A., Vandergheynst P. Spectrally approximating large graphs with smaller graphs. ICML. 2018. URL: <http://proceedings.mlr.press/v80/loukas18a.html>
30. Nemhauser G., Wolsey L. Maximizing submodular set functions. *Mathematical Programming*. 1978;14(1):265–294. URL: <https://thibaut.horel.org/submodularity/papers/nemhauser1978.pdf>
31. Panayiotou G., Magnani M., Pinaud B. Current challenges in multilayer network engineering. *Applied Network Science*. 2024;9(1). DOI: 10.1007/s41109-024-00686-4
32. Ghoniem M., Melançon G. The State of the Art in Multilayer Network Visualization. *Computer Graphics Forum*. 2019;38(16). DOI: 10.1111/cgf.13610

## ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

**Мария Вячеславовна Денисова** — магистрант, направление «Прикладная информатика», факультет информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация; BI-разработчик, ООО «Авито Тех», Москва, Российская Федерация

**Maria V. Denisova** — Master’s student, program “Applied Informatics”, Department of Artificial Intelligence, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation; BI Developer, Avito Tech, LLC, Moscow, Russia

<https://orcid.org/0009-0009-5719-1841>

[mardeni201101@mail.ru](mailto:mardeni201101@mail.ru)

**Расул Ахматович Кочкаров** — кандидат экономических наук, доцент кафедры искусственного интеллекта факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**Rasul A. Kochkarov** — Cand. Sci (Econ.), Assoc. Prof. of the Department of Artificial Intelligence, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

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

Автор для корреспонденции / Corresponding author:

[rkochkarov@fa.ru](mailto:rkochkarov@fa.ru)

### Заявленный вклад авторов:

**М.В. Денисова** — разработка общей концепции статьи, формализация алгоритма, написание текста.

**Р.А. Кочкаров** — научное руководство, методологическая поддержка, критический анализ результатов.

### Authors’ declared contributions:

**M.V. Denisova** — development of the general concept of the paper, algorithm formalisation, writing the manuscript.

**R.A. Kochkarov** — scientific supervision, methodological support.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

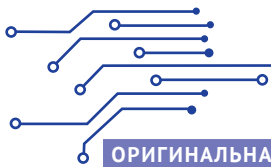
Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила 19.12.2025; после рецензирования 12.01.2026; принята к публикации 26.01.2026.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article was submitted on 19.12.2025; revised on 12.01.2026 and accepted for publication on 26.01.2026.

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

DOI: 10.26794/3033-7097-2026-2-1-28-34  
УДК 004.056.57(045)

# Недекларируемые возможности файловой архитектуры. Уязвимость формата PPTX Microsoft PowerPoint

А.А. Рыженко<sup>1</sup>, С.И. Козьминых<sup>2</sup><sup>1,2</sup> Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация;<sup>2</sup> Российский экономический университет им. Г.В. Плеханова, Москва, Российская Федерация

## АННОТАЦИЯ

Поскольку презентации прочно вошли в повседневную рабочую практику самых разных специалистов, форматы файлов для их создания неизбежно привлекают внимание злоумышленников — чем шире распространение, тем выше интерес к поиску уязвимостей. Что важно знать для эффективной защиты? **Предмет исследования** — уязвимость формата файлов PPTX (Microsoft PowerPoint), обусловленная его архитектурой как ZIP-контейнера с XML-компонентами. **Цель работы** — продемонстрировать возможность внедрения инъекций в PPTX-файлы без использования программирования, раскрыть механизмы эксплуатации и обозначить риски для информационной безопасности. **Основные аспекты анализа:** архитектурные особенности PPTX; механизм внешних ссылок через ActiveX-компоненты. Алгоритм инъекции: преобразование и модификация файла, маскировка подмены, восстановление исходного формата. Отражены практические риски, даны рекомендации по защите. **Научная и практическая значимость:** систематизация знаний о недеklarированных возможностях PPTX; демонстрация реального сценария атаки; формирование базы для разработки контрмер (DLP-правила, настройки Office). **Целевая аудитория:** специалисты по ИБ, разработчики офисного ПО, преподаватели курсов кибербезопасности, продвинутые пользователи. **Вывод:** в данной статье аналогично предыдущим сочетаются техническая глубина и доступность, подчеркивается необходимость комплексного подхода к защите офисных документов. Статья несет исключительно образовательные функции и предупреждения для действующих специалистов, не предполагая инструкции для нанесения вреда\* фирмам или организациям\*\*. **Ключевые слова:** инъекции; файловая система; презентации; безопасность; проникновение; глобальная сеть

**Для цитирования:** Рыженко А.А., Козьминых С.И. Недекларируемые возможности файловой архитектуры. Уязвимость формата PPTX Microsoft PowerPoint. *Цифровые решения и технологии искусственного интеллекта*. 2026;2(1):28-34. DOI: 10.26794/3033-7097-2026-2-1-28-34

## ORIGINAL PAPER

# Undeclared File Architecture Features. Vulnerability of Microsoft PowerPoint's PPTX Format

А.А. Ryzhenko<sup>1</sup>, S.I. Kozminych<sup>2</sup><sup>1,2</sup> Financial University under the Government of the Russian Federation, Moscow, Russian Federation;<sup>2</sup> Plekhanov Russian University of Economics, Moscow, Russian Federation

## ABSTRACT

Presentations have become an integral part of the core processes of many areas of professional activity. Nevertheless, the market for professional software packages that allow you to create presentations that meet modern requirements is not so saturated with competition. Microsoft's software products are the undisputed leader. Competitors are much inferior in quality. But, as mentioned in previous articles in this series, cross-platform file formats are not protected from simple injections (without programming knowledge). A detailed example for a word processor was demonstrated in the second part. In this article, the PPTX format will be analyzed in a similar scenario. A simple injection with a hidden link to the active element is a demonstration of the undeclared capabilities of this type of file formats. It should be remembered that these files can be run or opened on absolutely any (including mobile) smart device. This article, like the previous ones, does not provide instructions for harming\* firms or organizations\*\*. The articles have exclusively educational functions and warnings for current specialists. **Keywords:** injections; file system; presentations; security; penetration; global network

**For citation:** Ryzhenko A.A., Kozminych S.I. Undeclared file architecture features. Vulnerability of Microsoft PowerPoint's PPTX format. *Digital solutions and artificial intelligence technologies*. 2026;2(1):28-34. DOI: 10.26794/3030-7097-2026-2-1-28-34

\* Бирюков А. О преступлениях в ИТ простым языком. URL: <https://habr.com/ru/companies/otus/articles/804645/?ysclid=mlko2562y4238181887>

\*\* УК РФ, ст. 272. Неправомерный доступ к компьютерной информации; УК РФ, ст. 273. Создание, использование и распространение вредоносных компьютерных программ; УК РФ, ст. 274. Неправомерный доступ к компьютерной информации.

© Рыженко А.А., Козьминых С.И., 2026



## ВВЕДЕНИЕ

Важность обсуждения темы недекларируемых возможностей файловой архитектуры и уязвимостей формата PPTX обусловлена следующими факторами:

**1. Широкая распространенность формата.** PPTX — де-факто стандарт для презентаций в бизнесе, образовании и госуправлении; его используют миллионы людей ежедневно, что многократно увеличивает потенциальный масштаб угроз.

**2. Скрытость уязвимости.** Архитектура PPTX как ZIP-контейнера с XML-файлами позволяет внедрять вредоносный код без видимых признаков заражения, из-за чего атаки остаются незамеченными для большинства пользователей и систем защиты.

**3. Низкий порог эксплуатации.** Описанные инъекции реализуемы без программирования — достаточно базовых навыков работы с архивами и текстовыми редакторами, что делает атаку доступной даже для неопытных злоумышленников.

**4. Многовекторность угроз.** Уязвимость позволяет не только собирать данные через IP-логгеры, но и распространять фишинг, вредоносные скрипты или конфиденциальные payload-компоненты, встраиваемые прямо в структуру файла.

Презентации стали неотъемлемой частью основных процессов многих направлений профессиональной деятельности. Тем не менее рынок профессиональных пакетов программных продуктов, позволяющий создавать презентации, соответствующие современным требованиям, не так насыщен конкуренцией.

Несомненным лидером является программная продукция Microsoft. Конкуренты сильно уступают в качестве. Но, как было упомянуто в предыдущих статьях [1, 2], кросс-платформенные форматы файлов не защищены от простых инъекций (без знания программирования). В настоящей статье аналогично предыдущим сценариям будет разобран формат PPTX. Простая инъекция со скрытой ссылкой на активный элемент является демонстрацией недекларированных возможностей форматов файлов такого типа. При этом необходимо помнить, что данные файлы можно запустить или открыть на абсолютно любом (в том числе и мобильном) смартфоне. В данной статье, аналогично предыдущим, не предполагается инструкция для нанесения вреда<sup>1</sup> фирмам или организациям<sup>2</sup>. Статьи несут

образовательные функции и предупреждения для действующих специалистов.

Современный Microsoft Office способен сохранять файлы как в старом формате PPT, так и в новом формате PPTX. Как заражать старый формат, будет рассмотрено в последующих статьях. Современный кросс-платформенный расширенный формат, аналогично DOCX, представляет не закрытый файл, а ZIP-контейнер [1, 2]. При переименовании расширения PPTX в ZIP появляется возможность просмотра структуры файла. Используя данную возможность, можно создавать простые инъекции. В статье рассмотрен алгоритм простой инъекции, позволяющий изменять атрибуты активных компонентов презентации на примере YouTube<sup>3</sup> ролика<sup>4</sup>. Данная технология широко используется до сих пор на практике несмотря на то, что прямого доступа к видеороликам непосредственно из презентации практически невозможно. Так как ссылки заведомо ложные и являются редиректом IP-логгера<sup>5</sup>, сами видеоролики не запускаются.

## АЛГОРИТМ ИНЪЕКЦИИ В ПОПУЛЯРНЫЙ ФОРМАТ PPTX

Аналогично предыдущим инъекциям, создаем отдельную пустую папку для дальнейших экспериментов. Воспользуемся одним из самых популярных и пока доступных веб-ресурсом IP LOGGER<sup>6</sup>. Здесь сразу необходимо сделать небольшую поправку. Логгеры запоминают всю историю и хранят в log-файлах<sup>7</sup>. Пользоваться одним и тем же временным адресом логгера не рекомендуется. При обнаружении редирект-ссылки<sup>8</sup> браузеры заносят в черный список адрес и могут закрыть доступ к ресурсу.

Шаг 1. Заводим личный аккаунт в любом из IP-логгеров. Даже бесплатный одноразовый режим предоставляет доступ к достаточно большому количеству встроенных ресурсов (рис. 1).

и распространение вредоносных компьютерных программ; УК РФ, ст. 274. Неправомерный доступ к компьютерной информации.

<sup>3</sup> YouTube — интернет-видеоплатформа и социальная сеть. Запрещена на территории Российской Федерации.

<sup>4</sup> Эксперты ответили на сообщения о полной блокировке YouTube в России. URL: [https://www.rbc.ru/technology\\_and\\_media/10/02/2026/698b42499a79471c03a2bb74](https://www.rbc.ru/technology_and_media/10/02/2026/698b42499a79471c03a2bb74)

<sup>5</sup> Лучшие IP Logger Альтернативы. URL: <https://apkpure.net/ru/similar/com.linkslogger.iplogger>

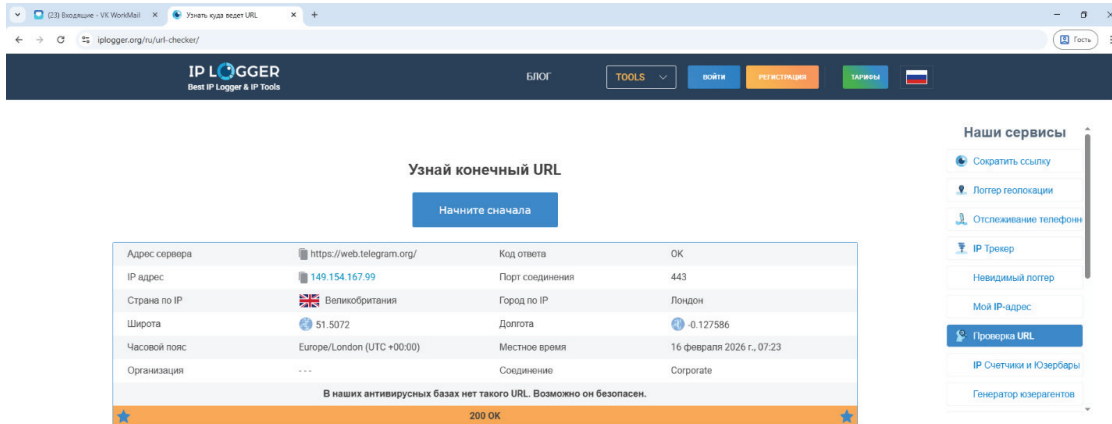
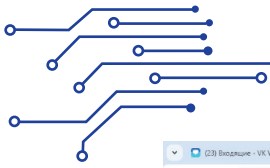
<sup>6</sup> IP Logger. Узнать чужой IP. URL: <https://iplogger.org/ru/>

<sup>7</sup> Как скрыться от IP логгеров? URL: <https://www.securitylab.ru/blog/personal/xiaomite-journal/355293.php>

<sup>8</sup> Что такое редирект. URL: <https://help.reg.ru/support/hosting/redirekty/chto-takoye-redirekt#0>

<sup>1</sup> Бирюков А. О преступлениях в ИТ простым языком. URL: <https://habr.com/ru/companies/otus/articles/804645/?ysclid=mlko2562y4238181887>

<sup>2</sup> УК РФ, ст. 272. Неправомерный доступ к компьютерной информации; УК РФ, ст. 273. Создание, использование



### О проверке URL

С помощью URL-чекера вы можете безопасно проверить ссылку и заранее узнать куда она ведет, обезопасив себя от возможных проблем. А перейдя сразу на конечную ссылку, вы не оставите следов при переходах. Теперь гораздо проще бороться с нелегальными спамерами и хакерами.

Рис. 1 / Fig. 1. Примеры IP Logger / IP Logger Examples

Источник / Source: открытая зона интернета / Open Internet Area.

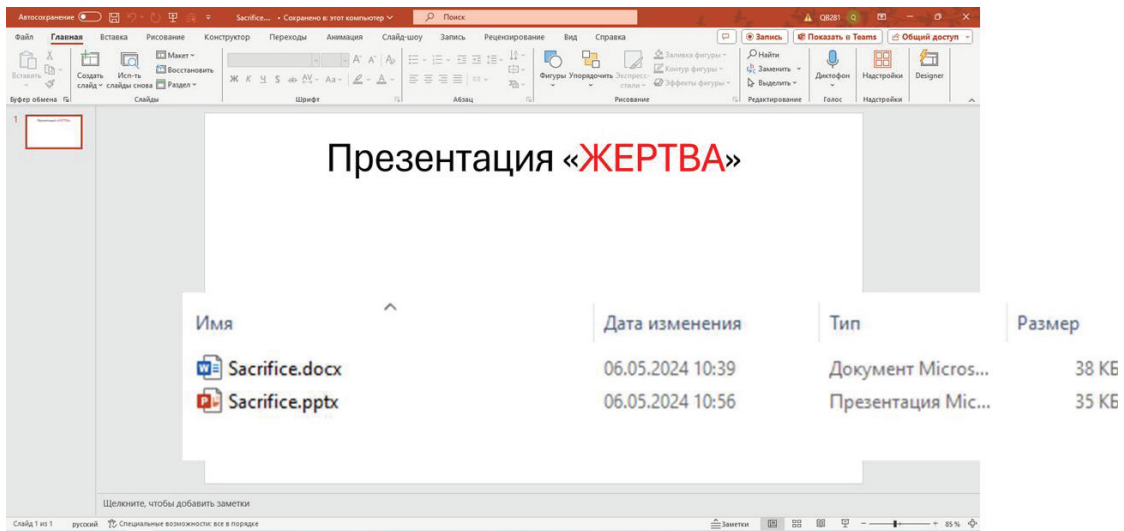


Рис. 2 / Fig. 2. Подготовка файла для инъекции / Preparing the File for Injection

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

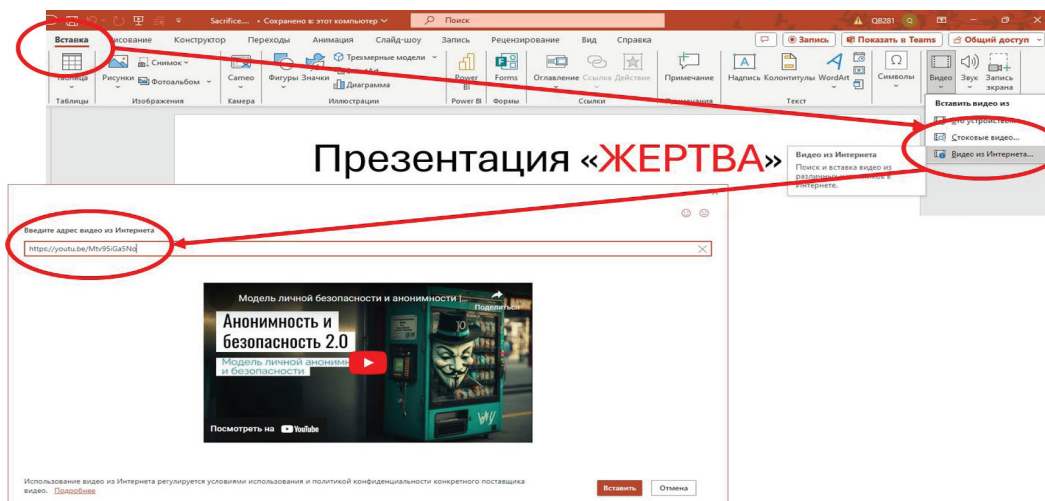


Рис. 3 / Fig. 3. Добавление ActiveX элемента / Adding an ActiveX Control

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

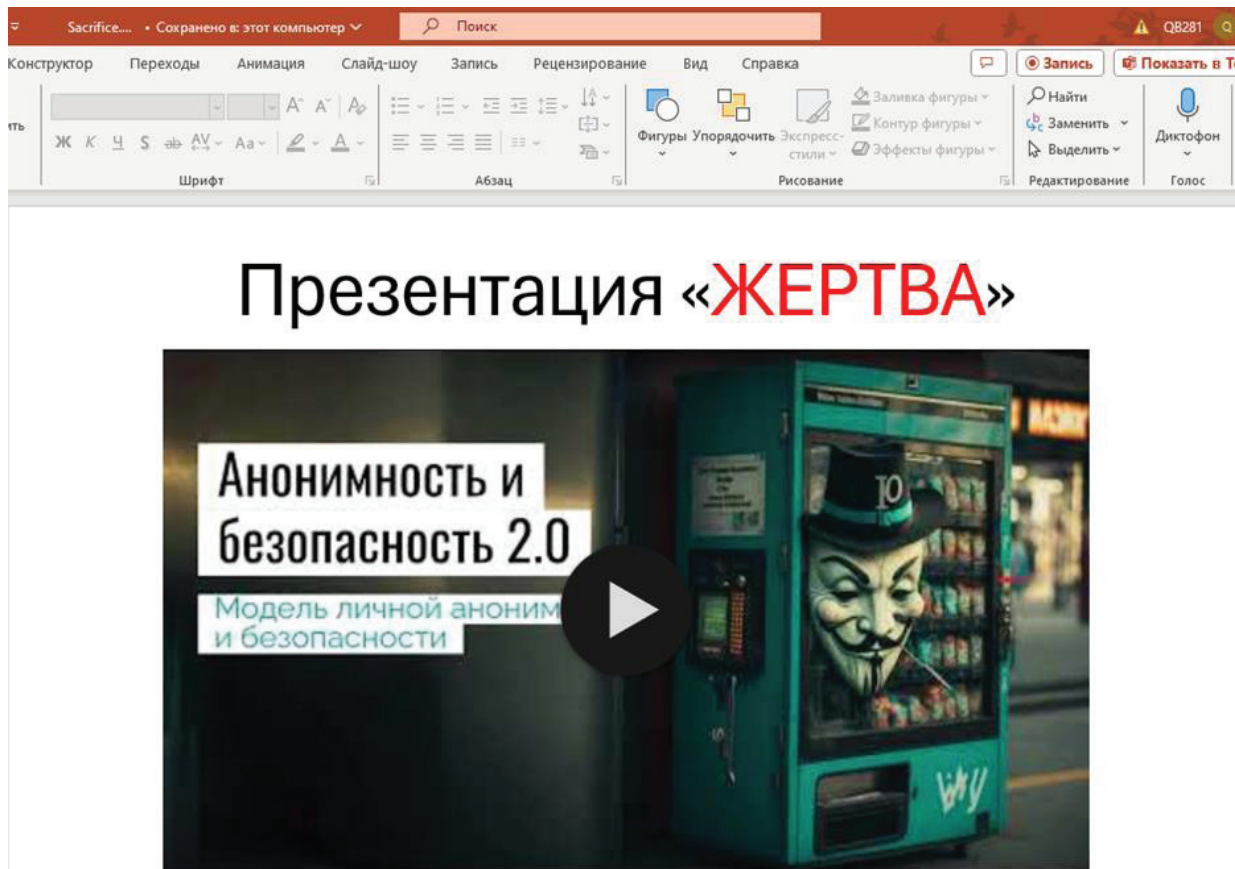
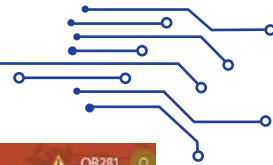


Рис. 4 / Fig. 4. Файл с активным элементом / The File with the Active Element

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

На втором шаге необходимо создать или использовать готовый.pptx файл, в который будем внедрять инъекцию. Можно использовать пустой файл без наполнения контентом. Интерес именно в файловой информации.

Шаг 2. Создаем документ жертву в формате.pptx (рис. 2)<sup>9</sup>.

Далее необходимо добавить активный элемент ActiveX<sup>10</sup> на слайд презентации. Открываем закладку «Вставка», находим группу «Видео».

Шаг 3. Добавляем в документ ActiveX компонент «Видео из интернета» (рис. 3).

При подключении используем произвольную ссылку на видеоролик YouTube. Попытки подключить видео с других каналов, таких как VK Video<sup>11</sup>, Yandex, RuTube<sup>12</sup> и т.д., потерпели неудачу. Данный

<sup>9</sup> Extract files or objects from a PowerPoint file. URL: <https://support.microsoft.com/en-us/office/extract-files-or-objects-from-a-powerpoint-file-85511e6f-9e76-41ad-8424-eab8a5bbc517>

<sup>10</sup> Включение и отключение параметров элементов ActiveX в файлах Office. URL: <https://support.microsoft.com/ru-ru/office/включение-и-отключение-параметров-элементов-office-f1303e08-a3f8-41c5-a17e-b0b8898743ed>

<sup>11</sup> URL: <https://vk.com/vkvideo>

<sup>12</sup> URL: <https://rutube.ru/>

элемент использует только один глобальный канал потоков видео данных.

Шаг 4. Просто сохраняем файл с активной вставкой (рис. 4).

Шаг 5. В папке с файлом открываем видимость расширений файлов. Меняем расширение.pptx на.zip (рис. 5).

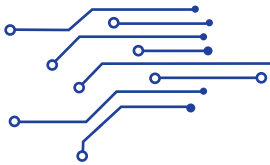
Если открыть полученный архив как папку, получим доступ к ряду файлов и папок. Основное расширение файлов атрибутов контента слайдов.xml. Дополнительные атрибуты неосновных элементов слайда хранятся в резервной папке Rels.

Шаг 6. Открываем архив Sacrifice.zip. Открываем папку ppt, затем slides. Открываем папку \_rels. Находим файл slide1.xml.rels (рис. 6).

Шаг 7. Находим в файле ссылку на Youtube-ролик и меняем на ссылку IP-ЛОГГЕР. Файл готов! (рис. 7).

Никаких нюансов здесь нет. Идет простая подмена ссылки. Можно ли скрыть подмену? Можно, воспользовавшись встроенной функцией «Подсказка». Если пользователь не выберет «Изменить ссылку», он никогда не узнает, что была подмена.

Шаг 8. Переименовываем файл обратно.zip в.pptx. Запускаем (рис. 8).



Имя	Дата изменения	Тип	Размер
Sacrifice.docx	06.05.2024 10:39	Документ Micros...	38 КБ
Sacrifice.pptx	06.05.2024 10:56	Презентация Мис...	35 КБ

Имя	Дата изменения	Тип	Размер
Sacrifice.docx	06.05.2024 10:39	Документ Micros...	38 КБ
Sacrifice.zip	06.05.2024 11:04	Архив ZIP - WinR...	151 КБ

Рис. 5 / Fig. 5. Меняем расширения презентации на расширение архиватора / Changing the Presentation Extensions to the Archiver Extension

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

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Relationships xmlns="http://schemas.openxmlformats.org/package/2006/relationships"><Relationship Id="rId3"
Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/image" Target="../media/image1.jpeg"/><Relationship Id="rId2"
Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/slideLayout" Target="../slideLayouts/slideLayout1.xml"/>
<Relationship Id="rId1" Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/video"
Target="https://www.youtube.com/embed/Mtv95iGa5No?feature=oembed" TargetMode="External"/></Relationships>
```

Рис. 6 / Fig. 6. Содержимое файла slide1.xml.rels / Contents of the slide1.xml.rels File

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

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Relationships xmlns="http://schemas.openxmlformats.org/package/2006/relationships"><Relationship Id="rId3"
Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/image" Target="../media/image1.jpeg"/><Relationship Id="rId2"
Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/slideLayout" Target="../slideLayouts/slideLayout1.xml"/>
<Relationship Id="rId1" Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/video"
Target="https://www.youtube.com/embed/Mtv95iGa5No?feature=oembed" TargetMode="External"/></Relationships>
```

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Relationships xmlns="http://schemas.openxmlformats.org/package/2006/relationships"><Relationship Id="rId3"
Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/image" Target="../media/image1.jpeg"/><Relationship Id="rId2"
Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/slideLayout" Target="../slideLayouts/slideLayout1.xml"/>
<Relationship Id="rId1" Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/video"
Target="https://iplogger.com/2Mrdq3" TargetMode="External"/></Relationships>
```

Рис. 7 / Fig. 7. Подмена ссылки на редирект / Substitution of the Redirect Link

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

На видеоролике нажимаем кнопку запуска PLAY. Система предложит запустить видео во внешнем браузере. Если пользователь согласится, то произойдет редирект IP-логгера на соответствующий ролик и система не «расскажет» про поддельную ссылку.

Аналогично предыдущим статьям, следует также обратить внимание на тот факт, что все манипуляции с целостностью файла.pptx не вызвали подозрения ни у одной установленной системы защиты информации.

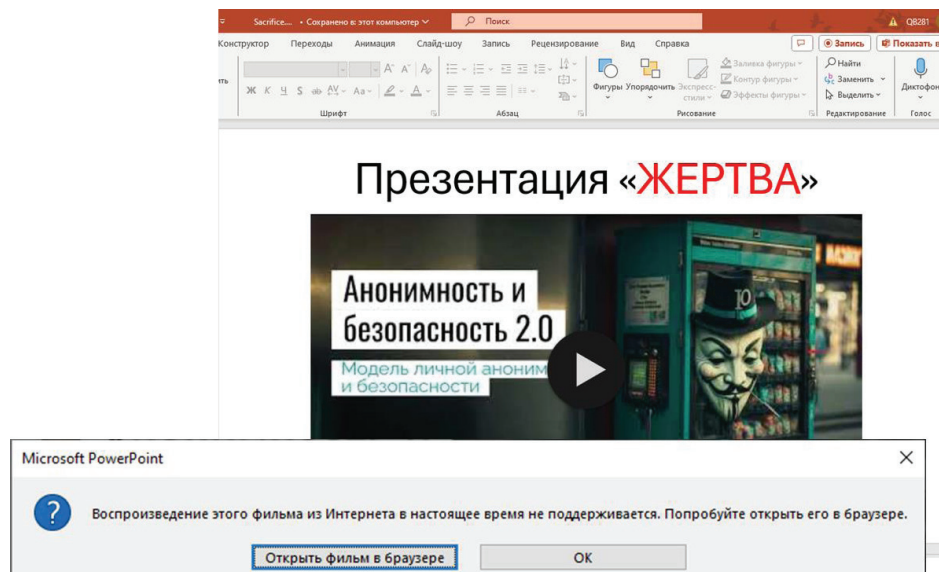
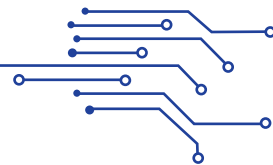


Рис. 8 / Fig. 8. Запуск редирект ссылки / Launching a Redirect Link

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

## ВЫВОДЫ

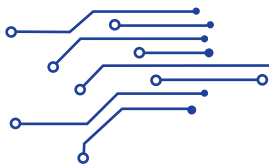
В статье рассмотрен еще один механизм простой инъекции в популярный формат файлов<sup>13</sup> авто-

<sup>13</sup> Спрятать файл внутри другого файла без потери работоспособности обоих. URL: <https://programmersforum.ru/showthread.php?t=48995>; Как спрятать файл внутри изображения. URL: <https://ru.wikihow.com/спрятать-файл-внутри-изображения>

матизированного офиса. Забегая вперед, можно отметить, что это не единственный метод простой инъекции [3–8]. В следующих статьях будут рассмотрены «игры в прятки» непосредственно на файловом уровне. В саму структуру файла презентации можно добавлять произвольные ресурсы, несущие в себе вирусы, послания и прочую чуждую информацию.

## СПИСОК ИСТОЧНИКОВ

1. Рыженко А.А., Козьминых С.И. Недекларируемые возможности файловой архитектуры: графические контейнеры. *Цифровые решения и технологии искусственного интеллекта*. 2025;1(3):55-61. DOI: 10.26794/3033-7097-2025-1-3-55-61
2. Рыженко А.А., Козьминых С.И. Недекларируемые возможности файловой архитектуры: Автоматизированный офис, текстовый редактор, формат Doc X. *Цифровые решения и технологии искусственного интеллекта*. 2025;1(4):60-68. DOI: 10.26794/3033-7097-2025-1-4-60-68
3. Petrosyan M.A., Usenko A.S. The concept and elements of criminalistics characteristics of unauthorized access to computer information. *Epomen. Global*. 2023;43:123-140. URL: <https://elibrary.ru/mfcqwu>
4. Елизаров М.М. Выявление уязвимостей и недекларированных возможностей в программном обеспечении. *Интернаука*. 2024;21-1(338):60-61. URL: <https://elibrary.ru/DLCUNL>
5. Бегаев А.Н., Кашин С.В., Маркевич Н.А., Марченко А.А. Выявление уязвимостей и недекларированных возможностей в программном обеспечении. СПб.: Университет ИТМО; 2020. 38 с. URL: <https://elibrary.ru/xfaznp>
6. Маркин Д.О., Макеев С.М., Санников И.А., Чунг Х.Т. Методика исследования системного программного обеспечения сетевого оборудования семейства Cisco на предмет наличия недекларируемых возможностей. *Ученые записки Орловского государственного университета*. 2020;3(88):215-221. URL: <https://elibrary.ru/xyjmtf>
7. Markin D.O., Makeev S.M., Ho Thai Trung. Security threat level estimation for untrusted software based on TrustZone technology. *Proceedings of the Institute for System Programming of RAS*. 2022;34(1):35-48. DOI: 10.15514/ISPRAS-2022-34(1)-3
8. Жук Р.В. Методика и алгоритмы определения актуальных угроз информационной безопасности в информационных системах персональных данных. Дис. ... канд. техн. наук. СПб.: Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича; 2021. URL: <https://elibrary.ru/prxzhx>



## REFERENCES

1. Ryzhenko A.A., Kozminykh S.I. Undeclared file architecture features: graphical containers. *Digital solutions and artificial intelligence technologies*. 2025;1(3):55-61. (In Russ.). DOI: 10.26794/3033-7097-2025-1-3-55-61
2. Ryzhenko A.A., Kozminykh S.I. Undeclared file architecture features: Automated office, text editor, DocX format. *Digital solutions and artificial intelligence technologies*. 2025;1(4):60-68. (In Russ.). DOI: 10.26794/3033-7097-2025-1-4-60-68
3. Petrosyan M.A., Usenko A.S. The concept and elements of criminalistics characteristics of unauthorized access to computer information. *Epomen. Global*. 2023;43:123-140. URL: <https://elibrary.ru/mfcqwu>
4. Elizarov M.M. Identification of vulnerabilities and undeclared features in software. *Internauka*. 2024;21-1(338):60-61. URL: <https://elibrary.ru/DLCUNL> (In Russ.).
5. Begaev A.N., Kashin S.V., Markevich N.A., Marchenko A.A. Identification of vulnerabilities and undeclared features in software. Saint Petersburg: ITMO University; 2020. 38 p. URL: <https://elibrary.ru/xfaznp> (In Russ.).
6. Markin D.O., Makeev S.M., Sannikov I.A., Chung H.T. Methodology for investigating the system software of Cisco network equipment for undeclared capabilities. *Scientific notes of the Orel State University*. 2020;3(88):215-221. URL: <https://elibrary.ru/xyjmtf> (In Russ.).
7. Markin D.O., Makeev S.M., Ho Thai Trung. Security threat level estimation for untrusted software based on TrustZone technology. *Proceedings of the Institute for System Programming of RAS*. 2022;34(1):35-48. DOI: 10.15514/ISPRAS-2022-34(1)-3
8. Zhuk R.V. Methodology and algorithms for determining actual threats to information security in personal data information systems. Cand. Sci. (Tech.) thesis. St. Petersburg: St. Petersburg State University of Telecommunications named after Prof. M.A. Bonch Bruevich, 2021. URL: <https://elibrary.ru/prxzhx> (In Russ.).

## ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

**Алексей Алексеевич Рыженко** — кандидат технических наук, доцент, доцент кафедры информационной безопасности, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**Alexey A. Ryzhenko** — Cand. Sci. (Tech.), Assoc. Prof. of the Information Safety Department, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0000-0002-7279-9929>

Автор для корреспонденции / Corresponding author:

[AAryzhenko@fa.ru](mailto:AAryzhenko@fa.ru)

**Сергей Игоревич Козьминых** — доктор технических наук, доцент, профессор кафедры прикладной информатики и информационной безопасности, Российский экономический университет им. Г.В. Плеханова, Москва, Российская Федерация; профессор кафедры информационной безопасности, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**Sergey I. Kozminykh** — Dr. Sci. (Tech.), Assoc. Prof., Prof. of the Department of Applied Informatics and Information Security, Plekhanov Russian University of Economics, Moscow, Russian Federation; Professor of Information Security Department, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0000-0003-3903-9562>

[kozminykh.si@rea.ru](mailto:kozminykh.si@rea.ru); [SIKozminykh@fa.ru](mailto:SIKozminykh@fa.ru)

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

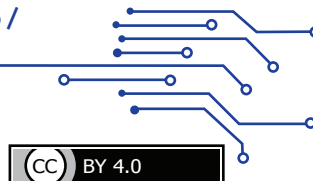
Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила 19.12.2025; после рецензирования 12.01.2026; принята к публикации 26.01.2026.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article was submitted on 19.12.2025; revised on 12.01.2026 and accepted for publication on 26.01.2026.

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



ОРИГИНАЛЬНАЯ СТАТЬЯ

DOI: 10.26794/3030-7097-2026-2-1-35-44  
УДК 004.56(045)

# Обзор инструментов для тестирования на проникновение: сравнение функционала и удобства использования

А.И. Любимов, С.А. Резниченко

Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

## АННОТАЦИЯ

В статье представлен детальный анализ ключевых инструментов пентестинга (Metasploit, Core Impact, Immunity Canvas, Security Forest и др.), востребованных в сфере информационной безопасности. **Цель исследования** – сравнить функциональность, удобство использования и специфику перечисленных решений, чтобы помочь специалистам выбрать оптимальный инструмент в зависимости от задач, бюджета и уровня квалификации. В работе раскрыта суть и значимость тестирования на проникновение (пентестинга) в контексте современной кибербезопасности; систематизированы этапы проведения пентеста (сбор информации, выявление уязвимостей, планирование атак, анализ результатов); приведены сравнительные характеристики инструментов в виде таблиц (поддерживаемые ОС, лицензии, интерфейсы, стоимость, языки программирования, особенности обновлений и отчетности); визуализированы процессы работы инструментов с помощью UML-диаграмм; затронуты этические и правовые аспекты применения пентестинга. **Основные выводы:** пентестинг – не разовая мера, а неотъемлемый элемент жизненного цикла системы безопасности, требующий регулярного применения; не существует универсального инструмента: автоматизированные сканеры (например, Nmap, Nessus) эффективны для первичного выявления уязвимостей, а фреймворки эксплуатации (Metasploit, Core Impact) – для подтверждения серьезности уязвимостей и демонстрации реального ущерба; коммерческие продукты предлагают глубину анализа и автоматизацию за высокую стоимость, а решения с открытым исходным кодом – гибкость при большей требовательности к квалификации специалиста; наилучшая результативность достигается при комбинировании автоматизированных и ручных методов проверки; решающий фактор успеха – квалификация и аналитическое мышление специалиста, действующего в правовых и этических рамках. Статья полезна специалистам по информационной безопасности, а также тем, кто изучает методы защиты данных и тестирования на проникновение.

**Ключевые слова:** тестирование на проникновение; пентестинг; информационная безопасность; уязвимость; инструменты тестирования; кибербезопасность; защита данных; хакерские атаки; безопасность сети; системы проникновения

**Для цитирования:** Любимов А.И., Резниченко С.А. Обзор инструментов для тестирования на проникновение: сравнение функционала и удобства использования. *Цифровые решения и технологии искусственного интеллекта*. 2026;2(1):35-44. DOI: 10.26794/3030-7097-2026-2-1-35-44

ORIGINAL PAPER

# Overview of Penetration Testing Tools: Comparison of Functionality and Ease of Use

A.I. Lyubimov, S.A. Reznichenko

Financial University under the Government of the Russian Federation, Moscow, Russian Federation

## ABSTRACT

This study provides a detailed analysis of penetration testing tools such as Metasploit, Core Impact, Immunity Canvas, and Security Forest. It compares their functionality, usability, and role in identifying vulnerabilities in information systems. The main focus is on the features of each tool, their strengths and weaknesses, and their areas of application. Penetration testing is an important element of a comprehensive approach to cybersecurity, allowing vulnerabilities to be identified at all stages of the information system lifecycle. The paper examines the stages of penetration testing, including information gathering, vulnerability identification, attack planning, and results analysis. Particular attention is paid to automated tools, which greatly simplify the testing process but require competent use. The article also discusses the ethical and legal aspects of penetration testing, emphasizing the need to comply with legislation and professional ethics. The article will be useful for information security specialists, as well as anyone interested in modern data protection methods. The paper emphasizes that the choice of tool depends on specific tasks and context, and that successful pentesting requires not only technical skills, but also a deep understanding of information protection processes.

© Любимов А.И., Резниченко С.А., 2026

**Keywords:** penetration testing; pentesting; information security; vulnerability; testing tools; cybersecurity; data protection; hacker attacks; network security; intrusion systems

**For citation:** Lyubimov A.I., Reznichenko S.A. Overview of penetration testing tools: Comparison of functionality and ease of use. *Digital solutions and artificial intelligence technologies*. 2026;2(1):35-44. DOI: 10.26794/3030-7097-2026-2-1-35-44

## ВВЕДЕНИЕ

Современный мир стремительно трансформируется, углубляясь в эпоху цифровых технологий, в которой огромное количество данных ежедневно создается, обрабатывается и хранится. С этой беспрецедентной цифровизацией возрастает и необходимость защищать наши данные от угроз, которые становятся сложнее, изощреннее и, увы, все более распространенными. Вопрос кибербезопасности сегодня поднят на совершенно новый уровень и требует не просто базовых мер предосторожности, а системного подхода с применением продвинутых методов и технологий.

Когда мы размышляем о защите информации, мы чаще всего обращаем внимание на обычные средства безопасности, такие как антивирусы, брандмауэры или сложные пароли. Однако все эти инструменты — всего лишь один из уровней многослойной системы обороны. Гораздо важнее выявить потенциальные слабые места самой структуры сети или программного обеспечения еще до того, как злонамеренные хакеры смогут воспользоваться ими. Именно здесь на помощь приходят методы тестирования на проникновение, или, как их называют профессионалы, пентесты. Это процесс, который представляет собой своего рода «управляемую атаку», проводимую с целью выявления уязвимостей системы безопасности.

Почему это приобретает столь высокую значимость? Потому что нельзя защитить то, что ты не знаешь, как атаковать. Любая система, остающаяся неизученной, становится идеальной мишенью для злоумышленников. Они постоянно совершенствуют свои стратегии, находят лазейки, используют уязвимости и эксплуатируют ошибки кода. Без тестирования на проникновение компания рискует потерять не только данные, но и доверие клиентов, средства и репутацию.

Этот процесс помогает организациям понять степень риска, которому они подвергаются, и в последующем принять соответствующие меры. Однако важно понимать, что пентест — это не единичное решение всех проблем. Это лишь элемент целого набора тактик, который должен работать в связке с другими средствами безопасности. Это подобно тренировочному полигону: чем больше вы знаете о своих слабых местах, тем эффективнее сможете отразить реальную атаку.

На сегодняшний день рынок изобилует инструментами для тестирования на проникновение,

которые способны решать самые разнообразные задачи — от анализа корпоративных сетей до оценки защищенности веб-приложений. В статье мы рассмотрим основные продукты, которые давно зарекомендовали себя как надежные союзники профессионалов в области кибербезопасности. Каждый из этих инструментов играет уникальную роль, дополняя общий комплекс мер по обеспечению защиты.

Одной из основных задач тестирования на проникновение является выявление уязвимостей до того, как это сделает злоумышленник, и их быстрое устранение. Злоумышленники используют различные автоматизированные программные средства и проводят сетевые атаки, чтобы получить доступ к системам.

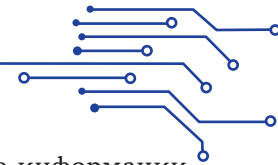
Благодаря тестированию на проникновение менеджеры могут получить представление о том, насколько защищена их сеть с точки зрения злоумышленника. Целью тестирования на проникновение является поиск методов обхода сетевых уязвимостей и их устранение до того, как хакер обнаружит их и воспользуется ими.

Во-вторых, даже зная о рисках, пользователи и сетевые администраторы могут нуждаться в отчете о тестировании на проникновение. С его помощью они способны убедить руководство в необходимости финансирования устранения проблем.

В-третьих, для подтверждения того, что наши сети настроены надежно, может быть проведено тестирование на проникновение. Отчет о тестировании на проникновение помогает проверить эффективность работы сотрудников службы безопасности нашей организации. Этот тест не повышает безопасность системы или сети; вместо этого он выявляет различия между предполагаемой и фактической реализацией.

В-четвертых, это помогает компаниям соблюдать законодательные требования, установленные правительством для ведения бизнеса.

Наконец, тестирование на проникновение используется для оценки новой технологии. Новая технология должна быть оценена перед запуском в производство. Тестирование на проникновение является простым и доступным инструментом, поскольку ни одна новая технология не может считаться полностью надежной без предварительной проверки.



## ЭТАПЫ И ИНСТРУМЕНТЫ ТЕСТИРОВАНИЯ НА ПРОНИКНОВЕНИЕ

Тестирование на проникновение, или пентест, представляет собой процесс оценки защищенности компьютерных систем и сетей путем имитации атак, потенциально возможных на данном объекте. В век цифровой трансформации и увеличивающегося числа кибератак такой подход становится не просто значимым инструментом в арсенале специалистов по безопасности, а необходимостью. Однако успешное проведение подобных тестов невозможно без специализированных инструментов, которые предоставляют средства для анализа, эксплуатации уязвимостей и разработки отчетов. В этой статье мы проведем обзор наиболее популярных инструментов для пентестинга, их функциональности, а также рассмотрим аспекты их использования.

Определенный набор входных данных, процедур и результатов называется методологией. Она дает нам инструкции о том, как перейти от входных данных к выходным.

На первом этапе мы выбираем, какая система, сеть или онлайн-приложение будут подвергнуты тестированию на проникновение.

Профиль злоумышленника, который будет использовать тестировщик, и продолжительность теста — это два аспекта, которые определяют масштаб теста.

Сбор информации — это следующий шаг. Как следует из названия, он предоставляет тестировщику информацию о различных целевых сетях. Наша цель — собрать как можно больше информации о нашей целевой сети, доступной для широкой публики. Для диагностики сети мы применяем такие инструменты, как ping, traceroute и ipconfig.

Выявление уязвимостей — это третий шаг [3]. Его основная цель — обнаружение слабых мест в целевой системе или сети, которые могут быть использованы злоумышленниками. Для этого применяются как стандартные процедуры, так и автоматизированные средства.

Тестировщик может вручную выявить распространенные ошибки настройки и дефекты в целевой сети или хосте с помощью ручного обнаружения уязвимостей. Сканеры уязвимостей — это коммерческие и бесплатные инструменты, доступные в рамках автоматизированной процедуры. Они позволяют проверить программное обеспечение целевой сети на наличие слабых мест. Такие сканеры выявляют уязвимости, потенциально пригодные для взлома наших систем и сетей.

Эти автоматизированные программы могут только выявлять уязвимости; тестирование на проникновение — это не то, что они могут сделать.

На четвертом этапе — анализе информации и планировании атаки — собираются данные, полученные на предыдущих этапах. Специалист по тестированию на проникновение может спланировать нападение на объект с помощью этой технической и общедоступной информации. Специалист по тестированию также определяет, какой объект нуждается в дальнейшем изучении.

Существует две составляющие нападения и проникновения [4]. Атака и проникновение составляют первую фазу. Следует попытаться использовать слабые места, выявленные в процессе анализа уязвимостей.

Самый распространённый способ, который применяют многие сканеры для сопоставления IP-адреса с сетевым хостом, — это ICMP-эхо-запрос (ping). Такие инструменты, как Nessus, могут задействовать пакеты TCP или UDP, чтобы определить активность хоста.

Опираясь на этот механизм, мы используем средства эксплуатации [5] для подтверждения наличия уязвимости. На рынке представлено большое число решений — как коммерческих, так и с открытым исходным кодом. Специалисты по тестированию на проникновение и злоумышленники зачастую применяют идентичный инструментарий. Перечисленные средства эксплуатации — лишь небольшая часть доступного арсенала. Большинство инструментов этой категории рассчитаны на однократное применение.

Как видно из *табл. 1*, инструменты первичного сканирования образуют класс доступного и специализированного ПО. Их ключевые различия лежат в плоскости поддерживаемых ОС и наличия узкоспециализированных функций (например, анализ логов в Unicornscan), что определяет выбор тестировщика в зависимости от конкретного сценария разведки.

В 2003 г. HD Moore разработал платформу Metasploit с открытым исходным кодом [6]. С ее помощью можно проводить исследования эксплойтов, тестирование на проникновение, создание сигнатур IDS, а также исследования и разработку уязвимостей.

Перед использованием эксплойта пользователь выбирает полезную нагрузку. Он может воспользоваться целевой удаленной службой. Ее можно использовать в сочетании со сценарием Meterpreter для управления программами, запускающими бэкдор. Платформа Metasploit отлично подходит для создания сценариев и проведения тестов на проникновение, предоставляя инструменты безопасности и эксплойты.

Представленная на *рис. 1* диаграмма иллюстрирует модульную архитектуру Metasploit, где



Инструменты для сканирования и обнаружения уязвимостей / Tools for Scanning and Detecting Vulnerabilities

Название	Назначение	ОС	Лицензия
Nmap	Сканирование сети, сканирование портов	Linux, Windows	Бесплатно
SuperScan	Обнаружение открытых TCP/UDP портов	Linux, Windows	Бесплатно
Hping	Сканирование портов, удаленное определение ОС	Windows	Бесплатно
AngryIP	Сканирование TCP, Открытие хостов	Windows	Бесплатно
Unicornscur	Сканер TCP/UDP портов, Анализ логов PCAP	Unix/Linux	Бесплатно
Advanced Port Scanner	Сканирование портов TCP, Многопоточность	Windows	Бесплатно

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

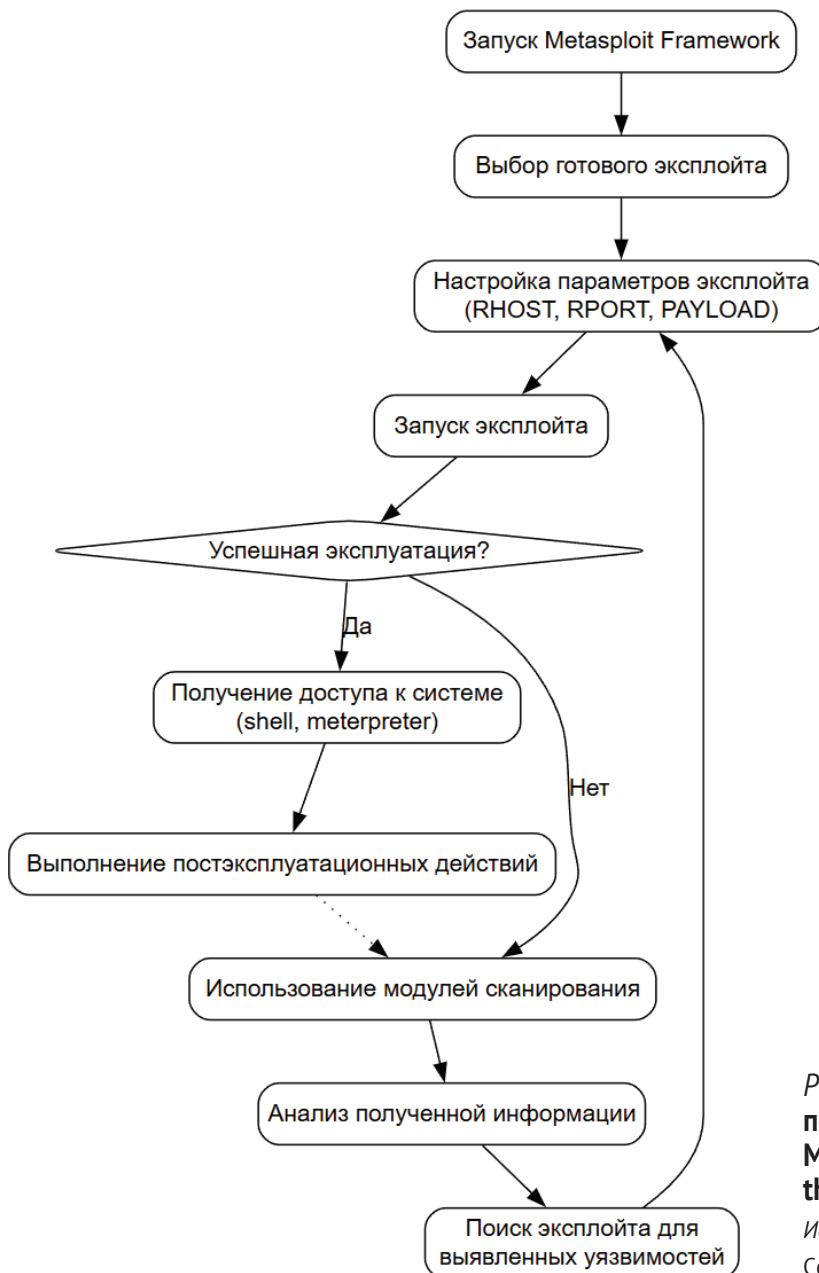


Рис. 1 / Fig. 1. UML – диаграмма, показывающая процесс работы Metasploit / Is a UML Diagram Showing the Metasploit Process

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

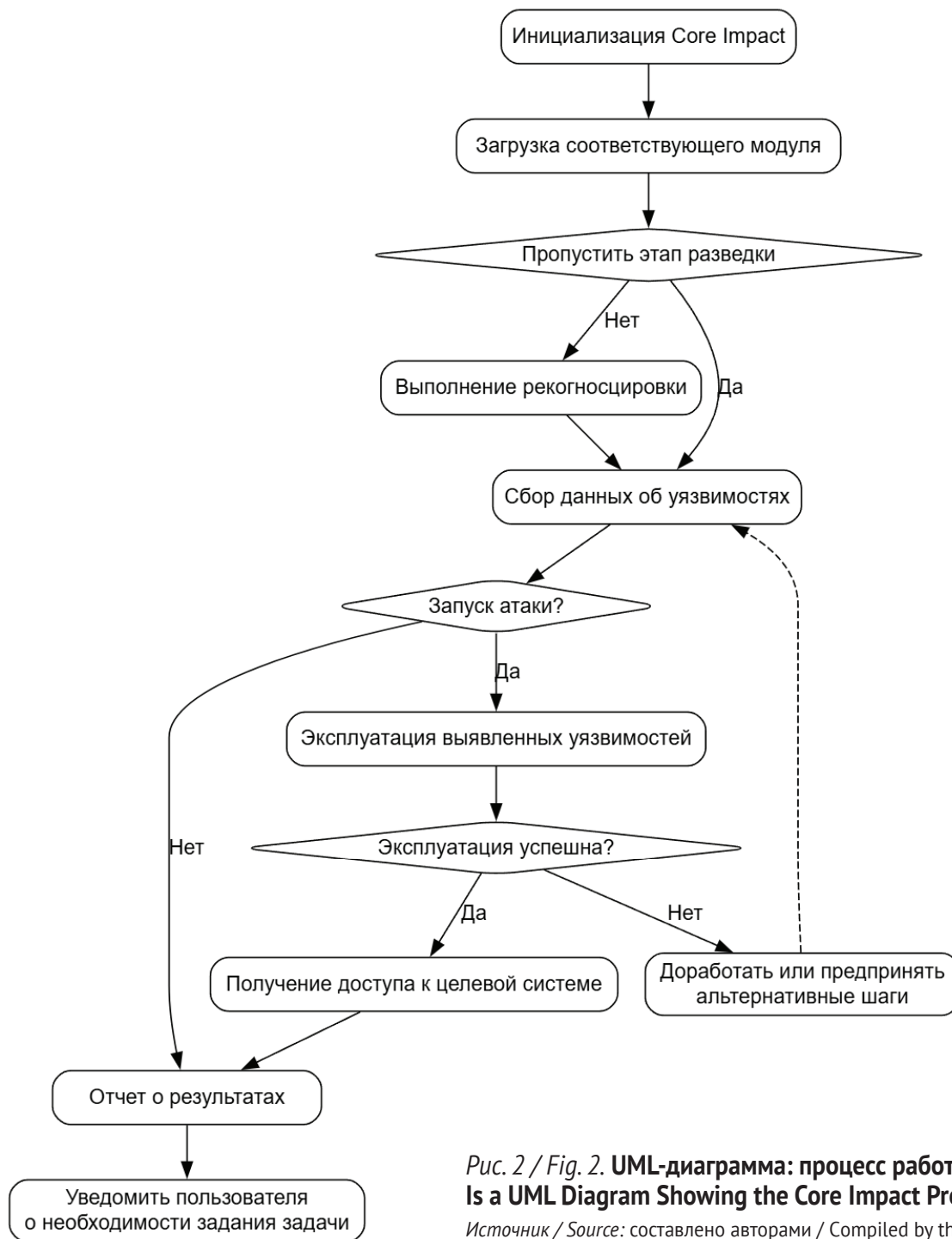
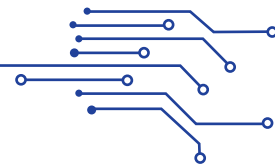


Рис. 2 / Fig. 2. UML-диаграмма: процесс работы Core Impact / Is a UML Diagram Showing the Core Impact Process

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

процесс эксплуатации реализован как последовательность строго детерминированных этапов. Такая структура обеспечивает гибкость при выборе вектора атаки и объясняет широкую распространенность фреймворка в качестве базового инструмента для идентификации уязвимостей.

Для платформы Metasploit доступны как база команд, так и графический интерфейс пользователя. Ее версия с графическим интерфейсом называется Armitage.

Core Impact — это коммерческая платформа компании Core Security Technologies [2]. С по-

мощью Core Impact Pro можно оценить степень безопасности и методы многовекторного тестирования для сетевых, онлайн-, мобильных и беспроводных сред.

Схема работы Core Impact (рис. 2) демонстрирует интегрированный подход, при котором этапы сканирования, эксплуатации и постэксплуатации объединены в единый автоматизированный контур. Это позволяет не только обнаруживать уязвимости, но и моделировать цепочки атак, оценивая реальный ущерб для информационной инфраструктуры.

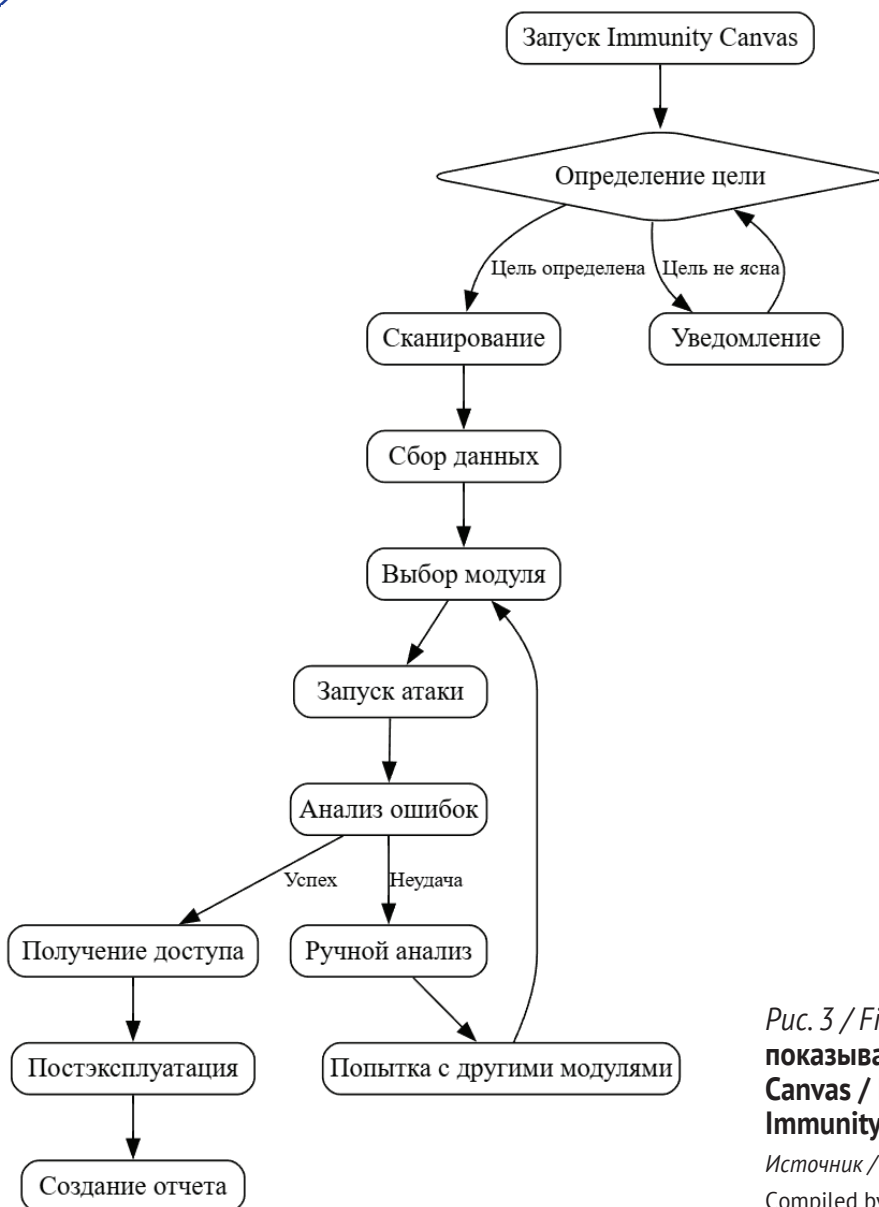


Рис. 3 / Fig. 3. UML – диаграмма, показывающая процесс работы Immunity Canvas / Is a UML Diagram Showing the Immunity Canvas Process

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

Архитектура фреймворка, показанная на диаграмме, оптимизирована для точного применения специфичных атакующих модулей, что позиционирует его как решение для целевых проверок устойчивости систем к сложным векторам атак.

ImmunitySec разработала коммерческий инструмент для устранения уязвимостей — Canvas. Последняя версия Immunity Canvas — 6,45. Он предназначен не для тестирования на проникновение, а, скорее всего, для тестирования безопасности и разработки эксплойтов. Canvas предлагает среду MOS-защиты, которая облегчает быстрое обнаружение уязвимостей. Компания может проверить конкретное представление о состоянии своей безопасности с помощью Immunity Canvas (рис. 3).

Платформа Security Forest Exploitation Framework — в ней также есть инструменты с открытым исходным кодом, которые могут быть использованы тестировщиками на проникновение.

Этот фреймворк использует набор кода эксплойта, известный как «Дерево эксплойтов». Его интерфейс позволяет тестировщику запускать код эксплойта через веб-браузер (рис. 4). Он поддерживает полный исходный код, а иногда даже включает эксплойты нулевого дня.

Сравнительные характеристики, описанные в табл. 2, позволяют увидеть различия фреймворков по критериям функциональной насыщенности и целевой аудиторией.

Наблюдается прямая корреляция между стоимостью решения и уровнем его автоматизации, в то время как open-source-альтернативы сохраняют преимущество в гибкости и доступности для кастомизации.

Для тестирования на проникновение доступно гораздо больше платформ, чем упомянуто выше. Wzaf — одна из них. По сути, это платформа для атак и аудита веб-приложений.



Рис. 4 / Fig. 4. UML – диаграмма, показывающая процесс работы Security Forest / Is a UML Diagram Showing the Security Forest Process

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

Ее можно применять для обнаружения уязвимостей в онлайн-приложениях и использования их слабых мест.

Большинство компаний и частных лиц сегодня работают с продуктами Microsoft, которые подвержены хорошо продуманным уязвимостям. Таким образом, защита данных и информации в настоящее время является наиболее сложной задачей.

Оценка сетевой безопасности или уязвимостей может быть в какой-то степени полезной, но она не обязательно показывает, как далеко могут зайти хакеры, чтобы воспользоваться уязвимостью. Хотя тестировщики на проникновение искренне пытаются в какой-то степени воспроизвести реальную атаку, они часто компрометируют систему, находя недостатки, которые можно эффективно использовать.

Хакеры и злоумышленники часто добиваются успеха в своих целях, поскольку им нужно найти лишь одну уязвимость, чтобы воспользоваться ею, – в то время как тестировщикам на проникновение может потребоваться обнаружить все уязвимости в сети. Учитывая, что тестирование на проникновение часто проводится в течение определенного периода времени, это сложный процесс.

Используемые в настоящее время платформы тестирования на проникновение недостаточно адаптированы для применения к различным типам систем или сетей. Ручные тесты, которые часто проводятся на таких платформах, как правило, представляют собой длительные и сложные процессы.

Из всего этого можно сделать вывод, что тест на проникновение сам по себе не повышает безопасность компьютера или сети.

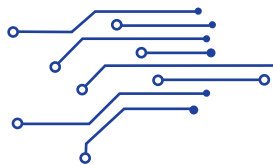
Для устранения уязвимостей, обнаруженных в ходе теста на проникновение, необходимы определенные действия.

Тестирование на проникновение может быть использовано для эффективной и успешной киберзащиты.

### Выводы

Проведенное исследование позволило сформулировать ряд ключевых выводов, основанных на сравнительном анализе инструментов и методологии тестирования на проникновение.

**Пентестинг – систематическая необходимость.** Тестирование на проникновение не является разовой мерой, а представляет собой неотъемлемый элемент жизненного цикла системы безопасности. Его необходимо регулярно применять при внедрении новых систем; после значительных обновлений инфраструктуры; для соблюдения нор-

Сравнительная таблица инструментов эксплуатации уязвимостей /  
Comparative Table of Vulnerability Exploitation Tools

Feature	Metasploit	Core Impact	Immunity Canvas	Security Forest
Количество эксплойтов	1467 эксплойтов	155 эксплойтов	800 эксплойтов	Огромное количество предустановленных эксплойтов
Интерфейс	Графический (GUI) и командная строка (CLI)	Графический (GUI)	Графический (GUI) и командная строка (CLI)	Графический (GUI) с ограниченными функциями
Платформа и установка	Независимая	Windows	Независимая	Windows
Стоимость	Бесплатно с ограничениями	2,5 млн руб. за одно рабочее место	300 тыс. руб. за 10 рабочих мест	Бесплатно
Язык программирования	Ruby, C	Python	Python	C, Python, Perl (для эксплойтов)
Обновления	Анонсируются на публичном сайте	Регулярные обновления доступны	Ежемесячные обновления	Периодические обновления
Выявление начальных хостов	Не поддерживается	Поддерживается	Не поддерживается	Не поддерживается
Функции отчетности	Бесплатно доступен, легко настраивается	Полностью автоматизирован, предлагает больше всего эксплойтов, профессиональный, но самый дорогой	Меньше эксплойтов, ниже стоимость, низкий уровень влияния	Предварительно скомпилированные и проиндексированные эксплойты

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

мативных требований; для независимой оценки эффективности защитных мер.

**Отсутствие универсального инструмента.** На рынке нет единого решения, покрывающего все потребности пентестинга. Различные инструменты дополняют друг друга: автоматизированные сканеры (Nmap, Nessus) эффективны для первичного выявления уязвимостей; фреймворки эксплуатации (Metasploit, Core Impact) необходимы для подтверждения серьезности уязвимостей и демонстрации потенциального ущерба.

**Выбор между коммерческими и open-source решениями.** Существует четкая дихотомия подходов: коммерческие продукты (например, Core Impact) обеспечивают глубокую автоматизацию и комплексный анализ, но имеют высокую стоимость; решения с открытым исходным кодом (например, Metasploit) предлагают гибкость и бесплатность, однако требуют от специалиста более высокой квалификации.

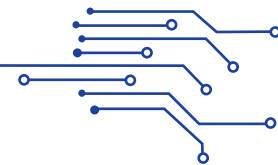
**Оптимальная стратегия** — комбинированный подход. Наибольшая эффективность достигается при интеграции: автоматизированных методов (быстрое сканирование, первичное выявление уязвимостей); ручных методов (глубокая проверка, подтверждение критичности уязвимостей).

**Строгая этапность процесса.** Пентестинг представляет собой последовательность четко определенных шагов: разведка и сбор информации; сканирование и выявление уязвимостей; эксплуатация уязвимостей для получения доступа; анализ полученных данных; составление отчета с рекомендациями по устранению недостатков.

**Ключевая роль специалиста.** Несмотря на развитие инструментов, решающим фактором успеха остается: квалификация тестировщика; его аналитическое мышление; способность действовать в правовых и этических рамках.

**Пентестинг как инструмент повышения киберустойчивости.** Грамотно проведенное тестирование на проникновение позволяет: выявить реальные уязвимости до их использования злоумышленниками; оценить реальный уровень защищенности инфраструктуры; обосновать необходимость инвестиций в безопасность; повысить общую киберустойчивость организации.

Таким образом, пентестинг — это не просто техническая процедура, а стратегический элемент системы информационной безопасности, требующий комплексного подхода, квалифицированных специалистов и регулярного применения.

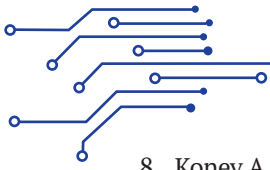


## СПИСОК ИСТОЧНИКОВ

1. Чуб В.С. Аудит безопасности информационной системы с использованием тестов на проникновение. *Молодой исследователь Дона*. 2018;6(15):88-90. URL: <https://www.elibrary.ru/ywbgdj>
2. Дворянкин О.А. Osint, Pentest и нетсталкинг — информационные технологии интернета. *Национальная ассоциация ученых*. 2022;84(2):6-13. URL: <https://www.elibrary.ru/lqlpwz>
3. Казыханов А.А., Байрушин Ф.Т. Pentest как основа обеспечения безопасности на средних и крупных предприятиях. *Символ науки*. 2016;10-2(22):50-51. URL: <https://www.elibrary.ru/wxdfbp>
4. Аверьянов В.С., Карцан И.Н. К вопросу выявления уязвимостей IPS/IDS систем. *Актуальные проблемы авиации и космонавтики*. 2020;2:191-197. URL: <https://www.elibrary.ru/rhmmin>
5. Алиева Е.М.К., Ширинова Ш.З.К. Актуальность атак с использованием SQL-инъекций. *In The World Of Science and Education*. 2025;3:163-167. DOI: 10.24412/3007-8946-2025-152-163-167
6. Шкрадюк А.Д. Оценка безопасности информационных систем с помощью тестирования на проникновение. *Умная цифровая экономика*. 2022;4(2):18-30. URL: <https://www.elibrary.ru/tshjtt>
7. Симбирцев Д.В., Жуков В.Г. Разработка автоматизированной системы анализа защищенности веб-ресурсов. *Актуальные проблемы авиации и космонавтики*. 2011;7:430. URL: <https://www.elibrary.ru/taozxn>
8. Конев А.А., Паюсова Т.И. Большие языковые модели в информационной безопасности и тестировании на проникновение: систематический обзор возможностей применения. *Научно-технический журнал информационных технологий, механики и оптики*. 2025;1:41-52. DOI: 10.17586/2226-1494-2025-25-1-42-52
9. Мордвинова А.Ю., Нуриев С.А. Исследование уязвимостей и угроз безопасности стандарта IEEE 802.11. *Современные инновации, системы и технологии*. 2023;3:117-131. DOI: 10.47813/2782-2818-2023-3-3-0117-0131
10. Гылыджова А., Пирлиев К., Ходжамбердиев С., Худайберенов Р. Анализ безопасности ALT LINUX с применением LYNIS в качестве опорной модели для проверки защищенности. *Символ науки*. 2024;4-2:71-74. URL: <https://www.elibrary.ru/bywprju>
11. Серов С.А., Серов С.С., Петрова И.В. Metasploit Framework как средство эксплуатации уязвимых серверов. *Форум молодых ученых*. 2021;5:1307-1313. URL: <https://www.elibrary.ru/honvex>
12. Дедов Д.О. Защита SSH-порта с использованием имитации уязвимостей Honeyrot. *Молодой исследователь Дона*. 2023;8-6(45):16-21. URL: <https://www.elibrary.ru/hhlcno>
13. Чемеркин Ю.С. Безопасность публичных сред облачных вычислений в условиях функциональной неопределенности. *T-Comm: телекоммуникации и транспорт*. 2014;6:56-60. URL: <https://media-publisher.ru/en/content-6-2014/>
14. Nahanova I.V. Method of Pentest Synthesis and Vulnerability Detection. *Радиоэлектроника и информатика*. 2012;4:68-73.
15. Samarov X.Q., Salimov Z.I.O., Qosimov I.S. Kiberxavfsizlikga oid fanlarga laboratoriya mashg'ulotlarini bajarish uchun platformani yaratishning dolzarbligi. *In The World Of Science and Education*. 2025;15. URL: <https://irc-els.com/docs/СБОРНИК%20МНЖ%2015%20ФЕВРАЛЯ%202025%20ТЕХНИЧЕСКИЕ%20НАУКИ.pdf>

## REFERENCES

1. Chub V.S. Information system security audit using penetration tests. *A young researcher of the Don*. 2018;6(15):88-90. URL: <https://www.elibrary.ru/ywbgdj> (In Russ.).
2. Dvoryankin O.A. Osint, Pentest and netstalking — information technologies of the Internet. *National Association of Scientists*. 2022;84(2):6-13. URL: <https://www.elibrary.ru/lqlpwz> (In Russ.).
3. Kazykhanov A.A., Bayrushin F. T. Pentest as a basis for ensuring security in medium and large enterprises. *A symbol of Science*. 2016;10-2(22):50-51. URL: <https://www.elibrary.ru/wxdfbp> (In Russ.).
4. Averyanov V.S., Kartsan I.N. On the issue of identifying vulnerabilities in IPS/IDS systems. *Actual Problems of Aviation and Cosmonautics*. 2020;2:191-197. URL: <https://www.elibrary.ru/rhmmin> (In Russ.).
5. Alieva E.M.K., Shirinova S.Z.K. The relevance of attacks using SQL injections. *In The World Of Science and Education*. 2025;3:163-167. (In Russ.). DOI: 10.24412/3007-8946-2025-152-163-167
6. Shkradyuk A.D. Information system security assessment using penetration testing. *Smart Digital Economy*. 2022;4(2):18-30. URL: <https://www.elibrary.ru/tshjtt> (In Russ.).
7. Simbirtsev D.V., Zhukov V.G. Development of an automated system for analyzing the security of web resources. *Actual Problems of Aviation and Cosmonautics*. 2011;7:430. URL: <https://www.elibrary.ru/taozxn> (In Russ.).



8. Konev A.A., Payusova T.I. Large language models in information security and penetration testing: a systematic review of application possibilities. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. 2025;1:41-52. (In Russ.). DOI: 10.17586/2226-1494-2025-25-1-42-52
9. Mordvinova A.Yu., Nuriev S.A. Investigation of vulnerabilities and security threats of the IEEE 802.11 standard. *Modern Innovations, Systems and Technologies*. 2023;3:117-131. (In Russ.). DOI: 10.47813/2782-2818-2023-3-3-0117-0131
10. Gylydzhova A., Pirliev K., Khodjamberdiev S., Khudaiberenov R. ALT LINUX security analysis using LYNIS as a reference model for security verification. *A Symbol of Science*. 2024;4-2-2:71-74. URL: <https://www.elibrary.ru/bywpju> (In Russ.).
11. Serov S.A., Serov S.S., Petrova I.V. Metasploit Framework as a means of exploiting vulnerable servers. *Forum of Young Scientists*. 2021;5:1307-1313. URL: <https://www.elibrary.ru/honvex> (In Russ.).
12. Dedov D.O. SSH port protection using imitation of Honeypot vulnerabilities. *A Young Researcher of the Don*. 2023;8-6(45):16-21. URL: <https://www.elibrary.ru/hhlcno> (In Russ.).
13. Chemerkin Yu.S. Security of public cloud computing environments in conditions of functional uncertainty. *T-Comm: Telecommunications and Transport*. 2014;6:56-60. URL: <https://media-publisher.ru/en/content-6-2014/> (In Russ.).
14. Hakhanova I.V. Method of Pentest Synthesis and Vulnerability Detection. *Radio Electronics and Computer Science*. 2012;4:68-73.
15. Samarov Kh.Q., Salimov Z.I.O., Kasimov I.S. The relevance of creating a platform for performing laboratory training in cybersecurity-related disciplines. *In the World of Science and Education*. 2025;15. URL: <https://irc-els.com/docs/СБОРНИК%20МНЖ%2015%20ФЕВРАЛЯ%202025%20ТЕХНИЧЕСКИЕ%20НАУКИ.pdf> (in Kz.).

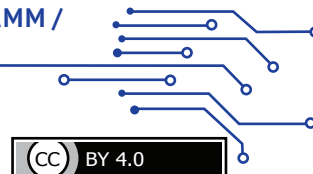
## ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

**Артем Игоревич Любимов** — студент факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация  
**Artem I. Lyubimov** — student of the Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation  
<https://orcid.org/0009-0005-2633-7909>  
[artem.lyubimov.harby@gmail.com](mailto:artem.lyubimov.harby@gmail.com)

**Сергей Анатольевич Резниченко** — кандидат технических наук, доцент, кафедры информационной безопасности факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация  
**Sergey A. Reznichenko** — Cand. Sci. (Tech), Assoc. Prof., Department of Information Security, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation  
<https://orcid.org/0000-0002-1539-0457>  
Автор для корреспонденции / Corresponding author:  
[sareznichenko@fa.ru](mailto:sareznichenko@fa.ru); [rsa\\_5@bk.ru](mailto:rsa_5@bk.ru)

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.  
Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила 19.01.2026; после рецензирования 02.02.2026; принята к публикации 12.02.2026.  
Авторы прочитали и одобрили окончательный вариант рукописи.  
The article was submitted on 19.01.2026; revised on 02.02.2026 and accepted for publication on 12.02.2026.  
The authors read and approved the final version of the manuscript.



## ОРИГИНАЛЬНАЯ СТАТЬЯ

DOI: 10.26794/3030-7097-2026-2-1-45-51  
УДК 004.852:330.43:311.14:339.9(045)

## Методы интеллектуального анализа данных в исследовании экономического развития стран

Л.Р. Борисова

Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

## АННОТАЦИЯ

**Цель исследования** – выявить взаимосвязи между относительными экономическими показателями (индексами развития) стран мира на основе данных Всемирного банка за 2023 г. Для решения задачи применены методы машинного обучения и анализа структурных уравнений. **Материалы и методы.** Исследование охватило 205 стран – при анализе методом взвешенных синдромов; 180 стран – при одно- и двухфакторном подтверждающем анализе. В качестве исходных данных использованы восемь ключевых относительных экономических показателей развития стран. Группирующим критерием выступил темп роста ВВП в 2023 г. (в %). Для анализа применены метод взвешенных синдромов; одно- и двухфакторный подтверждающий анализ; оценка качества моделей по индикатору ROC AUC, индексам сравнительного соответствия и Такера-Льюиса (TLI). **Результаты.** 1. Модель на основе метода взвешенных синдромов показала высокое качество распознавания: ROC AUC = 0,92. Диаграммы рассеяния подтвердили четкое разделение стран на две группы по анализируемому показателю. 2. При факторном анализе построены: однофакторная модель с низкими нагрузками ( $\lambda=0,258$ ;  $\lambda=0,131$ ), что указывает на слабую связь наблюдаемых переменных с латентным фактором; двухфакторная модель с неудовлетворительным индексом соответствия TLI = 0,474 (при пороге > 0,90–0,95). **Выводы.** Методы машинного обучения продемонстрировали преимущество при работе с данными, содержащими пропуски. Факторный анализ дал неудовлетворительные результаты: модели не объясняют дисперсию переменных и не соответствуют данным. Содержательная экономическая интерпретация латентных факторов не проведена из-за низкого качества факторных моделей. Сравнение методов машинного обучения и факторного анализа затруднено из-за разного объема пригодных данных.

**Ключевые слова:** машинное обучение; факторный анализ; экономические показатели; индексы развития; ВВП; ROC AUC; индекс Такера-Льюиса; взвешенные синдромы; структурные уравнения

**Для цитирования:** Борисова Л.Р. Методы интеллектуального анализа данных в исследовании экономического развития стран. *Цифровые решения и технологии искусственного интеллекта.* 2026;2(1):45-51. DOI: 10.26794/3030-7097-2026-2-1-45-51

## ORIGINAL PAPER

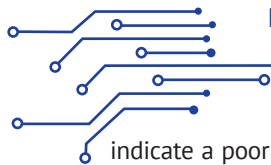
## Methods of Data Mining in the Study of Economic Development of Countries

L.R. Borisova

Financial University under the Government of the Russian Federation, Moscow, Russian Federation

## ABSTRACT

The paper presents a demonstration of the analysis of a specific data set using machine learning methods to solve the research problem of finding a link between relative economic indicators (development indices) of the world's countries according to the World Bank data for 2023 using machine learning methods and the use of structural equations. This approach was applied to the analysis of 205 countries using the weighted syndrome method and 180 countries using one- and two-factor confirmatory analysis. The advantage of using machine learning methods is the ability to use data with gaps, unlike regression models, which are the basis of factor analysis, when data gaps are not acceptable. To use the weighted syndromes method, eight main economic relative indicators of the countries' development were used. The % GDP growth rate for 2023 was chosen as the grouping indicator. The quality of the model was assessed by the ROC AUC indicator. This indicator is 0.92, which indicates that the selected features really make it possible to divide the countries. Scattering diagrams showing a clear division of countries into two groups according to the analyzed indicators also illustrate the quality of recognition. The use of factor analysis made it possible to build two models (one-factor and two-factor) using not eight indicators, but only four, so that the models measured by the indices of comparative conformity and Tucker-Lewis were statistically significant. However, the loads of the one-factor model ( $\lambda$ ) are extremely low (0.258, 0.131), which indicates a weak relationship between the observed variables and the latent factor. The factor practically does not explain the variance of the variables. These results



indicate a poor correspondence of the model to the data, especially for the two-factor model, since the value of the TLI index is 0.474, which also cannot be a satisfactory result with an adequate threshold of  $>0.90-0.95$ ). A meaningful economic interpretation of the latent factors obtained is not provided due to the poor correspondence of the results of factor analysis and the inability to compare machine learning and factor analysis methods with simultaneous analysis, since there was less data suitable for factor analysis than when using machine learning methods that demonstrated their adequacy.

**Keywords:** machine learning; factor analysis; economic indicators; development indices; GDP; ROC AUC; Tucker Lewis index; weighted syndromes; structural equations

**For citation:** Borisova L.R. Methods of data mining in the study of economic development of countries. *Digital solutions and artificial intelligence technologies*. 2026;2(1):45-51. DOI: 10.26794/3030-7097-2026-2-1-45-51

## ВВЕДЕНИЕ

Мировой опыт свидетельствует о существовании значительных диспропорций в развитии отдельных стран. Проблема оценки рационального развития национальной экономики становится актуальной, поскольку такие оценки могут предотвратить диспропорции в развитии между странами. Вследствие этого возникает необходимость использования методов математического моделирования для анализа относительных показателей экономического развития. Чаще всего применяются методы математической статистики: регрессионные модели, кластерный и факторный анализы [1].

**Обзор литературы.** В работе [2] использован кластерный анализ. Количество кластеров определено заранее — остановились на трех кластерах. Полученные результаты показывают, что странам кластера 1 требуется существенная поддержка в области управления и инфраструктуры, в то время как странам кластера 2 выгодны инвестиции в институциональную сферу и логистику. Кластер 3 демонстрирует примерные показатели управления и социально-экономических стандартов, предлагая модели устойчивого развития.

В работе [3] с помощью методов машинного обучения удалось установить, что учет ежемесячных данных о суровых климатических условиях — особенно информации, содержащейся в значениях, превышающих среднее (выше среднего) по экстремальным погодным условиям, — значительно превосходит не только базовую авторегрессионную модель, но и эконометрическую структуру, включающую макрофинансовые факторы, при прогнозировании темпов роста квартального реального валового накопления основного капитала.

В работе [4], внедряя принципы энергетической иерархии, обратной связи и самоорганизации Юджина Одума<sup>1</sup>, предложена модель для диагностики системных дисбалансов и разработки экологически обоснованных стратегий развития. Модель предлагает более последовательную альтернативу

традиционным экономически ориентированным показателям.

Помимо статистических методов, в математическом моделировании очень часто используется аппарат дифференциальных уравнений. В частности, в работе [5] представлена новая математическая модель для исследования влияния развития навыков на управление циклической безработицей. Сформирована и проанализирована система нелинейных дифференциальных уравнений. На основании полученных результатов сделан вывод, что развитие навыков у циклически безработных — эффективное средство снижения безработицы.

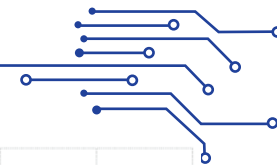
В работе [6] исследована взаимосвязь между экономическим ростом и теневой экономикой. Результаты моделирования показали, что качество институтов тесно связано с взаимосвязью между экономическим ростом и теневой экономикой: более высокий ВВП на душу населения соответствует меньшему уровню теневой экономики в странах с хорошим качеством институтов. Однако в странах с низким качеством институциональной среды увеличение ВВП на душу населения не влияет на масштабы теневой экономики.

В работе [7] выполнен учет безработицы и других экономических факторов при исследовании развития территориальных образований. Рассмотрена математическая модель линейной и многофакторной регрессии, описывающая взаимосвязь между количеством затраченных ресурсов и объемом произведенной продукции. Созданная модель четко показывает, что сельское хозяйство в местных сообществах может стать экономическим двигателем развития с учетом региональных условий.

Отечественные ученые в исследованиях делают основной акцент на использование методов математической статистики и эконометрики. Стоит отметить в этой связи работы В. С. Степанова с соавторами [8, 9], в которых построены интересные регрессионные модели, связывающие уровень жизни населения и уровень онкозаболеваемости с индикаторами социально-экономического развития в регионах России.

Также заслуживают внимания работы С. С. Матвеевского и Л. Р. Борисовой, в которых исследуется

<sup>1</sup> Американский ученый, занимающийся моделированием в экологии.



классификация и кластеризация отдельных регионов или стран по значениям инклюзивного роста [10] и электронных услуг и сервисов [1].

## МАТЕРИАЛЫ И МЕТОДЫ

Для исследования взяты данные с сайта Всемирного банка<sup>2</sup> за 2023 г., так как за 2024–2025 гг. данных недостаточно.

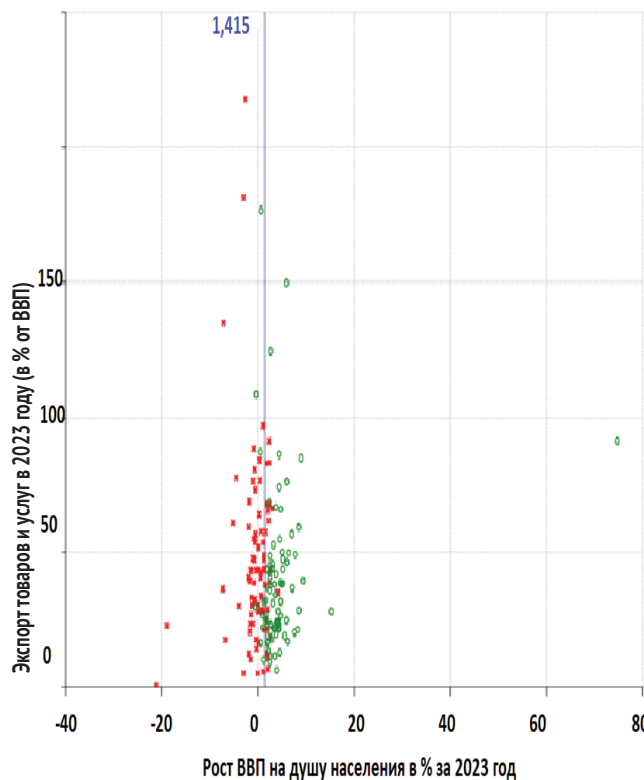
В базу данных для использования методов машинного обучения вошли 205 стран, для которых представлено большинство из следующих восьми показателей: рост ВВП в %; рост ВВП на душу населения в %; инфляция, измеряемая ростом потребительских цен (% в годовом исчислении); годовой прирост в промышленности в %; % от ВВП добавленной стоимости в промышленности; % от ВВП добавленной стоимости в обрабатывающей промышленности; импорт и экспорт товаров и услуг в виде % от ВВП; доступ к электричеству (% населения).

В качестве группирующего показателя при использовании методов машинного обучения был взят рост ВВП в % за 2023 г. В качестве метода распознавания применен метод статистически значимых синдромов (СВС) [11]. Метод СВС основан на процедуре взвешенного голосования по системам так называемых синдромов — областей признакового пространства, содержащих преимущественно объекты одного из классов [1]. Методы машинного обучения реализованы в системе анализа данных Data Master Azforus, использованной ранее в работе [12].

Подтверждающий факторный анализ задействован с применением статистического языка программирования R<sup>3</sup>.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

При выполнении скользящего контроля в результатах распознавания наблюдаются ошибки. Были отобраны случайным образом 42 страны из 205 (20%), число стран в каждой группе было по 21. Полученное значение показателя ROC AUC, равное 0,92 (близкое к 1), свидетельствует в пользу использования методов машинного обучения для классификации стран по показателям, важным для экономического развития. Заметим, что страны четко разделяются по таким основным показателям, как ВВП на душу населения и экспорт товаров и услуг (диаграмма рассеяния представлена на *рис. 1*). Гра-



*Рис. 1 / Fig. 1. Различие между двумя группами стран по зависимости экспорта товаров и услуг от роста ВВП в 2023 г. / The Difference Between the Two Groups of Countries in the Dependence of Exports of Goods and Services on GDP Growth in 2023*

*Источник / Source:* составлено автором / Compiled by the author.

ница, разделяющая страны с ростом ВВП за 2023 г. с уровнем, ниже медианного от стран с ростом ВВП выше медианного уровня, проходит на уровне 1,415%. Интересно заметить, что эта граница не изменится, если представить диаграмму зависимости импорта товаров и услуг от роста ВВП или зависимости годового прироста в промышленности (в %) от роста ВВП; % от ВВП добавленной стоимости в промышленности от роста ВВП.

### Факторный анализ

Моделирование структурными уравнениями охватывает широкий спектр линейных моделей, а именно линейную и многомерную регрессии, анализ траекторий, подтверждающий факторный анализ и структурную регрессию [1]. Из огромного количества экономических показателей стран, размещенных на сайте Всемирного банка, выбраны:

- показатель инфляции в виде роста потребительские цены (% в годовом исчислении) —  $x_1$ ;
- показатель добавленной стоимости в промышленности (включая строительство) (% от ВВП) —  $x_2$ ;

<sup>2</sup> URL: <https://www.worldbank.org/>

<sup>3</sup> Introduction to Structural Equation Modeling (SEM) in R with lavaan. URL: <https://stats.oarc.ucla.edu/r/seminars/rsem/>

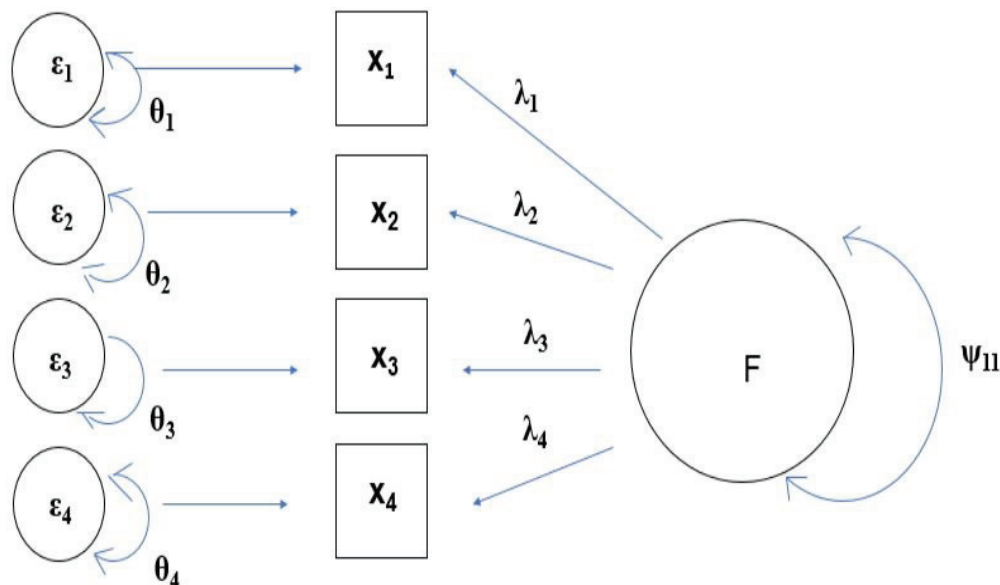


Рис. 2 / Fig. 2. Схема однофакторного подтверждающего анализа с четырьмя позициями /  
 Scheme of One-factor Confirmatory Analysis with four Positions

Источник / Source: составлено автором / Complied by the author.

- процент населения, имеющих доступ к электричеству, —  $x_3$ ;
- экспорт товаров и услуг (% от ВВП) —  $x_4$ .

Все показатели — за 2023 г. Факторный анализ используется для ответа на вопрос: насколько общая дисперсия является общей между элементами?

#### Однофакторная модель

Факторный анализ — это многомерная модель, в которой на каждого субъекта приходится столько результатов, сколько и элементов. В линейной регрессии для каждого субъекта существует только один исход. Наиболее фундаментальной моделью в подтверждающем факторном анализе является однофакторная модель, предполагающая, что ковариация между элементами обусловлена одним общим фактором [1]. Рассмотрим четыре показателя и один фактор (рис. 2).

Параметры:

- $\lambda_i$  ( $i = 1, \dots, 4$ ) — нагрузки, которые можно интерпретировать как корреляцию элемента с фактором F;
- $\varepsilon_i$  ( $i = 1, \dots, 4$ ) — остатки факторной модели, то, что остается после учета фактора;  $\theta_i$  ( $i = 1, \dots, 4$ ) — дисперсии (ковариации) остатков.

Основными индикаторами качества факторной модели являются хи-квадрат ( $\chi^2$ -распределение):

- статистика, получаемая из статистики максимального правдоподобия;
- индекс сравнительного соответствия CFI, значения которого могут варьироваться от 0 до 1 (значения больше 0,90 указывают на хорошее соответствие);

- индекс Такера-Льюиса TLI, также находящийся в диапазоне от 0 до 1 (значения больше 0,90 указывают на хорошее соответствие);

- среднеквадратичная ошибка аппроксимации RMSEA (если эта ошибка меньше 0,05, то соответствие модели данным хорошее).

Для анализируемых данных получены следующие значения индикаторов качества: CFI = 0,967, TLI = 0,901, RMSEA = 0,039. Как видим, модель статистически значимая. Осталось проанализировать корреляцию фактора с переменными. В табл. 1 представлены значения полученных при моделировании параметров [1].

Заметим, что в данной модели  $\psi_{11} = 1$ . Для всех дисперсий, кроме  $x_4$ , величина p-value = 0. Переменная  $x_4$  — экспорт товаров и услуг (в % к ВВП). В этом случае величина p-value = 0,006, т.е. близко к нулю. Если под фактором мы понимаем некий интегральный социально-экономический показатель, определяющий величину замедления роста экономики в регионах, то интересно отметить, что положительно на этот фактор влияют только два показателя; инфляция и показатель добавленной стоимости в промышленности (в % от ВВП). И наоборот, доступ к электричеству и экспорт товаров способствуют росту экономики (параметры  $\lambda_1$  и  $\lambda_2$  — положительные,  $\lambda_3$  и  $\lambda_4$  — отрицательные).

#### Двухфакторная модель

Рассмотрим четыре показателя и два фактора. Схема модели имеет вид, представленный на рис. 3.

**Значения параметров однофакторной модели / Values of the Parameters of the One-factor Model**

$\lambda$	$\theta$	$\lambda$	$\theta$
$\lambda_1 = 0,258$	$\theta_1 = 0,928$	$\lambda_3 = -0,446$	$\theta_3 = 0,795$
$\lambda_2 = 0,131$	$\theta_2 = 0,977$	$\lambda_4 = -0,595$	$\theta_4 = 0,641$

Источник / Source: составлено автором / Complied by the author.

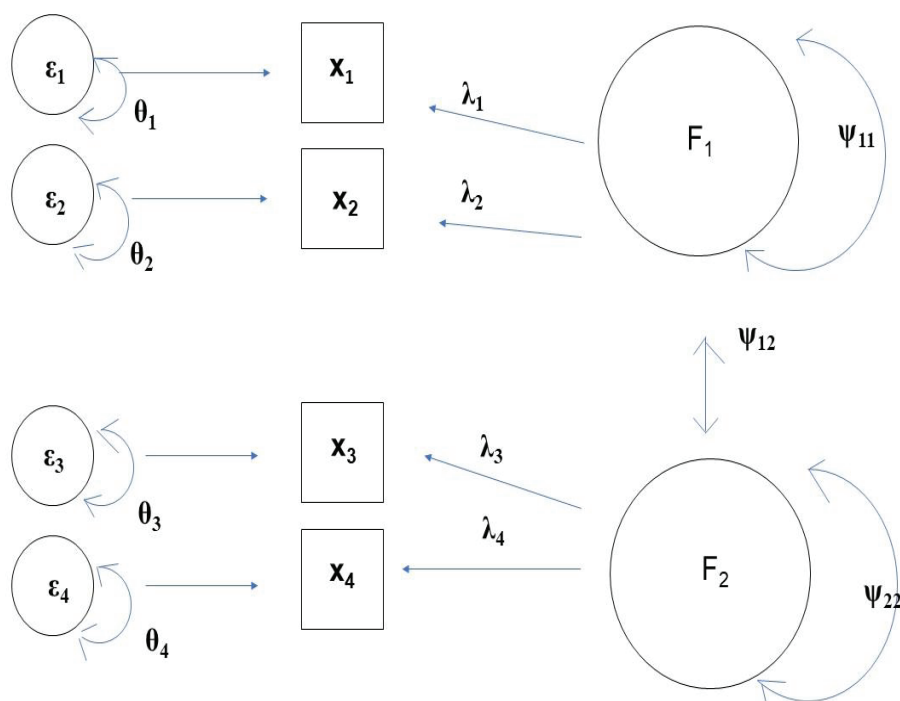


Рис. 3 / Fig. 3. Схема двухфакторного анализа с четырьмя позициями /  
The Scheme of Two-factor Confirmatory Analysis with Four Positions

Источник / Source: составлено автором / Complied by the author.

Как и в случае однофакторного анализа, моделирование в случае предположения о существовании двух скрытых факторов, влияющих на экономический региональный рост, было выполнено с использованием программы lavaan, имеющейся в библиотеке программ языка R. Для двухфакторного анализа зафиксировали две группы показателей: в первую вошли показатель инфляции и показатель добавленной стоимости в промышленности, во вторую — доступ к электричеству и экспорт

товаров. Оказалось, что два латентных фактора имеют отрицательную ковариацию  $\psi_{12} = -1,592$ . Для анализируемых данных получены следующие значения индикаторов качества: CFI = 0,912, TLI = 0,474, RMSEA = 0,09. Как видим, модель статистически значимая. Осталось проанализировать корреляцию двух скрытых факторов с исследуемыми переменными. В табл. 2 представлены значения полученных при моделировании параметров [1].

Таблица 2 / Table 2

**Значения параметров двухфакторной модели / Values of the Two-factor Model Parameters**

$\lambda$	$\theta$	$\lambda$	$\theta$
$\lambda_1 = 0,164$	$\theta_1 = 0,968$	$\lambda_3 = 0,442$	$\theta_3 = 0,799$
$\lambda_2 = 0,087$	$\theta_2 = 0,987$	$\lambda_4 = 0,597$	$\theta_4 = 0,638$

Источник / Source: составлено автором / Complied by the author.



Заметим, что в данной модели  $\psi_{11} = 1, \psi_{22} = 1$ . Интересно отметить, что наличие второго скрытого фактора изменило направление связи исследуемых признаков и фактора: все параметры  $\lambda_i (i = 1, \dots, 4)$  — положительные, но сами латентные факторы между собой отрицательно коррелируют ( $\psi_{12} = -1,592$ ).

### ВЫВОДЫ

В работе проанализированы данные за 2023 г. по экономическому развитию стран мира, представленные на сайте Всемирного банка двумя разными методами: методом машинного обучения и факторным анализом. В качестве метода классификации был использован метод статистически взвешенных синдромов, который продемонстрировал хорошее качество распознавания на контрольной выборке, составившей 20% от генеральной совокупности (площадь под РОК-кривой равна 0,92, что достаточно близко к 1). Следует отметить, что при применении более распространенных ал-

горитмов классификации — например, случайного леса или моделей бустинга — точность распознавания оказалась ниже по сравнению с методом статистически взвешенных синдромов.

Следовательно, применение технологий машинного обучения подтвердило, что выбранный набор экономических показателей пригоден для классификации стран.

Преимущество данного подхода заключается в том, что он позволяет включать в анализ страны, имеющие пропуски в данных.

После уменьшения числа показателей с 8 до 4 и уменьшения стран в базе данных для исследования были использованы структурные уравнения. Оказалось, что применение факторного анализа также возможно при получении статистически значимых результатов, которые заключаются в том, что существуют некоторые латентные факторы, участвующие во взаимосвязи между исследуемыми экономическими показателями.

### СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Кузнецова А.В., Борисова Л.Р., Кремер Н.Ш., Фридман М.Н. Сравнительный анализ дотационных регионов Российской Федерации методами машинного обучения по широкому набору показателей основных фондов. *Бизнес. Образование. Право*. 2025;1(70):20-28. DOI: 10.25683/VOLBI.2025.70.1182  
Kuznetsova A.V., Borisova L.R., Kremer N.S., Friedman M.N. Comparative analysis of subsidized regions of the Russian Federation using machine learning methods for a wide range of indicators of fixed assets. *Biznes. Obrazovanie. Pravo = Business. Education. Law*. 2025;1(70):20-28. (In Russ.). DOI: 10.25683/VOLBI.2025.70.1182
2. Visbal-Cadavid D., Delahoz-Dominquez E., Mendoza-Mendoza A. A multiple factor analysis and hierarchical clustering of global logistics governance and development. *Decision Analytics Journal*. 2025;(15):100579. DOI: 10.1016/j.dajour.2025.100579
3. Sheng X., Cepni O., Gupta R., Markovski M. Mixed Frequency Machine Learning Forecasting of the Growth of Real Gross Fixed Capital Formation in the United States: The Role of Extreme Weather Conditions 2025: Working Papers University of Pretoria. 2025;202520. URL: <https://EconPapers.repec.org/RePEc:pre:wpaper:202520>
4. Langa E.S., Giannetti B.F., Sevegnani f., Agostinho F., Almeida C. From theory to application: measuring development disparities in Mozambique through an Odum-inspired emergy framework. *Ecological Modeling*. 2025;(510):111287. DOI: 10.1016/j.ecolmodel.2025.111287
5. Yahyaoui M.El., Amine S. Mathematical modeling of unemployment dynamics with skills development and cyclical effects. *Partial differential equations in applied mathematics*. 2024; (911):100800. DOI: 10.1016/j.padiff.2024.100800
6. Baklouti N., Boujelbene Y. A simultaneous equation model of economic growth and shadow economy: Is there a difference between the developed and developing countries? *Economic Change and Restructuring*, Springer. 2020;53(1):151-170. DOI: 10.1007/s10644-018-9235-8
7. Bihun R., Lytvyn V., Oleksiv N. Mathematical modeling and analysis of the development of territorial communities. *Technology Audit and Production Reserves*. 2021;3(2(59)):6-12. DOI: 10.15587/2706-5448.2021.232788
8. Степанов В.С., Бобков В.Н., Шамаева Е.Ф., Одинцова Е.В. Построение модели, связывающей индикатор уровня жизни населения с комплексом показателей социально-экономической политики в регионах России. *Уровень жизни населения регионов России*. 2022;18(4):450-465. (In Russ.). DOI: 10.19181/lspr.2022.18.4.3



- Stepanov V.S., Bobkov V.N., Shamaeva E.F., Odintsova E.V. Building a model linking the indicator of the standard of living of the population with a set of indicators of socio-economic policy in the regions of Russia. *Living Standard of the Population in the Regions in Russia*. 2022;18(4):450-465. (In Russ.). DOI: 10.19181/lsprr.2022.18.4
9. Степанов В.С. Прогноз общей онкологической заболеваемости в регионах и муниципалитетах России на основе многофакторной модели. *Моделирование, оптимизация и информационные технологии*. 2023;11(1):1-17. DOI: 10.26102/2310-6018/2023.40.1.022  
Stepanov V.S. The forecast of cancer prevalence in the regions and municipalities of Russia based on a multivariate model. *Modeling, Optimization and Information Technology*. 2023;11(1):1-17. DOI: 10.26102/2310-6018/2023.40.1.022 (In Russ.).
10. Матвиевский С.С., Борисова Л.Р. Кластеризация стран Азиатско-Тихоокеанского региона по значениям инклюзивного экономического роста. *Вестник университета*. 2023;(1):112-121. DOI: 10.26425/1816-4277-2024-1-112-121  
Matviyevsky S.S., Borisova L.R. Clusterization of the countries of the Asia-Pacific region according to the values of inclusive economic growth. *Bulletin of the University*. 2023; (1):112-121. (In Russ.). DOI: 10.26425/1816-4277-2024-1-112-121
11. Борисова Л.Р. Сравнительный анализ регионов Российской Федерации методами машинного обучения по набору показателей электронных услуг и сервисов. *Цифровая социология*. 2024;7(4):33-43. DOI: 10.26425/2658-347X-2024-7-4-33-43  
Borisova L.R. Comparative analysis of the regions of the Russian Federation using machine learning methods for a set of indicators of electronic services and services. *Digital Sociology*. 2024;7(4):33-43. (In Russ.). DOI: 10.26425/2658-347X-2024-7-4-33-43
12. Senko O.V., Kuznetsova A.V. A recognition method based on collective decision-making using systems of regularities of various types. *Pattern Recogn. Image Anal.* 2010;20(2):152-162. DOI: 10.1134/S1054661810020069

## ИНФОРМАЦИЯ ОБ АВТОРЕ / ABOUT THE AUTOR

**Людмила Робертовна Борисова** — кандидат физико-математических наук, доцент кафедры математики и анализа данных факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**Lyudmila R. Borisova** — Cand. Sci. (Phys. And Math.) Assoc. Prof., Department of Mathematics and Data Analysis, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<http://orcid.org/0000-0002-5757-0341>

[lrborisova@fa.ru](mailto:lrborisova@fa.ru)

*Конфликт интересов: автор заявляет об отсутствии конфликта интересов.*

*Conflicts of Interest Statement: The author has no conflicts of interest to declare.*

*Статья поступила 23.09.2025; после рецензирования 31.10.2025; принята к публикации 15.12.2025.*

*Автор прочитала и одобрила окончательный вариант рукописи.*

*The article was submitted on 23.09.2025; revised on 31.10.2025 and accepted for publication on 15.12.2025.*

*The author read and approved the final version of the manuscript.*

DOI: 10.26794/3030-7097-2026-2-1-52-62  
УДК 519.17(045)

# Assessment of the Structural Stability of Transport and Logistics Systems Based on Graph Models and the Percolation Coefficient

D.V. Yatskin<sup>1</sup>, A.A. Kochkarov<sup>2</sup>, E.A. Okuneva<sup>3</sup><sup>1</sup>HSE University, Moscow, Russian Federation;<sup>2</sup>FIC "Fundamental Foundations of Biotechnology" of the Russian Academy of Sciences, Moscow, Russian Federation;<sup>3</sup>Financial University under the Government of the Russian Federation, Moscow, Russian Federation

## ABSTRACT

Global changes in the global economy and trade place new demands on transport and logistics systems (TLS). Increasing their resilience to disruptions and structural changes is becoming a critical task amid the growing share of logistics costs in Russia's GDP. Existing optimization methods often do not take into account the resilience of the network structure itself to destructive influences, which creates a gap in knowledge. **The purpose** of the study is to develop and test methodological tools for assessing and improving the structural stability of the TL. **The approach** integrates graph-theoretical modeling, multi-criteria optimization methods, and a new indicator, the percolation coefficient, which characterizes the network's ability to deliver goods to all destinations. The multi-criteria optimization problem of finding paths and flows is formalized. Sustainability was assessed through the coefficient of influence of structural changes on the effectiveness of solutions. A large-scale computational experiment was conducted with the generation of more than 1 million graph structures. A mathematical model of the radar has been developed based on a matrix of initial conditions, and an efficiency coefficient has been proposed for comparing alternative options. A close correlation has been established between network bandwidth, percolation coefficient, and solution efficiency. The barrier values of the coefficient of influence have been determined, which make it possible to classify the system as stable or unstable to a specific type of structural failure. The principles of building sustainable radar stations are formulated, the key of which is the availability of alternative routes with efficiency close to optimal. **The results** obtained lay the foundation for the creation of intelligent radar stations that are resistant to failures and load fluctuations.

**Keywords:** transport and logistics system; structural stability; graph theory; multicriteria optimization; percolation coefficient; network modeling

**For citation:** Yatskin D.V., Kochkarov A.A., Okuneva E.A. Assessment of the structural stability of transport and logistics systems based on graph models and the percolation coefficient. *Digital solutions and artificial intelligence technologies*. 2026;2(1):52-62. DOI: 10.26794/3030-7097-2026-2-1-52-62

## ОРИГИНАЛЬНАЯ СТАТЬЯ

# Оценка структурной устойчивости транспортно-логистических систем на основе графовых моделей и коэффициента просачиваемости

Д.В. Яцкин<sup>1</sup>, А.А. Кочкаров<sup>2</sup>, Э.А. Окунева<sup>3</sup><sup>1</sup>НИУ ВШЭ, Москва, Российская Федерация;<sup>2</sup>ФИЦ «Фундаментальные основы биотехнологии» РАН, Москва, Российская Федерация;<sup>3</sup>Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

## АННОТАЦИЯ

Глобальные изменения в мировой экономике и торговле предъявляют новые требования к транспортно-логистическим системам (ТЛС). Повышение их устойчивости к сбоям и структурным изменениям становится критической задачей на фоне роста доли логистических издержек в ВВП России. Существующие методы оптимизации часто не учитывают устойчивость самой сетевой структуры к деструктивным воздействиям, что создает пробел в знаниях. **Цель работы** – создать и апробировать методологический инструмент для оценки и усиления структурной устойчивости ТЛС.



В рамках предложенного подхода объединены: графо-теоретическое моделирование; методы многокритериальной оптимизации; новый показатель — коэффициент просачиваемости, отражающий способность сети обеспечивать доставку грузов во все пункты назначения. **В ходе исследования:** формализована многокритериальная оптимизационная задача, направленная на поиск оптимальных путей и потоков; разработана математическая модель ТЛС на базе матрицы начальных условий; введен коэффициент эффективности для сопоставления альтернативных вариантов решений. Проведена оценка устойчивости через коэффициент влияния структурных изменений на эффективность решений. Выполнен крупномасштабный вычислительный эксперимент, в рамках которого сгенерировано свыше 1 млн графовых структур. **Основные результаты:** выявлена тесная корреляция между пропускной способностью сети, коэффициентом просачиваемости и эффективностью принимаемых решений; определены барьерные значения коэффициента влияния, позволяющие классифицировать ТЛС как устойчивую или неустойчивую к конкретному типу структурного разрушения; сформулированы принципы построения устойчивых ТЛС, среди которых ключевым является наличие альтернативных путей с эффективностью, близкой к оптимальной. Полученные результаты формируют фундамент для разработки интеллектуальных транспортно-логистических систем, способных эффективно противостоять сбоям и колебаниям нагрузки. **Ключевые слова:** транспортно-логистическая система; структурная устойчивость; теория графов; многокритериальная оптимизация; коэффициент просачиваемости; сетевое моделирование

**Для цитирования:** Yatskin D.V., Kochkarov A.A., Okuneva E.A. Assessment of the structural stability of transport and logistics systems based on graph models and the percolation coefficient. *Цифровые решения и технологии искусственного интеллекта*. 2026;2(1):52-62. DOI: 10.26794/3030-7097-2026-2-1-52-62

## INTRODUCTION

The past decade has marked a new phase in global economic development, characterized by significant shifts in the international division of labor. These changes necessitate a fundamental review of development and modernization strategies for transport and logistics systems (TLS) at both global and regional levels, as the creation and modernization of TLS will be a key determinant of long-term regional economic development.

As market volumes expand, effectively solving increasingly complex logistics tasks — many of which are NP-hard problems [1] — becomes more critical. In practice, however, it is possible to formulate constrained problems that admit polynomial algorithms [2–4]. For each specific task, solution effectiveness is determined individually [5, 6]. Equally important is examining not only the effectiveness of solutions but also the system’s structural ability to deliver optimal outcomes — shifting the question from “how to find the optimum” to “how to build a system where solutions are optimal.”

Modern business requirements underscore the critical importance of supply chain sustainability [6]. Given frequent disruptions and their severe consequences, ensuring structural stability is paramount. Meanwhile, the integration of systems and technologies — a key trend in TLS evolution — gives rise to multi criteria decision making tasks underpinned by big data, amplifying the prevalence of NP-hard problems and demanding novel research methodologies [7].

In Russia, transport and logistics expenditures constitute approximately 20% of GDP, significantly exceeding the global average of about 11%, with potential annual savings of up to 180 billion USD if aligned with global benchmarks. Key development factors include warehouse class, the level of logistics

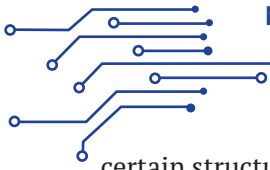
process optimization, and spatial uniformity of logistics centers.

This paper approaches TLS in an infrastructural sense, as a set of entities providing transportation, storage, and distribution of goods. The work aligns with research on managing structural dynamics of complex systems [7–9] and adapts methodological approaches originally developed for technical systems [10, 11] to analyze socio economic and organizational technical structures in TLS.

The primary objective is to develop a comprehensive toolkit for assessing the structural sustainability of TLS, addressing the following tasks: developing a graph theoretic model; formulating and solving a multi criteria optimization problem; determining structural failure processes; and calculating a quantitative stability coefficient. Stability analysis under network growth (e.g., expansion or scaling) falls outside the scope of this study and is reserved for future research [12].

## MATHEMATICAL DESCRIPTION OF THE TRANSPORT AND LOGISTICS SYSTEM

We propose a graph theoretic approach for modeling transport and logistics systems. In this framework, any such system can be represented as a network comprising nodes (vertices) and a set of pairwise connections (edges) between them. Practically, nodes may correspond to physical or functional entities such as transfer hubs, stations, warehouses, or key reference points within the logistics network. Since the system consists of a set of logistic nodes and connections between them, it is natural to describe it with a graph  $G(v, e)$ , where the vertices  $\{v\}$  are the nodes of the network, and the edges  $\{e\}$  are the connections (channels) between them. The main range of tasks is reduced to searching for



certain structures on the graph: flows, paths, routes, cycles, etc.

Considered a graph  $G = (V, E)$  in which there are two types of vertices:  $\{v \in V\}$  and  $\{e \in E\}$ , where

$$\{v^1 \in V^1\} \text{ и } \{v^2 \in V^2\}, V^1 \cap V^2 = \emptyset, V^1 \cup V^2 = V.$$

The vertices of the first type  $\{v^1\}$  correspond to important nodes in solving typical problems (for example, the source/drain or the start/end point). They may not be key to a particular task, but they have the potential to become one. The remaining vertices  $\{v^2\}$  are vertices of the second type. The vertex type is invariant with respect to particular problems and can change only for systemic and structural reasons.

In the graph  $G = (V, E)$  the edges model the cargo transfer channels. Two vertices  $v_i^1$  and  $v_j^2$  can be connected by several multiple edges  $\{e_{12}^{(1)}, e_{12}^{(2)}, \dots\}$ . The most important characteristics of edges are vector weight and throughput, which can be deterministic or non-deterministic. The weight reflects the resources for communication (cost, time, etc.). Bandwidth sets channel limits (number of shipments, total weight of cargo, etc.).

In mathematical terms, a product is formalized as an element of a task. A set of  $M$  elements of the problem is considered. For this set of problem elements, we define a collection of  $P$  features, which correspond to the weight coefficients assigned to the graph edges. The initial conditions of the system are represented by a four dimensional matrix  $I$ , whose elements are denoted as  $I_{ijmp}$ , where

$$i = \overline{1, N}, j = \overline{1, N}, m = \overline{1, M}, p = \overline{1, (P+1)}.$$

Indexes  $i$  and  $j$  define the edges,  $m$  — is the element of the problem, and  $p$  is the feature (the bandwidth of the edge for the element is added to the initial  $P$  features). The flow of task elements is denoted by  $f$ . The optimization criterion is written as  $Z(I, f)$ , and the task is represented as  $Z(I, f) \rightarrow opt$ .

For each flow  $f$  there is a solution  $s(I, f)$ , which is a graph structure (usually a path in a graph). If a different path is searched for each element, the task degenerates into  $M$  separate subtasks. When comparing different solutions, the value of the objective function  $Z(s)$  is evaluated. The optimal solution  $s_{opt}$  corresponds to the maximum achievable value of the criterion, denoted as  $Z_{opt} = Z(s_{opt})$ . The efficiency coefficient  $k_{eff}$  of any solutions  $s$  is defined as the ratio of its objective function value to that of the optimal solution:

$$k_{eff} = \frac{Z(s)}{Z_{opt}}.$$

The tasks in the system are divided into four groups:

1. The task of finding the (optimal) path between two key points.
2. The task of finding the optimal flow through the network, taking into account the bandwidth of the edges.
3. The combined task of taking into account the weights of the edges and their throughput.
4. Other tasks that do not require network modeling.

The first three groups are called *basic tasks*. The most common task is the third group, therefore, in the future, the *basic task* is understood as it is.

The weight and throughput of the edges are vector-based and may be non-deterministic. Structural disruptions can change the solutions to basic problems. It is assumed that the solution existing on the original graph structure is optimal (the reference). Comparison of other solutions is carried out with him.

## PERCOLATION THEORY

Percolation Theory<sup>1</sup> has not previously been used to describe transport and logistics problems [13]. This paper suggests a new approach that may be promising for global transportation planning.

One or two central vertices are highlighted on the graph as the source (for example, large network nodes) [14]. The peripheral vertices corresponding to the recipients of the goods are considered drains. This representation is applicable both for the tasks of finding the least expensive path and for optimizing flows.

The flow path  $P(V)$  of the graph  $G = (V, E)$  is defined as the shortest directed path from the central vertex to a given peripheral vertex  $v \in V$ . For each vertex  $v$ , the set of all such flow paths is finite and is denoted by  $V_p$ . Correspondingly, the stream chain  $S(V)$  is defined as the shortest undirected path from the central vertex to a peripheral vertex  $v$ . The finite set of all stream chains for a given vertex is denoted by  $V_s(v)$ . The Percolation coefficient is defined as

$$k_p = \frac{|V_p|}{|V_s|}.$$

If  $k_p > 0$  there is at least one streaming path, i.e. delivery is possible to one destination. At  $k_p = 1$  the flows reach all drains. The coefficient can vary over time due to structural changes (congestion, disasters) and is defined for different classes of graphs. This approach allows us to describe the problem using a graph model, which can be used to solve various classes of

<sup>1</sup> Work on network connectivity metrics, developing ideas close to the theory of infiltration.



problems, including the multi-criteria maximum flow problem, where the percolation coefficient becomes the key criterion.

### THE UTILITY FUNCTION OF THE SOLUTION

Each system is designed to solve specific tasks; the processes on it influence decisions. The most important process is changing the weights of the edges, since structural changes are a special case of it.

Two types of tasks are considered:

1. Minimizing resource costs:  $f(x) \rightarrow \min$ , where  $x$  is a vector set of features (time, cost, etc., possibly non-deterministic).

2. Load optimization:  $g(y) \rightarrow \min$ , where  $y$  is a vector of flows and throughput. The minimum corresponds to a uniform distribution.

In practice, simultaneous solutions are more often required, which leads to a third type of problem. The utility function of the solution is introduced, which is a combination of optimization functions with weighting coefficients:

$$P(x, y) \rightarrow \max.$$

The usefulness of the optimal solution is:

$$P_{opt} = \max_{x,y} (P(x, y)).$$

The utility coefficient of the solution found is

$$k_u = \frac{P}{P_{opt}}.$$

The dynamics of all key system processes are modeled via changes in vertex and edge parameters, including their weights and throughput. Each such change can be decomposed into a finite sequence of elementary transformations applied to the graph  $G$ .

Resistance to structural damage is being investigated. Any change in the structure entails a change in the matrix  $I$  (the weights of the edges of the complete graph). The correspondence of elementary structural changes to the modifications of matrix  $I$  is shown in Table.

Structural destruction is a set of elementary actions of removing vertices and edges of a graph. Deleting a vertex entails deleting incident edges.

A change in the structure leads to a change in the set of solutions  $\{s\}$  and the optimal solution  $s_{opt}$  relative to the criterion  $Z(s)$ . If a structural modification causes the optimal solution to shift from  $s_1$  to  $s_2$ , the coefficient of structural influence  $k_{inf}$  is defined as the ratio of their corresponding objective function values:

$$k_{inf} = \frac{Z(s_2)}{Z(s_1)}.$$

For processes representing structural degradation or removal (e.g., vertex/edge deletion), we have  $Z(s_2) \leq Z(s_1)$ , which implies  $k_{inf} \in [0, 1]$ . The magnitude of  $k_{inf}$  reflects the severity of the disruption: if it remains within a predefined acceptable threshold, the system can be considered structurally resilient to the given impact.

Table

#### Correspondence of Structural Changes to Changes in the Adjacency Matrix

No.	Structural Change	Corresponding Change in Matrix $I$
1	Removal of vertex $v_i$	Modification of all values $I_{ijmp}$ to 0 or $\infty$ depending on the semantic meaning of the corresponding value $p$
2	Addition of vertex $v_i$	Modification of all values $I_{ijmp}$ (from 0 or $\infty$ depending on the semantic meaning $p$ ) to actual values
3	Removal of edge $e_{ij}$	Modification of values $I_{ijmp}$ to 0 or $\infty$ depending on the semantic meaning of the corresponding value $p$
4	Addition of edge $e_{ij}$	Modification of all values $I_{ijmp}$ (from 0 or $\infty$ depending on the semantic meaning $p$ ) to actual values
5	Change in edge $e_{ij}$ throughput	Modification of values $I_{ijmp}$ (to corresponding updated values)
6	Change in vector weight of edge $e_{ij}$	Modification of values $I_{ijmp}$ (to corresponding updated values)

Source: compiled by the authors.

Calculating the efficiency coefficient  $k_{eff}$  with a modified structure allows us to judge the stability of a particular solution and the system as a whole. The method also allows you to study the dependence of efficiency on the flows of task elements, estimating the allowable load.

The approach has the potential to solve the inverse problem of building systems that are resilient to changes in the usefulness of solutions during structural changes or load scaling. The paper examines the structural stability of the model to external influences (for example, failures of nodes or channels, which corresponds to the removal of vertices or edges). This will make it possible to assess the viability of existing systems and formulate requirements for creating new ones that are optimally sustainable.

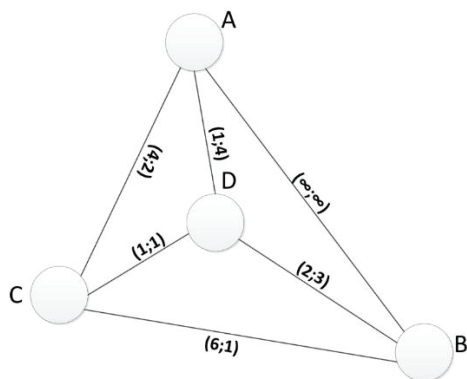
### A NUMERICAL EXAMPLE OF CALCULATING THE EFFICIENCY COEFFICIENT OF SOLUTIONS

A primary challenge in formulating the optimization problem lies in determining appropriate weighting coefficients for each resource. For illustration, consider two key resources: time  $T$  and transportation cost  $C$ . In this case, the optimization function can be expressed as  $f(x) = TC \rightarrow \min$ . Within the graph model, the total time and cost of delivery are computed as the sum of the corresponding values along the edges of the selected path. Accordingly, the utility function takes the form

$$(T, C) = \frac{C_0}{TC}, \text{ where } C_0 = \text{const.}$$

To demonstrate, we analyze a system with predefined edge weights representing cost and time. The objective is to determine the optimal path from vertex A to vertex B.

The path ADB is identified as optimal. The proposed method enables comparison of alternative solutions against this optimum via the efficiency coefficient  $k_{eff}$ .



For instance, the path ACB exhibits a 30% reduction in efficiency relative to the optimal solution, while the path ADCB incurs efficiency losses exceeding 50%. Fig. 1 shows the conditions of the problem and the calculation of the efficiency coefficient for different solutions.

Next, the combined task of optimizing the flow, taking into account the bandwidth, is considered. There are two goods; the capacity of the edges is set in the format  $(c_1, c_2)$ . It is necessary to deliver a batch  $(t_1, t_2)$  from A and B. The bandwidth of the path is  $c = (c_1, c_2)$ . The optimization function for the flow is defined as:

$$g(c_i) = \begin{cases} 1, & c_i \geq t_i; \\ \lceil t_i / c_i \rceil + 1, & c_i < t_i, \end{cases}$$

where  $g(c_i)$  is the component of the vector  $g(c)$  corresponding to the  $i$ -th product.

The utility function for a correct solution (for incorrect ones,  $P = 0$ ) has the form:

$$P(T, C, c) = \frac{840}{\sum_{i=1}^2 (TC)_i \cdot g(c_i)},$$

where  $(TC)_i$  may be different for each product if they go their separate ways. Fig. 2 shows the conditions of the problem and calculations of the utility function for different solutions.

The optimal solution is to send product 1 via ACB, and product 2 via ADB. Then:

$$P_{opt} = \frac{840}{3 \cdot 7 \cdot 2 + 10 \cdot 3 \cdot 1} = 11.7.$$

Both items cannot be shipped via ACB ( $P = 0$ ). When sending both over ADB:

$$P = 10, k_{eff} = \frac{10}{11.7} = 0.86.$$

Path	$C$ , cost	$T$ , time	$P(x) = \frac{840}{TC}$	$k_{eff}$
AB	$\infty$	$\infty$	0	—
ADB	3	7	40	1
ACB	10	3	28	0.7
ADCB	8	6	17.5	0.4375

Fig. 1. The Task of Finding the Optimal path From Vertex A to Vertex B

Source: Author's calculations.



## CONSTRUCTION OF TRANSPORT AND LOGISTICS SYSTEMS WITH SPECIFIED PARAMETERS OF STRUCTURAL STABILITY

The concept of structural stability makes sense only with respect to a specific task  $Z(I, f) \rightarrow opt$  and a given barrier value of  $k_{bar}$  which allows us to assess the impact of changes on the solution.

The opposite problem is relevant – the construction of systems resistant to structural damage. With a given  $k_{bar}$  for each flow  $f$  the stability of system  $I$  can be estimated. Flow  $f$  determines the departure/destination nodes and cargo volumes. The function  $Z(I, f) \rightarrow min$  is non-decreasing with increasing traffic volume. If the system is stable for flow  $f$ , then it is stable for flow  $f_0$  with smaller or equal volumes.

In practice, it is important to determine the maximum flows of goods that the system can handle. To assert the stability of system  $I$  at given maximum volumes and  $k_{bar}$  it is necessary to investigate solutions  $Z(I, f) \rightarrow min$  and estimate  $k_{inf}$  for all possible – routes.

A large-scale computer experiment was conducted involving the random generation of over 1 million instances of the initial condition matrix  $I_{ijmp}$  for graphs comprising 50 to 100 vertices. The results indicate that a significant proportion of the generated systems demonstrate resilience to elementary structural damage, defined as the removal of individual edges or vertices. Specifically, with a stability threshold set at  $k_{bar} = 0.81$ , more than 88% of the systems were classified as resistant.

The highest degree of stability was observed in systems possessing one or more vertex disjoint alternative solutions, where the objective function value  $Z(s)$  of each such solution deviated from the optimal

value by no more than 5.1%. Remarkably, among systems featuring two or more edge disjoint solutions, the proportion of stable configurations reached 99.94%.

## INVESTIGATION OF THE CONGESTION FORMATION PROCESS

The introduced definitions make it possible to investigate the relationship between the characteristics of the system and structural changes with the coefficient of infiltration  $k_p$ . This makes it possible to assess the impact of congestion (which changes the capacity of edges) on the effectiveness of solutions.

For the experiment, a model of a connected network of 100 nodes was generated, described by a complete graph (the bandwidth of some edges is 0, which corresponds to a lack of connectivity). Congestion can lead to zero throughput at a discrete point in time, which is equivalent to removing an edge.

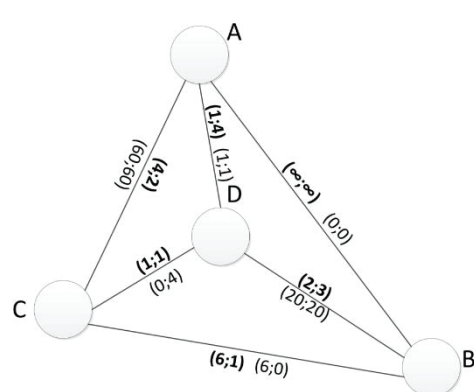
Next, random iterative changes in edge weights are modeled in the direction of increasing (decreasing throughput), corresponding to the formation of congestion. The model is anisotropic: the process of weight reduction is reversed without loss of accuracy, which makes it possible to describe both the formation and “resolution” of congestion. For the same limit values, the difference in averages is statistically insignificant.

Structure  $I$  covers 9900 typical delivery tasks with different departure and destination nodes. For each task,  $Z_{opt} = Z(s_{opt})$  and  $k_p$  are defined.

At each iteration,  $Z(s)$  for the new optimal solution and  $k_p$  are calculated.

The result of the simulation is the value of

$$k_{eff} = \frac{Z(s)}{Z_{opt}}$$



Path	C, cost	T, time	$c_1$	$c_2$	$\frac{840}{TCg(c_1)}$	$\frac{840}{TCg(c_2)}$	P
AB	$\infty$	$\infty$	0	0	0	0	0
ADB	3	7	1	1	20	20	10
ACB	10	3	6	0	28	0	0
ADCB	8	6	0	0	0	0	0

Fig. 2. The Combined Task of Finding the Optimal Path from Vertex A to Vertex B

Source: Author’s calculations.



veraged over all tasks. A correlation was found between the average throughput of the edges (normalized  $n$ ) and the values of  $k_p$  (Fig. 3).

The dependence of  $k_p$  on  $n$  is well approximated by a third-degree polynomial (correlation coefficient 0.976). With this level of correlation, it is possible to study the dependence of  $k_{eff}$  and  $k_p$  without taking into account the detailed throughput of individual edges, which simplifies calculations (Fig. 4).

At first, the change in  $k_p$  has little effect on  $k_{eff}$  then there is a sharp exit to a plateau ( $k_p$  range from 0.4 to 0.65), where  $k_{eff}$  stabilizes, after which it increases again.

The experiment was repeated for different configurations of parameters and graphs. The graph type is preserved, the plateau is always the same. The values of  $k_{eff}$  on the plateau lie within the boundaries of [0.615; 0.72]. This allows us to draw a parallel with

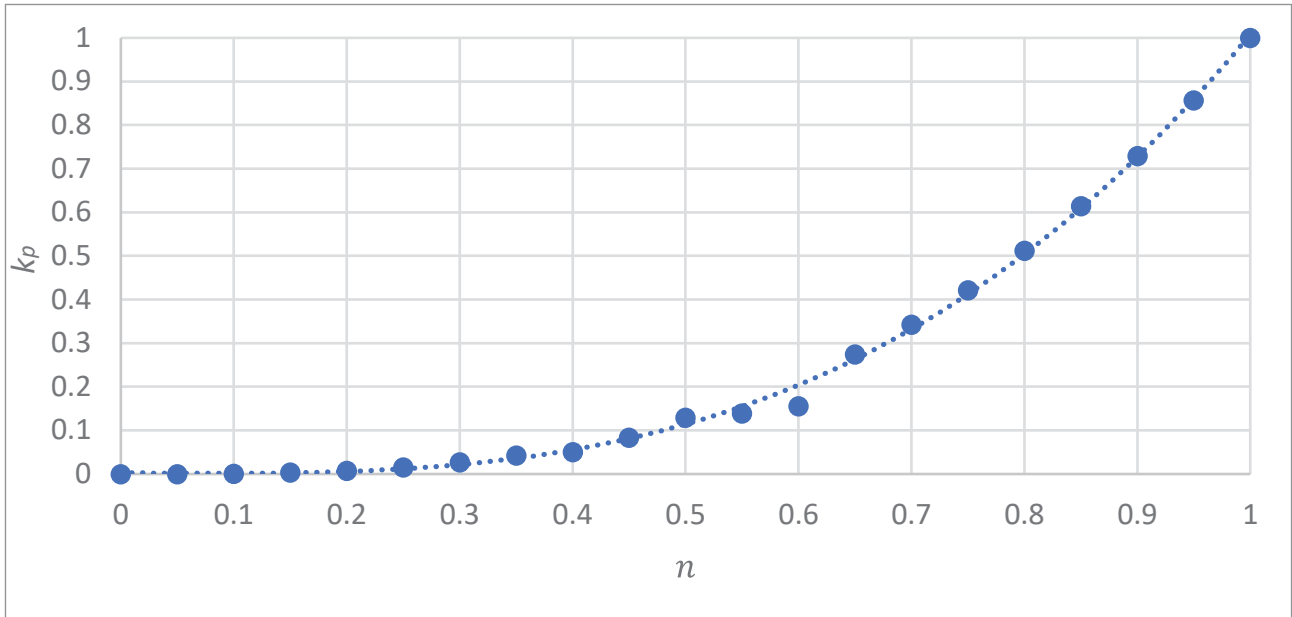


Fig. 3. Dependence of  $k_p$  on the Normalized Average Throughput of Edges  $n$

Source: Author's calculations.

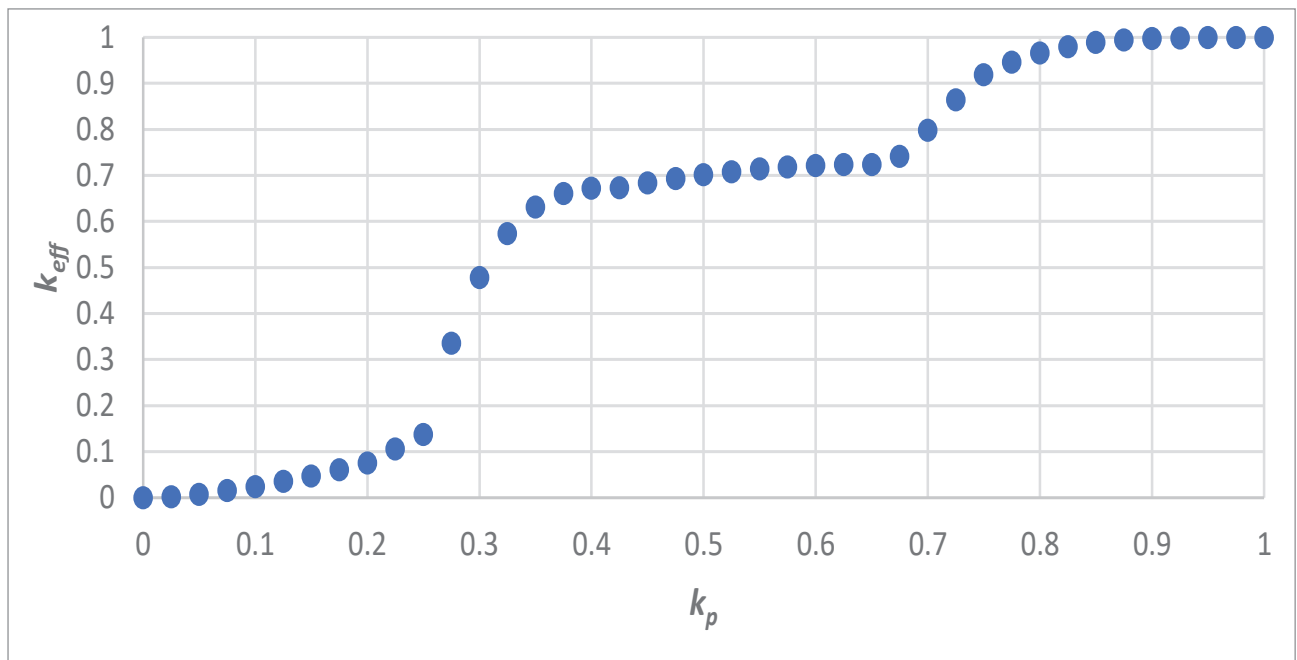


Fig. 4. Dependence of  $k_{eff}$  on  $k_p$

Source: Author's calculations.

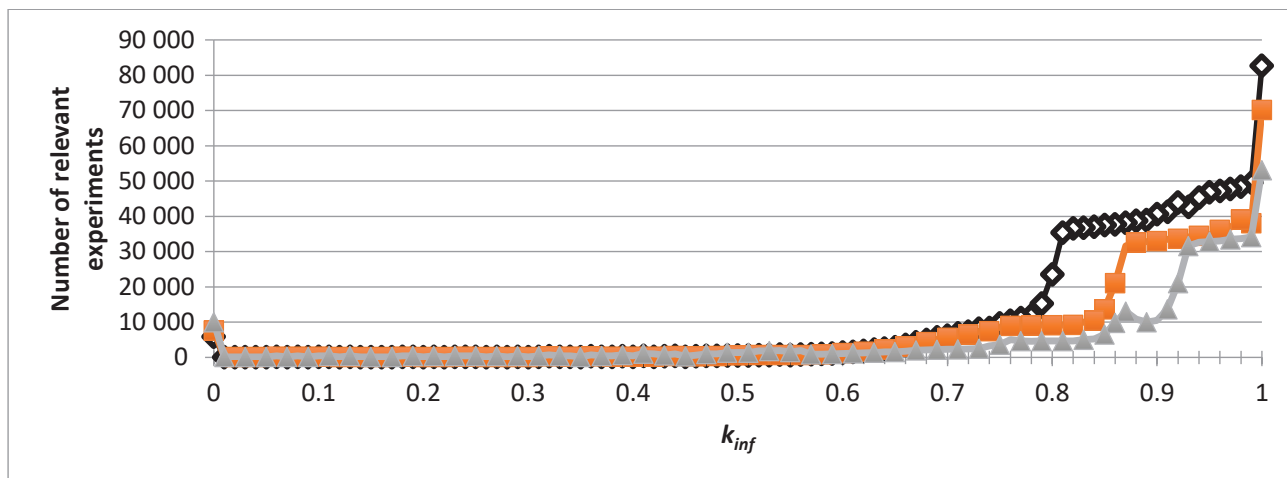


Fig. 5. Number of Experiments and  $k_{inf}$ .

Source: Author's calculations.

Note: Upper graph – removal of one element; Middle graph – removal of two elements; Lower graph – removal of three elements.

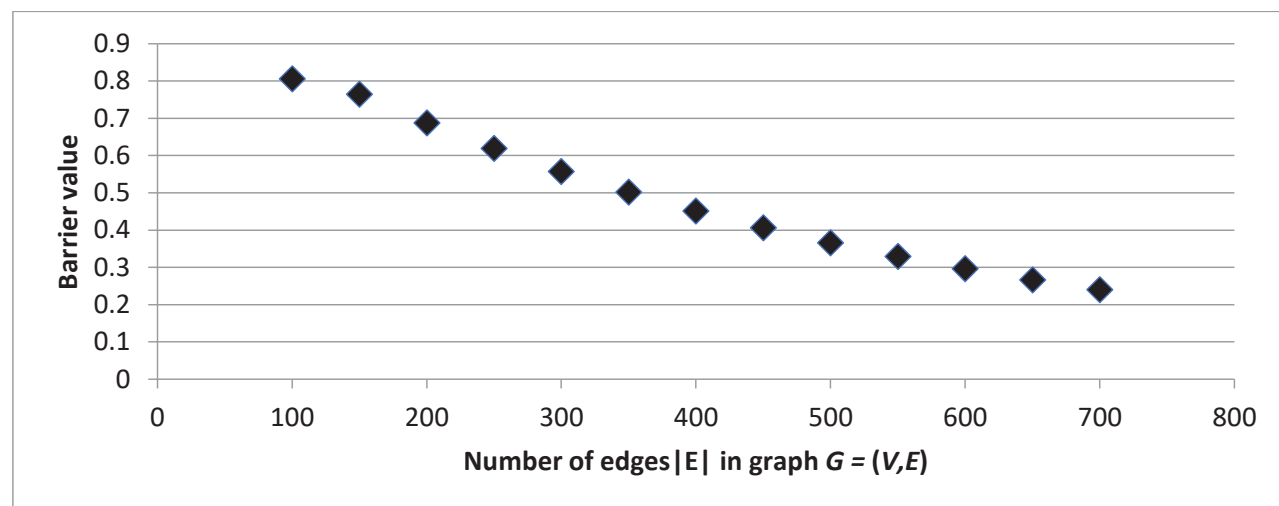
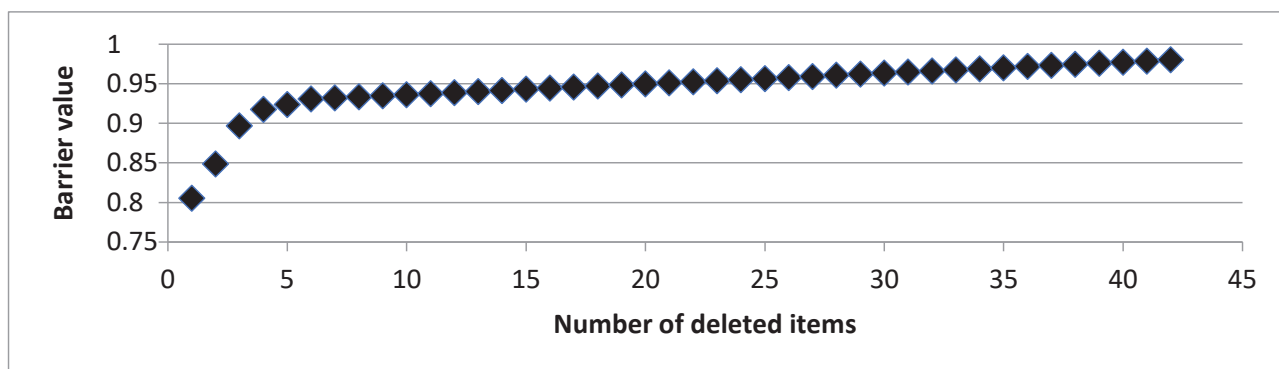


Fig. 6. Number of Experiments and values  $k_{inf}$

Source: Author's calculations.

Note: The upper graph is the dependence of the barrier value on the number of elements to be removed. The lower graph is the dependence of the barrier value on  $|E|$ .

the barrier values of the coefficient of resistance to structural damage, since changes in  $k_p$  are associated with changes in the throughput of the edges, and structural destruction is their degenerate case [15].

Thus, the model allows not only to assess the impact of changes in throughput on problem solving, but also to make a priori forecasts of efficiency changes with characteristic changes in throughput.



## COMPUTER MODEL CALCULATIONS

To determine the barrier value of  $k_{bar}$  a computer experiment was conducted to model transport and logistics systems and structural failure processes. 1,050,000 structures were generated with graphs  $G(V, E)$  of size  $|V|=100$ , matrices of initial conditions  $I$  and optimization criteria  $Z(s)$ .

The structural resistance to the removal of one element (vertex or edge) was studied using the coefficient of influence  $k_{inf}$ . The dependence of the number of systems on the value of  $k_{inf}$  is shown in Fig. 5. Jumps at  $k_{inf} = 0$  and  $k_{inf} = 1$  correspond to a complete loss of solvability or lack of influence on the optimal solution.

The appropriate range of acceptable values of  $k_{inf}$  for stability is considered to be  $[0.81; 1]$ . In 83.25% of the experiments  $874\ 114\ k_{inf} \geq 0.81$ . The point 0.81 corresponds to a loss of efficiency of up to 19%. In the range  $[0.81; 0.83]$  there is a sharp jump in the derivative, which divides all systems into two groups. Thus,  $k_{inf} = 0.81$  is recommended as a barrier value for destruction of the “removal of one element” type. Systems with  $k_{inf} < 0.81$  are considered unstable.

Additionally, resistance to removal of two and three elements was investigated. The barrier values were 0.86 and 0.90, respectively. The dependence of the barrier value on the number of elements to be deleted and the size of the graph is shown in Fig. 6: the barrier value increases with the number of elements to be deleted and decreases with the size of the graph.

The coefficients  $k_{inf}$  and  $k_{bar}$  provide a numerical measure of structural stability, allow us to analyze specific systems and solve the inverse problem of finding the optimal structure resistant to specified damage.

## CONCLUSION

This paper develops a comprehensive methodological framework for studying transport and logistics systems (TLS), integrating principles from optimal control theory, multi criteria optimization, and graph theory. We introduce key formal concepts, including structural graph destruction, the combined transport logistics problem, and two central quantitative metrics: the solution efficiency coefficient  $k_{eff}$  and the structural influence coefficient  $k_{inf}$ .

A series of computer simulations modeling various structural damage scenarios were conducted. The experimental results enabled us to characterize the distribution of  $k_{inf}$  for elementary failure processes and to establish a threshold for acceptable performance degradation. Specifically, a system is considered to remain in a stable state if  $k_{inf} \geq 0.81$  for the removal of a single network element (vertex or edge).

The theoretical significance of this work lies in its potential application to the design and analysis of resilient transport and logistics networks. The practical value stems from the development of accessible tools for assessing the structural sustainability of existing systems and for identifying critical vulnerabilities.

The results lay the foundation for further research on the dependence of stability on the characteristics of systems and the maximum volume of commodity flows. The proposed approach allows us to solve the inverse problem of building systems with a given structural stability, as well as to investigate the effectiveness of solutions depending on the load.

The prospect is to test the model on real data from logistics companies, as well as expand it to a multiparametric formulation with fuzzy weights and develop appropriate algorithms. The model is considered as an element of future intelligent transport and logistics systems in demand in the global transportation market.

## REFERENCES

1. Costa Y., Duarte L., Pereira R. The impact of global supply chain reconfiguration on international trade patterns: A network analysis. *Research in Transportation Business & Management*. 2024;52:100811. DOI: 10.1016/j.rtbm.2024.100811
2. Gavish B., Graves S.C. The Travelling Salesman Problem and Related Problems. *INFORMS Journal on Computing*. 2023;35(2):319–335. DOI: 10.1287/ijoc.2022.0123
3. Cota P.M., de Lima F.S., de Araújo O.C.B. A two-stage integer programming model for cargo allocation with transshipment time considerations. *Computers & Industrial Engineering*. 2023;175:108852. DOI: 10.1016/j.cie.2023.108852
4. Roy S.K., Mula P. A stochastic transportation problem with multiple choice of cost and requirement. *International Journal of Operational Research*. 2022;44(3):365–382. DOI: 10.1504/IJOR.2022.120211
5. Pournader M., Kach A., Fahimnia B. A Review of Transport Network Resilience: Concepts, Models, and Future Research Directions. *Transport Reviews*. 2023;43(4):689–715. DOI: 10.1080/01441647.2023.2201771
6. Ivanov D., Dolgui A. Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. *International Journal of Production Research*. 2022;60(1):1–17. DOI: 10.1080/00207543.2020.1750727



7. Dubey R., Gunasekaran A., Bryde D.J., Dwivedi Y.K., Papadopoulos T. Blockchain technology for enhancing supply chain resilience and sustainability. *International Journal of Production Research*. 2023;61(7):2266–2285. DOI: 10.1080/00207543.2022.2101771
8. Dmitriev A., Shnurkov M. Graph Theory Application in Modern Supply Chain Network Design: A Systematic Review. *Transportation Research Part E: Logistics and Transportation Review*. 2023;175:103–145. DOI: 10.1016/j.tre.2023.103145
9. Zhang Y., Li X., Wang Q. Modeling Multimodal Transportation Networks as Directed Hypergraphs for Enhanced Logistics Planning. *Computers & Industrial Engineering*. 2024;187:109821. DOI: 10.1016/j.cie.2024.109821
10. Chen, L., Pietrabissa, A. A Robust Network Flow Model for Strategic Logistics Hub Classification under Demand Uncertainty. *European Journal of Operational Research*. 2022;301(1):283–297. DOI: 10.1016/j.ejor.2021.09.021
11. Garcia-Herrero S., Mar-Ortiz J., & De-La-Peña J.L. An Integrated Multi-Objective Optimization Model for Green Vehicle Routing and Scheduling with Heterogeneous Fleet and Time-Dependent Capacities. *Journal of Cleaner Production*. 2023;414:137503. DOI: 10.1016/j.jclepro.2023.137503
12. Ivanov D., Dolgui A. A Digital Supply Chain Twin for Managing the Disruption Risks and Resilience in the Era of Industry 4.0. *Production Planning & Control*. 2021;32(9):775–788. DOI: 10.1080/09537287.2021.1922771
13. Newman M.E.J. A Measure of Structural Redundancy for Semantically Annotated Networks. *Scientific Reports*. 2022;12(1):13487. DOI: 10.1038/s41598-022-17191-81
14. Katarzyniak M., Mulawka D. Identification of Critical Nodes in Transport Networks Using Centrality Measures and Vulnerability Scenarios. *IEEE Access*. 2023;11:45678-45691. DOI: 10.1109/ACCESS.2023.3281171
15. Ghavamifar A., Saberi M. Robustness of Interdependent Logistics Networks: A Percolation-Based Analysis of Cascade Failures. *Transportation Research Part E: Logistics and Transportation Review*. 2024;181:103367. DOI: 10.1016/j.tre.2024.103367

## ABOUT THE AUTHORS / ИНФОРМАЦИЯ ОБ АВТОРАХ

**Danil V. Yatskin** — Leading Expert at the Institute for Statistical Research and Economics of Knowledge, Higher School of Economics, Moscow, Russian Federation

**Данил Владиленович Яцкин** — ведущий эксперт, Институт статистических исследований и экономики знаний НИУ ВШЭ, Москва, Российская Федерация  
<https://orcid.org/0000-0003-2914-0907>  
[d.iatskin@gmail.com](mailto:d.iatskin@gmail.com)

**Azret A. Kochkarov** — Dr. Sci (Tech.), Deputy Director for Innovation at the Fundamental Principles of Biotechnology Research Center of the Russian Academy of Sciences, Moscow, Russian Federation

**Азрет Ахматович Кочкаров** — доктор технических наук, заместитель директора по инновационной работе, ФИЦ «Фундаментальные основы биотехнологии» РАН, Москва, Российская Федерация  
<https://orcid.org/0000-0002-3232-5331>  
[akochkar@fbras.ru](mailto:akochkar@fbras.ru)

**Evelina A. Okuneva** — assistant of the Department of Mathematics and Data Analysis, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

**Эвелина Александровна Окунева** — ассистент кафедры математики и анализа данных факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация  
<https://orcid.org/0009-0006-4385-4462>

*Corresponding author / Автор для корреспонденции:*  
[eaokuneva@fa.ru](mailto:eaokuneva@fa.ru)



**Authors' declared contributions:**

**D.V. Yatskin** — development of a general methodology and research concept, formalization of a mathematical model of a transport and logistics system based on graph theory, formulation of a multi-criteria optimization problem, definition of key indicators, writing and editing of the main sections of the article.

**A.A. Kochkarov** — development of algorithms, generation of test data, modeling of structural damage and overload processes, statistical analysis and visualization of the results, preparation of sections devoted to the results of modeling.

**E.A. Okuneva** — analytical review of literary sources, preparation of numerical examples and tables, formulation of practical conclusions and principles of building sustainable systems, preparation of a list of references, as well as participation in writing and editing the introductory and final parts of the article.

**Заявленный вклад авторов:**

**Д.В. Яцкин** — разработка общей методологии и концепции исследования, формализация математической модели транспортно-логистической системы на основе теории графов, постановка многокритериальной оптимизационной задачи, определение ключевых показателей, написание и редактирование основных разделов статьи.

**А.А. Кочкаров** — разработка алгоритмов, генерация тестовых данных, моделирование процессов структурных разрушений и перегрузок, статистический анализ и визуализация результатов, подготовка разделов, посвященных результатам моделирования.

**Э.А. Окунева** — аналитический обзор литературных источников, подготовка численных примеров и таблиц, формулировка практических выводов и принципов построения устойчивых систем, оформление списка литературы, а также участие в написании и редактировании вводной и заключительной частей статьи.

*Conflicts of Interest Statement: The authors have no conflicts of interest to declare.*

*Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.*

*The article was submitted on 19.01.2026; revised on 02.02.2026 and accepted for publication on 12.02.2026.*

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

*Статья поступила 19.01.2026; после рецензирования 02.02.2026; принята к публикации 12.02.2026.*

*Авторы прочитали и одобрили окончательный вариант рукописи.*



# Анализ подходов и решений для разработки системы хранения клиентских данных банковской экосистемы

Н.А. Бурыкин<sup>1</sup>, Н.В. Гринева<sup>1</sup>, П.Е. Голосов<sup>2</sup>

<sup>1</sup> Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация;

<sup>2</sup> Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Москва, Российская Федерация

## АННОТАЦИЯ

Статья посвящена анализу подходов и решений для разработки единого, масштабируемого и высоконагруженного хранилища клиентских данных в рамках цифровой экосистемы банка. В условиях цифровой трансформации и развития цифровых экосистем крупные организации, в том числе банки, сталкиваются с тем, что хранение и обработка клиентских данных становится критически важным вопросом для обеспечения бизнес-процессов компании, персонализации сервисов для клиентов, соответствия регуляторным требованиям, поддержания конкурентного преимущества. **Цель исследования** — определить оптимальный набор архитектурных принципов и технологических решений для построения масштабируемой, отказоустойчивой и высокопроизводительной системы хранения клиентских данных. **Методы исследования:** системный анализ и синтез — для изучения специфики предметной области банковской цифровой экосистемы; сравнительный анализ — для оценки технологий хранения данных по следующим критериям: производительность, масштабируемость, уровень согласованности данных, отказоустойчивость; анализ научной литературы и информационных источников — для изучения предметной области, современных решений и технологий в системах хранения данных. **Результаты исследования:** выбраны оптимальные паттерны и технологии для реализации платформы хранения клиентских данных; сформулированы рекомендации по проектированию архитектуры единого клиентского хранилища, способного эффективно функционировать в динамичном ИТ-ландшафте современной банковской экосистемы. **Выводы.** Результаты исследования подтверждают, что использование гибридных хранилищ данных с применением различных технологий и решений наиболее оптимально для построения высоконагруженных и масштабируемых систем хранения и обработки данных (что также согласуется с выводами других научных работ). **Перспективы исследования** включают дальнейшее углубленное изучение интеллектуальных систем управления данными; технологий хранения и обработки данных, используемых в машинном обучении и ИИ; подходов Data Mesh и Data Fabric; возможностей применения перечисленных технологий в банковской сфере. **Ключевые слова:** банковская экосистема; клиентские данные; высоконагруженные системы; NoSQL; реляционные базы данных; In-Методу хранилища; архитектура информационных систем

**Для цитирования:** Бурыкин Н.А., Гринева Н.В., Голосов П.Е. Анализ подходов и решений для разработки системы хранения клиентских данных банковской экосистемы. *Цифровые решения и технологии искусственного интеллекта.* 2026;2(1):63-72. DOI: 10.26794/3030-7097-2026-2-1-63-72

# Analysis of Approaches and Solutions for Developing a Customer Data Storage System for the Banking Ecosystem

N.A. Burykin<sup>1</sup>, N.V. Grineva<sup>1</sup>, P.E. Golosov<sup>2</sup>

<sup>1</sup> Financial University under the Government of the Russian Federation, Moscow, Russia;

<sup>2</sup> Russian Presidential Academy of National Economy and Public Administration, Moscow, Russia

## ABSTRACT

The article is devoted to the analysis of approaches and solutions for the development of a single, scalable and highly loaded customer data warehouse within the bank's digital ecosystem. In the context of digital transformation and the development of digital ecosystems, large organizations, including banks, are faced with the fact that the storage and processing of customer data is becoming a critical issue for ensuring company processes, personalization of services for customers, compliance with regulatory requirements and maintaining a competitive advantage. The purpose of the research is to determine the optimal set of architectural principles and technological solutions applicable to building a scalable, fault-tolerant and high-performance customer data storage system. The following methods were used to conduct the research: system analysis and synthesis to study the specifics of the subject area of the banking digital ecosystem, comparative analysis to evaluate data storage technologies based on criteria such as performance, scalability, data consistency, fault tolerance, and analysis of scientific literature and information sources to study the subject area, modern solutions and technologies in systems data storage.

Based on the conducted research, optimal patterns and technologies for implementing a customer data storage platform have been selected and recommendations have been formulated for designing a single customer storage architecture capable of operating effectively in the dynamic IT landscape of the modern banking ecosystem. Based on the results of the study, it was concluded that other scientific papers and articles also confirm that the use of hybrid data warehouses using different technologies and solutions is most optimal for building highly loaded and scalable data storage and processing systems. The prospects of the research include further in-depth study of intelligent data management systems, data storage and processing used in machine learning and AI, Data Mesh and Data Fabric approaches and their application in the banking sector.

**Keywords:** banking ecosystem; client data; high-load systems; NoSQL; relational databases; In-Memory storage; information systems architecture

**For citation:** Burykin N.A., Grineva N.V., Golosov P.E. Analysis of approaches and solutions for developing a customer data storage system for the banking ecosystem. *Digital solutions and artificial intelligence technologies*. 2026;2(1):63-72. DOI:10.26794/3030-7097-2026-2-1-63-72

## ВВЕДЕНИЕ

Современные банковские цифровые экосистемы находятся в состоянии цифровой трансформации, в рамках которой данные о клиентах превращаются из пассивного актива в ключевой стратегический ресурс, что позволяет создавать персонализированные сервисы, управлять рисками и обеспечить кредитной организации конкурентное преимущество. Процесс связан с обработкой растущих объемов различной информации. Помимо структурированных транзакционных данных, банки вынуждены анализировать массивы неструктурированных и слабоструктурированных данных, включая поведенческую аналитику, данные о взаимодействии и данные со сторонних платформ экосистемы [1]. Задача усложняется ужесточением регуляторных требований и растущих ожиданий клиентов в отношении скорости, доступности и персонализации сервисов в режиме 24/7.

В таких условиях классические централизованные и монолитные архитектуры систем хранения данных демонстрируют свою несостоятельность, становясь «узким местом» в развитии экосистемы. Их ограниченная масштабируемость, высокая стоимость владения и негибкость при интеграции новых источников данных формируют принципиально новые требования к инфраструктуре, где главными задачами становятся: обеспечение бесперебойной работы в условиях пиковых нагрузок, обработка данных в реальном времени, гарантированная согласованность и безопасность при распределенной обработке. Это соответствует принципам построения высоконагруженных отказоустойчивых систем, подробно исследуемых в фундаментальных работах по проектированию программного обеспечения.

**Актуальность данного исследования** обусловлена острой необходимостью в систематизации и поиске оптимальных архитектурных решений для построения распределенных высоконагруженных систем хранения данных, способных удовлетворить требования цифровой банковской экосистемы. Как отмечают исследователи, современные системы должны сочетать в себе гарантии согласованности

ACID и гибкость принципов BASE для обработки больших данных и обеспечения высокой доступности [2]. Необходим комплексный подход, объединяющий достижения в области микросервисной архитектуры, контейнеризации, сегментирования (sharding) данных, а также синергетического использования различных классов систем хранения.

**Научная новизна** работы заключается в комплексном и системном анализе архитектурных паттернов и технологических решений для построения распределенных высоконагруженных систем хранения клиентских данных. Исследование предлагает целостную модель, рассматривающую проблему через призму интеграции разнородных систем хранения (SQL, NoSQL, in-memory), при одновременном соблюдении строгих требований к производительности, надежности и соответствию регуляторным нормам.

**Цель исследования** — разработка теоретико-методологического подхода и выбор решений для построения надежной высоконагруженной системы хранения данных клиентов с высокой производительностью, удовлетворяющей все требования к системам хранения клиентских данных в крупных организациях.

**Задачи исследования:** проанализировать предметные области и выявить специфические требования к современным системам хранения данных; выполнить сравнительный анализ современных подходов к организации распределенных систем хранения данных и выявить их применимость в контексте банковской цифровой экосистемы; оценить и систематизировать современные технологии хранения данных и подходы к разработке систем хранения клиентских данных.

## ИССЛЕДОВАНИЕ СУЩЕСТВУЮЩИХ ПОДХОДОВ К ХРАНЕНИЮ И ОБРАБОТКЕ ДАННЫХ ДЛЯ РАЗРАБОТКИ ВЫСОКОНАГРУЖЕННЫХ СИСТЕМ ХРАНЕНИЯ ДАННЫХ

Разработка высоконагруженной распределенной системы хранения для цифровой экосистемы банка



представляет собой комплексную задачу, требующую соответствия фундаментальным ограничениям, сформулированным в теореме CAP (Брюера). Согласно данной теореме, в распределенной системе невозможно одновременно обеспечить более двух из трех свойств: согласованности данных (Consistency), доступности (Availability) и устойчивости к разделению сети (Partition Tolerance) [3].

Этот принцип, как отмечают Таненбаум и ван Стейен, является краеугольным камнем при проектировании и делит все системы на два основных класса: CP-ориентированные (с акцентом на согласованность и устойчивость к разделению) и AP-ориентированные (с акцентом на доступность и устойчивость к разделению) [4].

Исторически основой для построения надежных систем хранения в финансовой сфере являлись реляционные СУБД (Oracle Database, PostgreSQL), основанные на принципах ACID (Atomicity, Consistency, Isolation, Durability — атомарность, согласованность, изолированность, долговечность). Они обеспечивают строгую транзакционную семантику, что критически важно для банковских систем, где целостность данных является неприкосновенной [5].

В ответ на вызовы масштабируемости и высокой доступности в эпоху Big Data и веб-сервисов сформировалась альтернативная парадигма BASE<sup>1</sup>. Данный набор принципов, подробно описанный Садаладжем и Фаулером, лежит в основе NoSQL-систем, которые жертвуют строгой согласованностью в пользу доступности и горизонтального масштабирования [5].

Однако, как подчеркивает Мартин Фаулер, при масштабировании такие системы сталкиваются с «бутылочным горлышком» в виде единой точки отказа и сложностями горизонтального масштабирования (sharding), что требует сложных процедур рещардинга и может негативно сказываться на производительности при высоких нагрузках [5].

Для различных моделей данных и сценариев использования разработаны специализированные типы хранилищ:

- «ключ-значение» (Redis или Apache Ignite) — применяются в высоконагруженных проектах для кэширования и работы с данными в оперативной памяти, требующими минимальной задержки;
- ширококолоночные (Apache Cassandra). Отсутствие единой точки отказа и предсказуемая производительность при линейном масштаби-

ровании делают их привлекательными для задач, требующих записи и чтения больших объемов данных с временными рядами или событийными журналами [6];

- документоориентированные (MongoDB или Yandex Database). YDB от Яндекса — отличный пример современной гибридной системы, поддерживающей как документную модель, так и реляционные запросы с распределенными транзакциями<sup>2</sup>.

Интересный тренд — конвергенция подходов, выраженная в появлении класса NewSQL-систем. Эти системы, такие как Google Spanner [7] и его открытые аналоги (CockroachDB и YDB), сочетают гарантии ACID реляционных баз данных с горизонтальной масштабируемостью NoSQL. В их основе лежат алгоритмы распределенного консенсуса, такие как Paxos или Raft, для управления репликацией данных, что обеспечивает строгую согласованность в распределенной среде [8].

Еще один важный архитектурный паттерн — CQRS (Command Query Responsibility Segregation — разделение ответственности команд и запросов) в сочетании с Event Sourcing (источники событий)<sup>3</sup>. CQRS разделяет модели для операций обновления (команд) и чтения (запросов), позволяя независимо масштабировать и оптимизировать каждую из них.

Event Sourcing предполагает сохранение не конечного состояния объекта, а последовательности всех событий, которые к этому состоянию привели. Это не только обеспечивает полный аудит изменений, что крайне важно для банковской деятельности, но и позволяет строить различные материализованные представления (проекции) данных, оптимальные для конкретных сценариев чтения.

На *рис. 1* схематически представлено разделение моделей данных с применением подхода CQRS, где можно увидеть изоляцию моделей данных на уровне хранилищ данных: выполнение запросов на запись — в одно хранилище с конкретной моделью данных, выполнение запросов на чтение — в другое хранилище с другой моделью данных<sup>4</sup>.

Активное развитие искусственного интеллекта (ИИ) и его внедрение в автоматизацию бизнес-процессов и производство в рамках цифровой банковской экосистемы сформировали новый подход к хранению и обработке данных — RAG (Retrieval-Augmented Generation), применяемый в интеллект-

<sup>2</sup> Документация YDB. URL: <https://ydb.tech/docs/ru/?version=v25.2>

<sup>3</sup> CQRS Documents. URL: [https://cQRS.wordpress.com/wp-content/uploads/2010/11/cQRS\\_documents.pdf](https://cQRS.wordpress.com/wp-content/uploads/2010/11/cQRS_documents.pdf)

<sup>4</sup> Там же.

<sup>1</sup> BASE. Basically Available, Soft state, Eventual consistency — фундаментальная доступность, неустойчивое состояние, потенциальная согласованность.

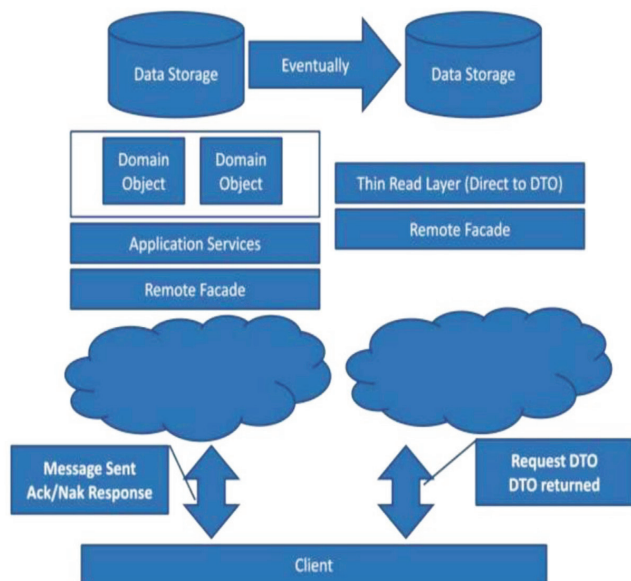


Рис. 1 / Fig. 1. Разделение моделей данных с применением подхода CRQS / Separation of Data Models Using the CRQS Approach

Источник / Source: документы CRQS / CQRS Documents.

туальных сервисах, где для извлечения информации используются запросы в виде натурального языка, сгенерированные большими языковыми моделями (LLM) или непосредственно человеком.

RAG разработан для повышения точности, актуальности и проверяемости выходных данных LLM за счет их обогащения релевантной информацией, извлеченной из внешних корпусов знаний (knowledge bases) [9]. Перед записью в RAG данные

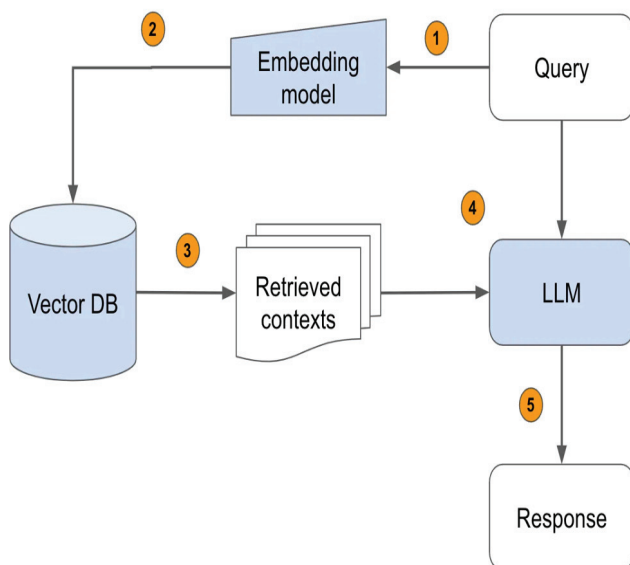


Рис. 2 / Fig. 2. Схематическое описание работы RAG / Schematic description of RAG operation

Источник / Source: RAG (Retrieval-Augmented Generation): основы и продвинутое техники\*.

Примечание / Note: \*URL: <https://habr.com/ru/articles/871226/>

индексируются путем преобразования в векторное представление данных (*embedding*) с помощью специализированных моделей машинного обучения.

Для извлечения данных из RAG запросы, полученные от пользователя (человек или LLM), также преобразуются в *embedding* для дальнейшего вычисления векторного расстояния между проиндексированными данными и поступившим от пользователя запросом — с целью извлечения информации, наиболее близкой по векторному расстоянию к поступившему запросу и смысловому значению.

Распространенными решениями векторных баз данных, применяемых в RAG, являются Elasticsearch, Qdrant, плагин PG Vector для СУБД PostgreSQL и Milvus.

На рис. 2 представлено схематическое описание работы RAG.

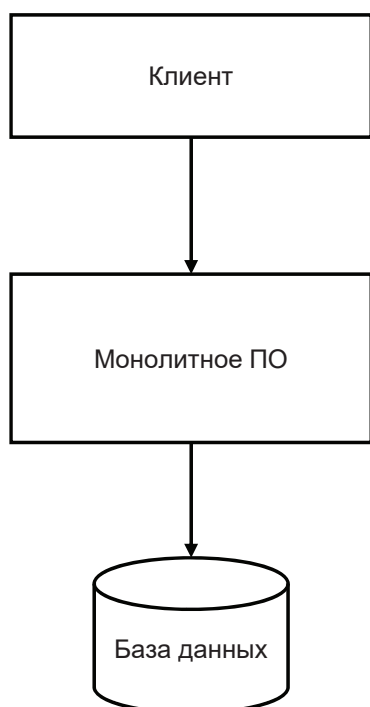
Помимо технологий и решений хранения данных необходимо также рассмотреть архитектурные паттерны для разработки распределенных систем хранения. Глобально можно выделить главные архитектурные подходы разработки программного обеспечения (ПО) — монолитная и микросервисная архитектуры.

Монолитная архитектура характерна тем, что вся логика приложения, включая пользовательский интерфейс, бизнес-логику и доступ к данным, находится внутри одной единицы развертывания приложения. При монолитной архитектуре разработка, тестирование и развертывание монолитного приложения может быть относительно проще, так как все части находятся в одном месте из-за чего изменение или добавление новых функций может быть сложным, а масштабирование приложения может достигнуть своих пределов при росте объема данных или пользовательской нагрузки [10].

На рис. 3 представлен пример классической монолитной архитектуры, где компоненты и слои приложения программного обеспечения объединены в единицу развертывания, а взаимодействие между компонентами осуществляется с помощью локальных вызовов.

Рассмотрим альтернативу монолитной архитектуре — микросервисную архитектуру — еще один подход к разработке программного обеспечения, при котором приложение строится как набор небольших, независимых и легко масштабируемых сервисов, взаимодействующих между собой через API (Application Program Interface). Каждый микросервис представляет собой отдельную и независимую функциональность или бизнес-задачу [10].

На рис. 4 изображен пример микросервисной архитектуры. На схеме видно отсутствие общей единицы развертывания для ПО, как в случае мо-



нолитной архитектуры, и разделение ПО на микросервисы, выполняющие свои независимые функции и передающие вызовы другим микросервисам через API.

Анализ существующих подходов демонстрирует отсутствие универсального решения. Выбор архитектуры системы хранения для цифровой экосистемы банка должен определяться спецификой бизнес-требований к каждому типу данных, что предполагает переход от монолитной базы данных к полиглотному хранению (Polyglot Persistence)<sup>5</sup> — стратегии, где в рамках одной экосистемы используются различные системы хранения, наиболее подходящие для решения конкретных задач. Например, транзакционные данные могут храниться в СР-системе (NewSQL, такая как YDB), кэш-сессии — в кластере ключ-значение (Redis или Tarantool), а журналы операций — в ширококолонном хранилище (Apache Cassandra). Ключевая задача здесь — обеспечение надежной оркестрации (на основе таких платформ, как Kubernetes), мониторинга и согласованности данных между компонентами [11].

Рис. 3 / Fig. 3. Монолитная архитектура ПО / Monolithic Software Architecture

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

<sup>5</sup> Polyglot Persistence. URL: <https://martinfowler.com/bliki/PolyglotPersistence.html>

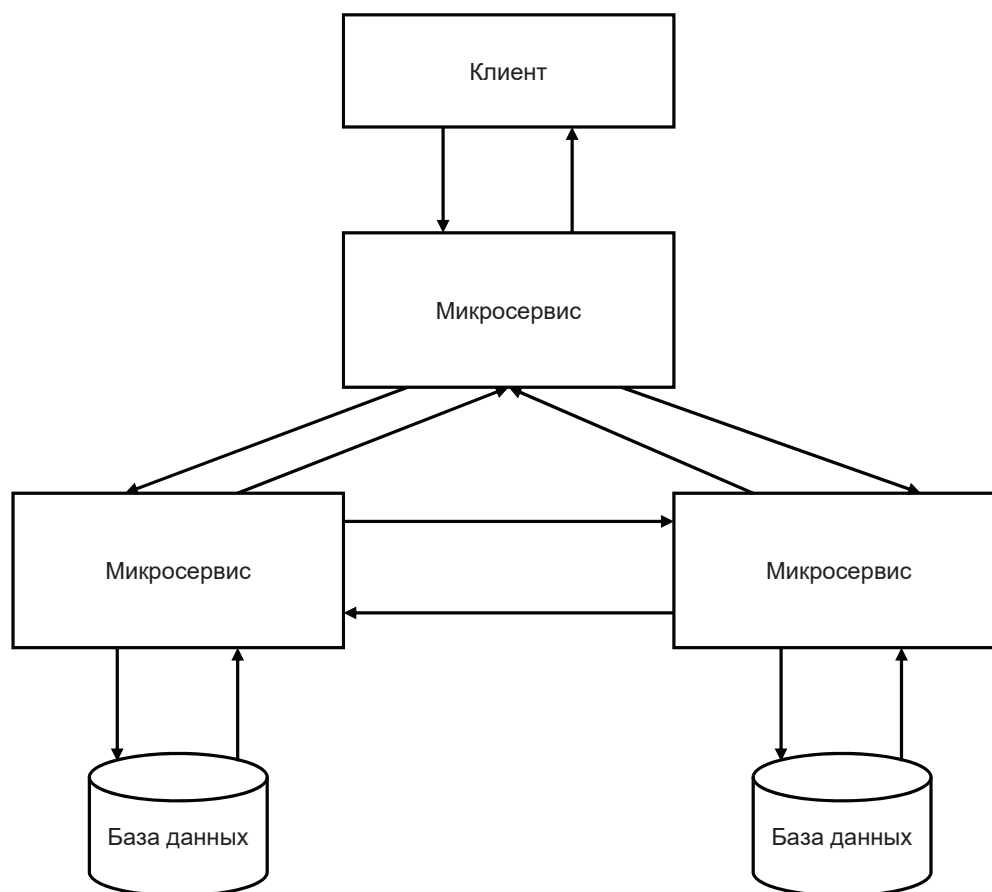


Рис. 4 / Fig. 4. Микросервисная архитектура ПО / Software Microservice Architecture

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

## СРАВНИТЕЛЬНЫЙ АНАЛИЗ РЕШЕНИЙ И ПОДХОДОВ К РАЗРАБОТКЕ СИСТЕМ ХРАНЕНИЯ И ОБРАБОТКИ ДАННЫХ

В ходе исследования проведен сравнительный анализ подходов и технологий хранения данных, таких как традиционные реляционные СУДБ,

NoSQL- и NewSQL-системы. Выбраны ключевые критерии сравнения при выборе решения для разработки распределенной и высоконагруженной систем хранения данных:

- теоретическая основа и модель данных;
- архитектурный паттерн масштабирования;

Таблица 1 / Table 1

Сравнительный анализ подходов и решений к хранению данных / Comparative Analysis of Approaches and Solutions to Data Storage

Критерий / Criteria	Реляционные СУДБ / Relational databases	NoSQL-системы / NoSQL systems	NewSQL-системы / NewSQL systems
Теоретическая основа и модель данных	Реляционная алгебра, теория множеств. Жесткая, предопределенная схема. ACID – неотъемлемое свойство	Отказ от реляционной модели в пользу специализированных структур: документной, «ключ-значение», колоночной, графовой. Гибкая или отсутствующая схема. Базируются на принципе BASE	Сохранение реляционной модели и SQL-интерфейса как фундамента. Переосмысление архитектурного уровня для распределенного выполнения с сохранением ACID
Архитектурный паттерн масштабирования	Преимущественно используется вертикальное масштабирование. Кластерные конфигурации ориентированы на отказоустойчивость, а не на линейный прирост производительности при записи	Горизонтальное масштабирование как основополагающий принцип. Шардирование и партиционирование данных и репликация реализованы на архитектурном уровне	Горизонтальное масштабирование как базовая возможность. Автоматическое шардирование и ребалансировка данных в рамках единого логического кластера
Позиционирование относительно теоремы CAP	Приоритет CA (Consistency, Availability). В классических одноузловых и кластерных конфигурациях с синхронной репликацией устойчивость к разделению сети не достигается	Явный выбор AP (Availability, Partition Tolerance) или CP (Consistency, Partition Tolerance) в зависимости от предназначения системы	Стратегическая цель – CA-системы с минимальным компромиссом по P благодаря инновационным распределенным протоколам и управляемой задержки репликации
Модель транзакций и консистентность	Полноценная поддержка ACID-транзакций, включающих множественные операции над разнородными данными	Транзакционность часто ограничена операциями в рамках одного документа, ключа или партиции. Глобальные распределенные транзакции либо отсутствуют	Ключевое архитектурное достижение: поддержка распределенных ACID-транзакций на уровне всего кластера

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



- позиционирование относительно теоремы CAP;
- модель транзакций и консистентность.

Результаты сравнительного анализа представлены в *табл. 1*.

Выбор варианта хранения данных — не просто оптимизационная задача, а стратегическое решение, определяющее онтологические свойства будущей системы. При создании хранилища для банковской цифровой экосистемы приоритет отдается не единой парадигме, а стратифицированной

архитектуре, где каждый подход используется там, где его свойства и ограничения дают наилучший результат [12]. Для высоконагруженных распределенных систем в исследовании сопоставлены монолитная и микросервисная архитектуры. Итоги сравнения представлены в *табл. 2*.

Результаты сравнительного анализа показывают: и микросервисная, и монолитная архитектуры ПО обладают как преимуществами, так и недостатками — в зависимости от критериев оценки. При этом выбор архитектуры определяется функцио-

Таблица 2 / Table 2

### Сравнительный анализ монолитной и микросервисной архитектур ПО / Comparative Analysis of Monolithic and Microservice Software Architectures

Критерий / Criterion	Архитектура / Architecture	
	монолитная / Monolithic	микросервисная / Microservice
Структурная организация	Единая, тесно связанная кодовая база. Все функциональные модули компилируются и развертываются как единое целое	Набор слабосвязанных, независимо развертываемых сервисов. Каждый сервис инкапсулирует специфическую бизнес-возможность и обладает автономным циклом данных
Взаимосвязанность компонентов	Жесткая взаимосвязанность между модулями через общие структуры данных, вызовы функций и единую базу данных, что приводит к распространению изменений	Взаимодействие через четко определенные, стабильные API (REST/gRPC, асинхронные взаимодействия). Распространение изменений ограничено контрактом API
Масштабируемость	Вертикальное масштабирование: развертывание дополнительных идентичных экземпляров всего монолита. Горизонтальное масштабирование неэффективно, так как копируется вся функциональность, даже если нагрузка пришлась на один модуль	Гранулярное горизонтальное масштабирование: возможность независимо масштабировать только те сервисы, которые испытывают высокую нагрузку. Позволяет оптимизировать использование ресурсов и стоимостные показатели
Гибкость технологического стека	Единый технологический стек для всего приложения. Технологическая эволюция затруднена, требует глобального обновления	Принцип полиглотного стека: каждый сервис реализуется на наиболее подходящем языке и используется специализированное хранилище данных. Ускоряет инновации и адаптацию новых технологий
Отказоустойчивость	Единая точка отказа: сбой в любом модуле приводит к падению всего приложения	Распределение и изоляция сбоев повышает общую доступность системы

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

нальными и нефункциональными требованиями к системе хранения данных. В частности, для цифровой банковской экосистемы при разработке системы хранения клиентских данных эффективнее использовать микросервисную архитектуру: она обеспечивает необходимое масштабирование и отказоустойчивость.

## ВЫВОДЫ

Проведенное исследование позволяет сформулировать комплексный подход к проектированию и разработке распределенной высоконагруженной системы хранения клиентских данных в условиях современной цифровой банковской экосистемы. Анализ эволюции требований к подобным системам — экспоненциальный рост объемов и разнообразия данных, необходимость обработки информации в режиме реального времени, ужесточение регуляторных норм и повышенные ожидания клиентов к доступности и персонализации сервисов — выявил несостоятельность традиционных подходов к построению систем хранения данных.

Сравнительный анализ технологических парадигм хранения данных — реляционных СУБД, NoSQL- и NewSQL-систем — подтвердил отсутствие универсального решения. Каждый подход — уникальный компромисс в рамках теоремы CAP (Брюера), что детерминирует его применимость для конкретных типов данных и сценариев использования внутри экосистемы. Транзакционные данные, требующие строгой ACID-семантики, могут эффективно обслуживаться современными

NewSQL-системами (например, CockroachDB, YDB), в то время как для задач кэширования и работы с временными рядами оптимальны хранилища «ключ-значение» (Redis, Apache Ignite) и широко-колонные СУБД (Apache Cassandra).

Архитектурный анализ продемонстрировал, что микросервисная парадигма, в сравнении с монолитной, является более подходящим фундаментом для построения масштабируемой и отказоустойчивой системы. Преимущества данного подхода — слабая связанность компонентов, гранулярное горизонтальное масштабирование, технологическая гибкость и изоляция сбоев — прямо отвечают на вызовы динамичной банковской среды. Ключевым архитектурным принципом становится стратегия Polyglot Persistence, предполагающая синергетическое использование разнородных систем хранения, наиболее подходящих для решения конкретных бизнес-задач.

Таким образом, оптимальная архитектура единого хранилища клиентских данных для цифровой банковской экосистемы представляет собой стратифицированную, сервисно-ориентированную систему. Ее ядро должно сочетать гарантии целостности критически важных данных (через NewSQL или реляционные кластеры) с высокой доступностью и производительностью для аналитических и операционных нагрузок (через специализированные NoSQL-решения и in-memory-хранилища). Реализация таких архитектурных паттернов, как CQRS и Event Sourcing, позволит дополнительно повысить масштабируемость и поддержку разнообразных материализованных представлений данных.

## СПИСОК ИСТОЧНИКОВ

1. Клеппман М. Высоконагруженные приложения программирование, масштабирование, поддержка. Пер. с англ. И. Пальти, А. Тумаркин. СПб.: Питер; 2021. URL: [https://rusneb.ru/catalog/000200\\_000018\\_RU\\_NLR\\_BIBL\\_A\\_012416514/](https://rusneb.ru/catalog/000200_000018_RU_NLR_BIBL_A_012416514/)
2. Мартин Р. Чистая архитектура. Искусство разработки программного обеспечения. СПб.: Питер; 2021. URL: <https://search.rsl.ru/ru/record/01010779241>
3. Wiling B. Scientific Study of CAP Theorem and Understanding its Different Implementation Methods. *Mathematical Statistician and Engineering Applications*. 2022;(1);133-137. DOI: 10.17762/msea.v7i1i.55
4. Таненбаум Э., ван Стейен М. Распределенные системы. Принципы и парадигмы. СПб.: Питер; 2003. URL: [https://rusneb.ru/catalog/000199\\_000009\\_002157753/?ysclid=mmw6lsm0tr819862687](https://rusneb.ru/catalog/000199_000009_002157753/?ysclid=mmw6lsm0tr819862687)
5. Садаладж П., Фаулер М. NoSQL: новая методология разработки нереляционных баз данных. М. Вильямс; 2013. 192 с. URL: <https://search.rsl.ru/ru/record/01006568860>
6. Lakshman, A., Malik, P. Cassandra: A Decentralized Structured Storage System. *ACM SIGOPS Operating Systems Review*. 2021;(2):35-40. DOI: 10.1145/1773912.1773922
7. Corbett J. C. Spanner: Google's Globally-Distributed Database. *OSDI*. 2023;(12):251-264. URL: <https://sayedalesawy.hashnode.dev/spanner-googles-globally-distributed-database>
8. Игнатенко И. Д., Астахов В. В., Акинина Ю. С. Сравнительный анализ подходов к реализации распределенных транзакций. В сб.: Будущее науки — 2025. 2025;(1):209-216. URL: <https://www.elibrary.ru/eoxwmm>
9. Lewis P., Perez E., Piktus A. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*. 2021;(33):9459-9474. DOI: 10.48550/arXiv.2005.11401



10. Ньюмен С. От монолита к микросервисам. СПб.: БХВ-Петербург; 2021. 272 с. URL: [https://rusneb.ru/catalog/000200\\_000018\\_RU\\_NLR\\_BIBL\\_A\\_012548166/?ysclid=mmw6qujxqc209793652](https://rusneb.ru/catalog/000200_000018_RU_NLR_BIBL_A_012548166/?ysclid=mmw6qujxqc209793652)
11. Куликова О. М., Суворова С. Д. Облачные технологии: основа построения корпоративной архитектуры. *Инновационная экономика: перспективы развития и совершенствования*. 2021;(4):65-70. URL: <https://www.elibrary.ru/xfjvta>
12. Косарев В. Е., Городецкая О. Ю., Гобарева Я. Л., Рычаго М. Е. Интеграция распределенных облачных вычислений и модульных информационных систем для повышения эффективности управления промышленным производством. *Кузнечно-штамповочное производство. Обработка материалов давлением*. 2025;(11):107-114. URL: <https://www.elibrary.ru/qgcazs>

## REFERENCES

1. Kleppman M. Highly loaded applications programming, scaling, support. Transl. from Eng. by I. Palti, A. Tumarkin. Saint Petersburg: St. Petersburg; 2021. URL: [https://rusneb.ru/catalog/000200\\_000018\\_RU\\_NLR\\_BIBL\\_A\\_012416514/](https://rusneb.ru/catalog/000200_000018_RU_NLR_BIBL_A_012416514/) (In Russ.).
2. Martin R. Pure architecture. The art of software development. Saint Petersburg: St. Petersburg; 2021. URL: <https://search.rsl.ru/ru/record/01010779241> (In Russ.).
3. Wiling B. Scientific Study of CAP Theorem and Understanding its Different Implementation Methods. *Mathematical Statistician and Engineering Applications*. 2022;(1);133-137. DOI: 10.17762/msea.v71i1.55
4. Tanenbaum E., van Steyen M. Distributed systems. Principles and paradigms. Saint Petersburg: Peter; 2003. URL: [https://rusneb.ru/catalog/000199\\_000009\\_002157753/?ysclid=mmw6lsm0tr819862687](https://rusneb.ru/catalog/000199_000009_002157753/?ysclid=mmw6lsm0tr819862687) (In Russ.).
5. Sadalaj P., Fowler M. NoSQL: a new methodology for developing non-relational databases. Moscow: Williams; 2013. 192 p. URL: <https://search.rsl.ru/ru/record/01006568860> (In Russ.).
6. Lakshman A., Malik, P. Cassandra: A Decentralized Structured Storage System. *ACM SIGOPS Operating Systems Review*. 2021;(2):35-40. DOI: 10.1145/1773912.1773922
7. Corbett J. C. Spanner: Google's Globally-Distributed Database. *OSDI*. 2023;(12):251-264. URL: <https://sayedalesawy.hashnode.dev/spanner-googles-globally-distributed-database> (In Russ.).
8. Ignatenko I. D., Astakhov V. V., Akinina Yu. S. Comparative analysis of approaches to the implementation of distributed transactions. In the collection: The future of science — 2025. 2025;(1):209-216. URL: <https://www.elibrary.ru/eoxwmm> (In Russ.).
9. Lewis P., Perez E., Piktus A. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*. 2021;(33):9459-9474. URL: <https://doi.org/10.48550/arXiv.2005.11401>
10. Newman S. From monolith to microservices. Saint Petersburg: BHV-Petersburg; 2021. 272 p. URL: [https://rusneb.ru/catalog/000200\\_000018\\_RU\\_NLR\\_BIBL\\_A\\_012548166/?ysclid=mmw6qujxqc209793652](https://rusneb.ru/catalog/000200_000018_RU_NLR_BIBL_A_012548166/?ysclid=mmw6qujxqc209793652) (In Russ.).
11. Kulikova O. M., Suvorova S. D. Cloud technologies: the basis for building a corporate architecture. *Innovative economy: prospects for development and improvement*. 2021;(4):65-70. URL: <https://www.elibrary.ru/xfjvta> (In Russ.).
12. Kosarev V. E., Gorodetskaya O. Yu., Gobareva Ya. L., Rychago M. E. Integration of distributed cloud computing and modular information systems to improve the efficiency of industrial production management. *Forging and stamping production. Pressure treatment of materials*. 2025;(11):107-114. URL: <https://www.elibrary.ru/qgcazs> (In Russ.).

## ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

**Никита Андреевич Бурыкин** — студент магистратуры факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**Nikita A. Burykin** — Master's degree student, Department of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation  
<https://orcid.org/0009-0009-6631-230X>

Автор для корреспонденции / Corresponding author:  
244461@edu.fa.ru



**Наталья Владимировна Гринева** — кандидат экономических наук, доцент, доцент кафедры информационных технологий факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

**Natalia V. Grineva** — Cand. Sci. (Econ.), Assoc. Prof., Assoc. Prof. of the Department of Information Technology, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0000-0001-7647-5967>

[ngrineva@fa.ru](mailto:ngrineva@fa.ru)

**Павел Евгеньевич Голосов** — кандидат технических наук, директор Института общественных наук РАНХиГС, Москва, Российская Федерация

**Pavel E. Golosov** — Cand. Sci. (Tech.), Director of the Institute of Social Sciences RANEPА, Moscow, Russian Federation

<https://orcid.org/0000-0003-4313-0887>

[golosov-pe@ranepa.ru](mailto:golosov-pe@ranepa.ru)

#### **Заявленный вклад авторов:**

**Н. А. Бурыкин** — сравнительный анализ архитектурных подходов и технологических решений для разработки системы хранения клиентских данных, написание разделов «Введение», «Выводы».

**Н. В. Гринева** — разработка общей концепции статьи, исследование существующих подходов хранения и обработки данных.

**П. Е. Голосов** — исследование существующих технологических решений хранения данных, список литературы.

#### **Authors' contributions:**

**N. A. Burykin** — comparative analysis of architectural approaches and technological solutions for the development of a customer data storage system, writing the sections “Introduction”, “Conclusions”.

**N. V. Grineva** — development of the general concept of the article, research of existing approaches to data storage and processing.

**P. E. Golosov** — research of existing technological solutions for data storage, references.

*Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.*

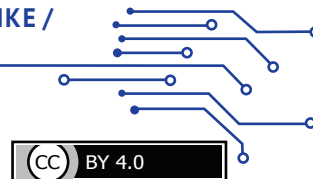
*Conflicts of Interest Statement: The authors have no conflicts of interest to declare.*

*Статья поступила 19.01.2026; после рецензирования 02.03.2026; принята к публикации 13.03.2026.*

*Авторы прочитали и одобрили окончательный вариант рукописи.*

*The article was submitted on 19.01.2026; revised on 02.03.2026 and accepted for publication on 13.03.2026.*

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



# Исследование реальных фракталов опережающих индикаторов

И.Ю. Варьяш

Научно-исследовательский финансовый институт (НИФИ), Москва, Российская Федерация

## АННОТАЦИЯ

В статье исследуется применимость фрактального анализа в социологии и биржевой аналитике. Автор рассматривает возможность использования принципа фрактальности для изучения социоэкономических показателей и прогнозирования биржевой активности. **Цель исследования** – продемонстрировать потенциал фрактальной методологии как инструмента междисциплинарного анализа, объединяющего социологические наблюдения, экономические индикаторы и технические методы биржевого анализа. **Основные задачи:** проанализировать теоретические основы фрактального подхода; выявить особенности проявления фрактальных структур в социальных и экономических процессах; оценить эффективность фрактального анализа в сравнении с традиционными методами биржевого прогнозирования. **Методология** включает: обзор научной литературы по фрактальному анализу в финансах и социологии; применение концепции масштабируемой инвариантности и дробной размерности; корреляционный анализ динамических рядов; моделирование циклической фрактальности на разных горизонтах ожиданий (микро-, медиа- и макроциклы). **Результаты:** подтверждена циклическая природа социальных фракталов и их способность формировать последовательности на кривой медианных значений показателей; выявлена масштабируемость фрактальных структур, позволяющая расширять горизонт прогнозирования опережающих индикаторов; рассчитан коэффициент корреляции динамических рядов, демонстрирующий устойчивость подобия независимо от горизонта ожиданий; показано, что микроциклы демонстрируют более высокое подобие к макроциклам, тогда как сходство с медиациклами выражено слабее. **Выводы:** фрактальный анализ обладает преимуществами перед традиционными методами биржевого анализа (технический, фундаментальный) благодаря численной оценке вероятности событий, вектору волатильности, обозначению периодов перелома трендов, масштабируемости инвариантов; метод перспективен для макроэкономического прогнозирования, стратегического планирования и управления рисками, особенно в сочетании с другими индикаторами. Однако фрактальный анализ не является универсальным: его эффективность зависит от контекста и требует учета ограничений (например, относительности социального времени, мультифрактальности данных). **Практическая значимость** заключается в возможности использования фрактальных моделей для: уточнения биржевых прогнозов; синтеза опережающих данных; заполнения информационных промежутков между регулярными отчетами (еженедельными, ежемесячными).

**Ключевые слова:** фрактальный анализ; социальные фракталы; циклическая фрактальность; масштабирование; биржевой анализ; опережающие индикаторы; корреляционный анализ; прогнозирование

**Для цитирования:** Варьяш И.Ю. Исследование реальных фракталов опережающих индикаторов. *Цифровые решения и технологии искусственного интеллекта*. 2026;2(1):73-82. DOI: 10.26794/3030-7097-2026-2-1-73-82

## Research of Real Fractals of Leading Indicators

I.Yu. Varjas

Scientific Research Financial Institute (NIFI), Moscow, Russian Federation

## ABSTRACT

The article explores the applicability of fractal analysis in sociology and stock market analytics. The author considers the possibility of using the fractal principle to study socio-economic indicators and forecast stock market activity. The purpose of the study is to demonstrate the potential of fractal methodology as an interdisciplinary analysis tool combining sociological observations, economic indicators and technical methods of stock market analysis. Main tasks: to analyze the theoretical foundations of the fractal approach; to identify the features of fractal structures in social and economic processes; to evaluate the effectiveness of fractal analysis in comparison with traditional methods of stock forecasting. The methodology includes: a review of the scientific literature on fractal analysis in finance and sociology; application of the concept of scalable invariance and fractional dimension; correlation analysis of dynamic

series; modeling cyclic fractality on different horizons of expectations (micro, media and macrocycles). Results: The cyclical nature of social fractals and their ability to form sequences on the curve of median values of indicators have been confirmed. The scalability of fractal structures has been revealed, which makes it possible to expand the horizon of forecasting leading indicators. The correlation coefficient of dynamic series is calculated, demonstrating the stability of similarity regardless of the expectation horizon. It is shown that microcycles show a higher similarity to macrocycles, while the similarity to media cycles is less pronounced. Conclusions. Fractal analysis has advantages over traditional stock analysis methods (technical, fundamental) due to the numerical estimation of the probability of events, the volatility vector, the designation of trend reversal periods, and the scalability of invariants. The method is promising for macroeconomic forecasting, strategic planning and risk management, especially in combination with other indicators. However, fractal analysis is not universal: its effectiveness depends on the context and requires consideration of limitations (for example, the relativity of social time, the multifractality of data). The practical significance lies in the possibility of using fractal models to: refine stock forecasts; synthesize leading data; fill in the information gaps between regular reports (weekly, monthly).

**Keywords:** fractal analysis; social fractals; cyclical fractality; scaling; stock market analysis; leading indicators; correlation analysis; forecasting

**For citation:** Varjas I.Yu. Aggregation of weakly connected components and influence bridges in multilayer social graphs. *Digital solutions and artificial intelligence technologies*. 2026;2(1):73-82. DOI: 10.26794/3030-7097-2026-2-1-73-82

## ВВЕДЕНИЕ

Объектом социологического исследования фрактального измерения являются события, понятия, последовательности численных значений, выстраивающиеся в структуры подобных геометрических фигур [1–3]. Происходит масштабирование исходного компонента, определяемого как масштабируемая инвариантность [1, 4]. Однако этот привлекательный, на первый взгляд, успешно работающий в биржевом техническом анализе метод должен был бы противоречить условию неравенства любых общностей, личностей, тем более событий в социологии. Принцип подобия в социуме более, чем относителен. Единственное, что остается здесь подобным, — это дробные части четвертого измерения социума — периоды времени. При всей условности квантирования пространственно-временного континуума социума время остается, хотя и условно, равномерно распределенным по нему во множестве размерностей.

## ОБЗОР ЛИТЕРАТУРЫ

Принцип фрактальности применяется в техническом биржевом анализе во всем мире. В российском биржевом сообществе можно найти первые попытки применения фрактальной теории на организованном рынке. В основном используется наследие Б. Мандельброта и его математический аппарат [1–3, 5, 6]. Но до сих пор он не развит в анализе экономических и финансовых процессов, хотя некоторые предпосылки для этого постепенно складываются [7, 8]<sup>1</sup>.

<sup>1</sup> В начале 2000-х гг. увлечение «эконофизикой» направило внимание экономистов, социологов, философов и математиков на исследование фракталов в общественных процессах. Впоследствии эта тема не развивалась из-за слишком очевидной метафоричности переноса методов естественных наук в область обществознания. В связи с развитием фрактального метода в биологии и социологии было нача-

Теория фрактального анализа рынков изложена Б. Мандельбротом в соавторстве с Р.Л. Хадсоном в работе «(Не)послушные рынки: фрактальная революция в финансах» [1]. Б. Вильямс предположил нечто общее в движении хаотических систем, к которым причислил рынок, а именно — присущее всем рынкам и любым таймфреймам движение с одинаковой структурой цен со структурой, схожей с движением объектов в природе и биологии [9]. Выявленные Б. Вильямсом фракталы изменения численных показателей инструментов биржевой активности на рынке Форекс позволили перейти к ее компьютерной симуляции. В то же время, как подтвердило дальнейшее исследование, выявилась недостаточность линейных функций для анализа. Некорректным представляется и вывод о том, что хаос, вероятно, является постоянным состоянием рынка. Фрактал назван присущим стоп-лоссам, что противоречит предположению хаоса рынка.

Одни из последователей применения в России фрактальной теории Мандельброта–Хадсона для разработки стратегии на финансовых рынках А.А. Алмазов обратил внимание на фундаментальность циклов изменения цен [10]. Им была предложена фрактальная функция в составе математической программы Fractan. Автор применил математическое понятие начальных условий, аттракторов, непериодических циклов, размерностей и др. к графическому представлению анализа структуры биржевых рынков. И, тем не менее, по его мнению, указанный математический аппарат все же оказался недостаточным для прогнозирования ценообразования.

то преодоление этого отставания в ходе технического анализа, а ныне в синтезе социологических и экономических динамических рядов [3, 9].



А.Е. Сериков [5, с. 176] предлагает модель вычисления «нормированного размаха» методом накопленного отклонения (по Чижевскому) для вычисления квантирования временного ряда. Но он опирается лишь на архивные статистические данные, вовсе не предназначенные для вычисления будущих численных значений экономических показателей. Отсюда и неуместность автоматической линейной аппроксимации, допускающей произвольное элиминирование отдельных точек на кривой. В будущем «явно выпадающих из общей тенденции» точек нет и быть не может. В противном случае вероятность аппроксимации заведомо подвергается обоснованному сомнению. То же самое относится и к показателю Херста — меры «изломанности» процесса. Стремление к сведению модели к линейной зависимости представляется неуместным по отношению к социальным событиям, не подчиняющимся нормированию. Так, при  $D = 2 - H$ , где  $D > 1,5$  и  $H < 0,5$  пропадает автокорреляция, т.е. значения ряда оказываются независимыми, что некорректно. Исследования А.А. Давыдова показывают, что ряды социологических данных характеризуются мультифрактальностью и их размерность по Херсту — изменчивостью [11].

### 1. Исследование фрактальности социальноэкономических показателей

Использование принципа подобия в циклических процессах позволяет применять характерное для фракталов свойство масштабирования. Это дает

возможность расширять горизонт прогнозирования опережающих показателей [12].

В рассматриваемых случаях самоподобие проявляется как преобразование, при котором структура сохраняется при изменении масштаба. Такое свойство позволяет на практике либо увеличивать детализацию анализа (масштабировать), либо упрощать его (квантовать) — в зависимости от поставленных задач.

В экономике чаще используют масштабирование, например, при долгосрочном прогнозировании и планировании. Однако дробление также применяется, например, при разбивке прогнозных периодов на календарные планы, дорожные карты и другие инструменты управления.

Рассмотрим проявление цикличности в социальной фрактальности на примере рынка FOREX (рис. 1).

Рисунок 1, демонстрирующий геометрию фрактальности опережающего индикатора биржевой активности рынка USD/RUB<sup>2</sup>, позволяет предположить возможность дополнения технического и фундаментального анализа биржевого рынка социологией фракталов как самоподобных циклических процессов, проявляющихся, в том числе, в изменении мнений активных участников рынка.

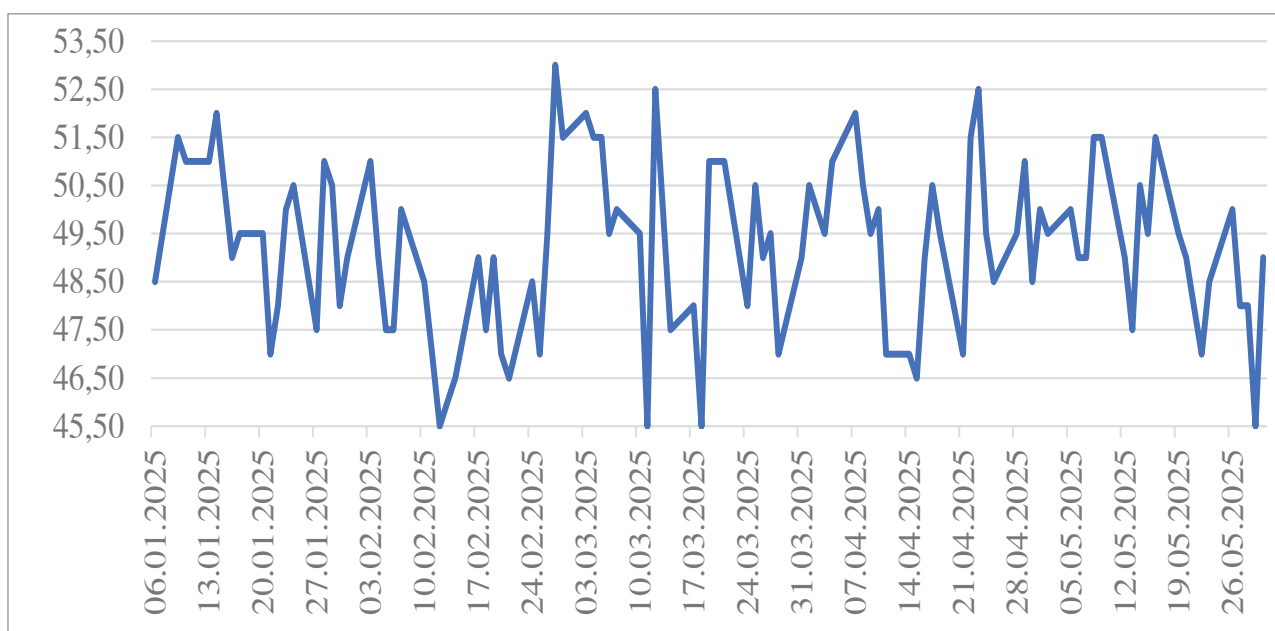


Рис. 1 / Fig. 1. Фрактальная динамика опережающего индикатора курса USD/RUB. Январь-май 2025 г. / Fractal Dynamics of the USD/RUB Leading Indicator. January-May 2025

Источник / Source: составлено автором / Compiled by the author.

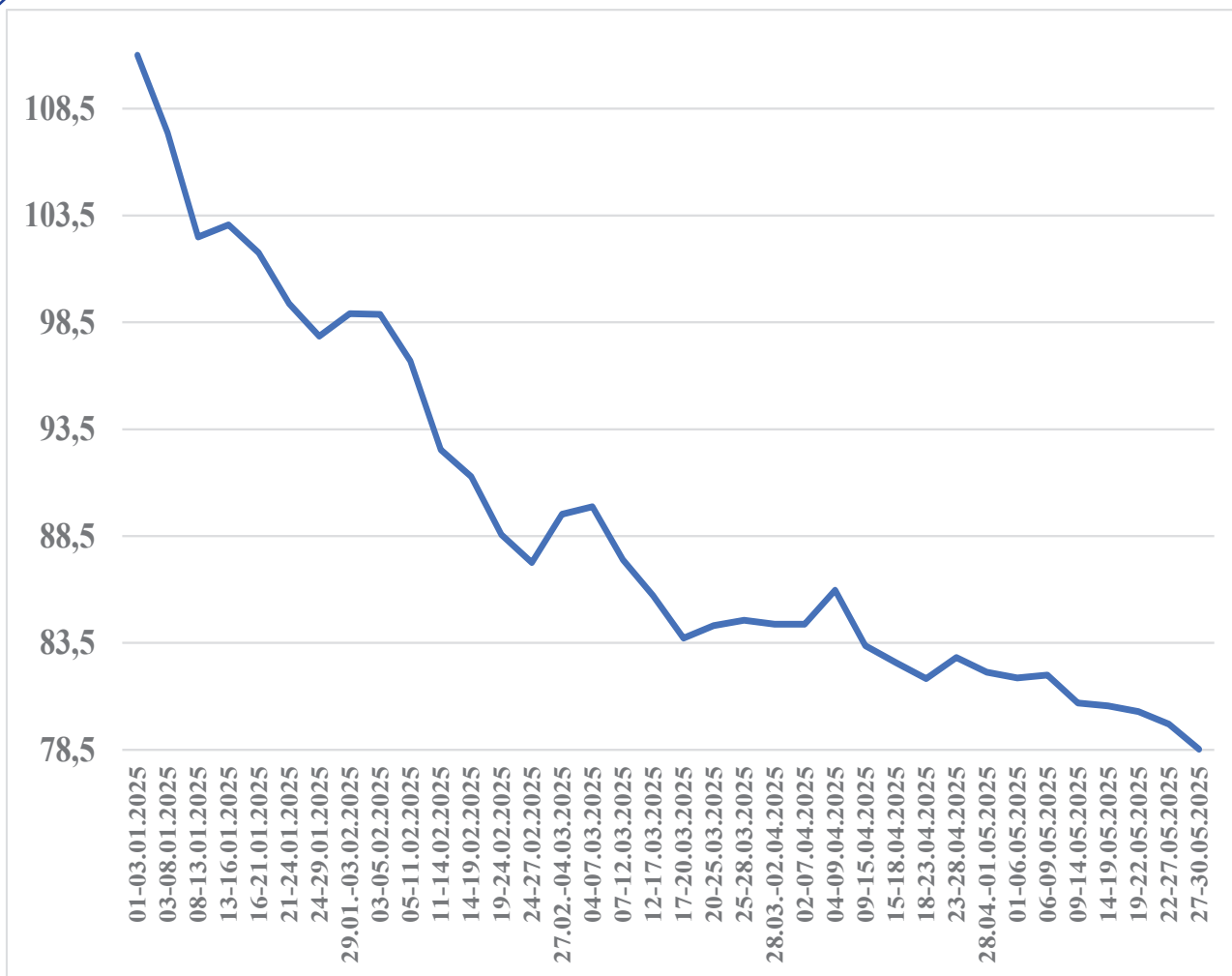


Рис. 2 / Fig. 2. Индекс изменения стоимости доллара США в российских рублях /  
 The Index of Changes in the Value of the USD in RUB

Источник / Source: составлено автором по данным Московской биржи / Compiled by the author according to the data of the MOEX\*.

Примечание / Note: \*Московская биржа. URL: <https://www.exchange-rates.org/ru>

## 2. Циклическая фрактальность

Структурная социология является частным случаем социологии процессов, поскольку оперирует имплицитным допущением неизменности социальных состояний, что обуславливается проектным требованием одновременности социологических опросов. Тем не менее структура «социологии знания» в информатике может быть развернута в систему процессов. Наиболее ясно это видно при сопоставлении ответов на одни и те же вопросы разных групп (поколенческих, образовательных, профессионально-квалификационных, статусных), т.е. всех тех, чей состав сменяется новыми возрастными группами (прежде всего, поколениями) в биржевой торговле. Остается отделить фундаментальные и конъюнктурные компоненты личностного и группового сознания в качестве неизменных и изменяющихся переменных.

Необходимо учитывать *разноразмерность* объектов социологического наблюдения. При этом время в социуме можно считать относительно *равноразмерным*: его допустимо квантовать на самоподобные отрезки (фракталы). Даже с учетом относительности социального времени эта относительность сохраняет *равноразмерность*. Речь идет о фрактальности, определяемой параметрами временных циклов (срочности) и пространственным распределением данных в многомерной модели (радиус-векторов барицентров в *n*-мерном пространстве) [13].

Существенное отличие фрактальных структур от фрактальной динамики заключается в последовательности фигур, сходных по трем критериям:

- срочности;
- количеству событий (числовых значений показателя);



• пространственному соотношению элементов.

Аналогичная закономерность наблюдается в динамических рядах соответствующих показателей, сформированных на основе цикличности (рис. 2).

Как видим, структура фракталов цепного индекса статистики подобна структуре фракталов опережающих индикаторов по всем трем признакам, а также и по четвертому признаку — одинаковой размерности показателей. Для изменения курса используется цепной индекс, а для опережающего индикатора — индекс ЕМІ (индекс ожиданий рынка), отражающий отношение изменения мнений участников рынка о будущих численных значениях биржевых показателей относительно прошлого периода — в данном случае будущего микроцикла к предшествующему микроциклу<sup>3</sup>.

Впрочем, отвлекаться от ускорения социально-экономического развития не стоит. В нем состоит относительность социального времени и в восприятии общества, и в восприятии личности, когда время будто бы сжимается, и события или несутся вскачь, или растягиваются, когда кажется, что время остановилось. Конечно, изменяется не размерность времени, а ее восприятие как наиболее нестабильная компонента относительности социального времени. Тем не менее в экономическом управлении и в политике относительность четвертого измерения чувствуется в наименьшей степени благодаря несравнимой с ними медленности фундаментальных социальных процессов.

### 3. Масштабирование циклической фрактальности

Следуя направлению исследования масштабирования фракталов в социологии и политологии [5, 6], сказанное ранее проиллюстрируем результатами анализа подобия динамики опережающих индикаторов и статистических данных применения циклического фрактального анализа для построения модели опережающих данных биржевого рынка на различных горизонтах ожиданий (см. таблицу).

Исследование масштабирования циклической фрактальности дало нетривиальные результаты.

Количественная оценка подобия, измеряемая коэффициентом корреляции динамических рядов, составляет 0,72. Это значение остается стабильным в пределах статистической ошибки независимо от горизонта ожиданий.

При этом наблюдается следующая закономерность:

- микроциклы демонстрируют более высокое подобие к макроциклам (0,72 против 0,74);
- сходство с медиациклами (состоящими из 7 и 9 числовых значений показателей) выражено слабее.

Анализ семидневных циклов показал, что их сходимость не достигает статистически значимого уровня корреляции (0,49). В девятидневных циклах коэффициент корреляции составляет 0,69, что лишь незначительно (на 0,01) ниже условного порога значимости (0,70). В контексте экономически значимого горизонта ожиданий это отклонение можно считать несущественным.

Далее необходимо построить графическую схему масштабирования фракталов и представить схему масштабирования циклических фракталов опережающих индикаторов (рис. 3).

Как видим, у рядов фракталов появляется еще одно измерение — глубина, что указывается на  $n + 1$  измерений.

Как и любой другой математический метод, фрактальный анализ в одних ситуациях эффективен, а в других — не очень [14]. В этом смысле он не лучше и не хуже, к примеру, метода японских свечей<sup>4</sup>: он просто немного иной и срabатывает в иных случаях.

Иногда результативен метод японских свечей, иногда — построение и анализ фракталов, а иногда первым сигнал подает какой-либо иной подход. Поэтому строить торговлю, руководствуясь только результатами фрактального анализа, нельзя — так же, как нельзя отказаться от фундаментального анализа и торговать, опираясь лишь на технический.

Обычно с помощью данного метода трейдер уточняет свой прогноз. При этом нельзя недооценивать фрактальный анализ: в сочетании с другими индикаторами это довольно мощный и эффективный инструмент. Этому способствует также множественность измерений фрактального представления, предоставляющая дополнительные возможности для синтеза опережающих данных [15, 16].

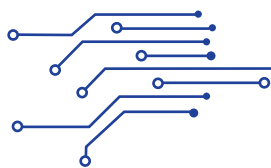
## Выводы

На основе проведенного исследования можно сформулировать следующие ключевые выводы:

**1. Фрактальность социоэкономических процессов подтверждена.** Социальные фракталы де-

<sup>3</sup> Подробнее о цикличности изменения экономических показателей см. [12, 13].

<sup>4</sup> URL: <https://www.rbc.ru/quote/news/article/6931636c9a7947b3da867402>



**Корреляционный анализ подобия опережающих индикаторов и статистики на разных горизонтах ожиданий / Correlation Analysis of the Similarity of leading Indicators and Statistics on Different Horizons of Expectations**

USD/RUB		IMOEX		GOLD		BRENT		Median*
Макроциклы (27 б/дн**)								
периоды	корр.	периоды	корр.	периоды	корр.	периоды	корр.	0,72
22.05–27.06.2025	0,92	28.04–03.06.2025	0,46	06.05–11.06.2025	0,78	24.04–30.05.2025	0,36	
27.06–04.08.2025	0,88	03.06–09.07.2025	0,68	11.06–17.07.2025	0,79	30.05–07.07.2025	0,57	
04.08–09.09.2025	0,97	09.07–14.08.2025	0,94	17.07–22.08.2025	0,78	07.07–12.08.2025	0,49	
09.09–15.10.2025	0,83	14.08–19.09.2025	0,78	22.08–29.09.2025	0,75	12.08–17.09.2025	0,52	
22.05.–15.10.2024	0,90	28.04.–19.09.2025	0,73	06.05.–29.09.2025	0,78	24.04.–17.09.2024	0,51	
Медиациклы (9 б/дн**)								
периоды	корр.	периоды	корр.	периоды	корр.	периоды	корр.	0,69
28.08–09.09.2025	0,72	05.09–15.09.2025	0,58	01.09–11.09.2025	0,80	22.08–03.09.2025	0,65	
09.09–19.09.2025	0,76	15.09–23.09.2025	0,68	11.09–23.09.2025	0,53	03.09–15.09.2025	0,77	
19.09–01.10.2025	0,80	23.09–01.10.2025	0,61	23.09–03.10.2025	0,63	15.09–25.09.2025	0,68	
01.10–13.10.2025	0,77	01.10–09.10.2025	0,66	03.10–15.10.2025	0,76	25.09–07.10.2025	0,66	
28.08.–13.10.2025	0,77	26.08.–09.10.2025	0,63	01.09.–15.10.2025	0,69	22.08.–07.10.2025	0,67	
Микроциклы (3 б/дн**)								
периоды	корр.	периоды	корр.	периоды	корр.	периоды	корр.	0,74
03.10–07.10.2025	0,78	05.09–15.09.2025	0,83	11.09–19.09.2025	0,65	03.09–11.09.2025	0,64	
07.10–09.10.2025	0,72	15.09–23.09.2025	0,87	19.09–29.09.2025	0,63	11.09–19.09.2025	0,58	
09.10–13.10.2025	0,81	23.09–01.10.2025	0,88	29.09–07.10.2025	0,61	19.09–29.09.2025	0,68	
13.10–15.10.2025	0,85	01.10–09.10.2025	0,89	07.10–15.10.2025	0,60	29.09–07.10.2025	0,72	
03–15.10.2025	0,80	01–13.10.2025	0,88	07–15.10.2025	0,62	03.09.–13.10.2025	0,66	
МЕДИАНА								0,72
Справочно								
Медиациклы (7 б/дн**)								
Период	корр.	Период	корр.	Период	корр.	Период	корр.	0,49
09.09–17.09.2025	0,80	05.09–15.09.2025	0,63	11.09–19.09.2025	0,50	03.09–11.09.2025	0,64	
17.09–25.09.2025	0,77	15.09–23.09.2025	0,57	19.09–29.09.2025	0,74	11.09–19.09.2025	0,64	
25.09–03.10.2025	0,78	23.09–01.10.2025	0,55	29.09–07.10.2025	0,55	19.09–29.09.2025	0,64	
03.10–13.10.2025	0,74	01.10–09.10.2025	0,79	07.10–15.10.2025	0,65	29.09–07.10.2025	0,84	
09.09.–13.10.2025	0,77	05.09.–09.10.2025	0,6	11.09.–15.10.2025	0,6	03.–07.10.2025	0,64	

Источник / Source: составлено автором по результатам расчетов синтезированных опережающих данных изменения конъюнктуры Мосбиржи / Compiled by the author based on the results of calculations of synthesized advanced data on changes in the MOEX.

Примечание / Note: \*коэффициенты корреляции / correlation coefficients; \*\*биржевых дней в цикле / Stock exchange days in the cycle.

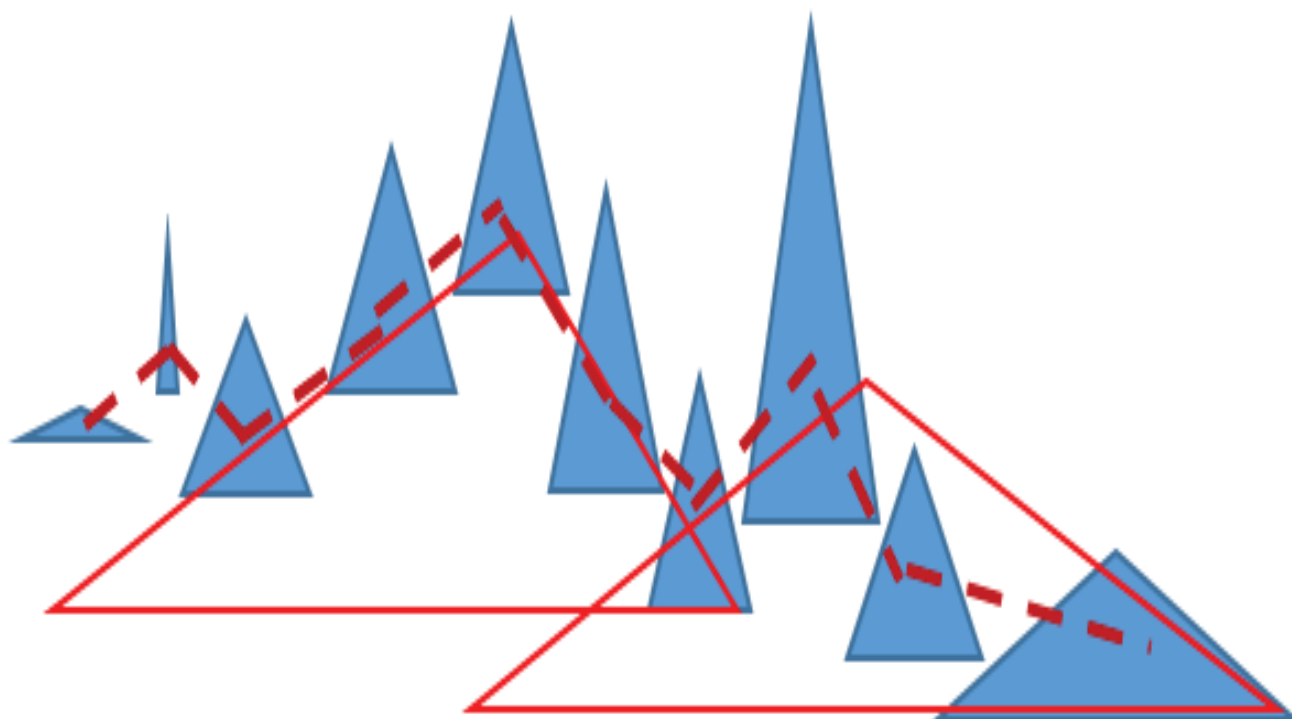
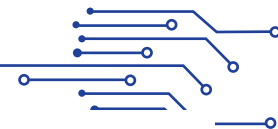


Рис. 3 / Fig. 3. Схема масштабирования фракталов опережающих индикаторов /  
The Scheme of Scaling Fractals of Leading Indicators

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

монстрируют циклическую природу: их структура сохраняется при изменении масштаба (самоподобие). Выявлена последовательность фракталов на кривой медианных значений показателей, что подтверждает их системность. Дробная размерность социальных фракталов (по Хаусдорфу–Безиковичу) и масштабируемая инвариантность указывают на фрактальный характер динамики социологических и экономических показателей.

**2. Преимущества фрактального анализа перед традиционными методами.** Численная оценка вероятности событий позволяет количественно измерять риски. Вектор волатильности дает возможность отслеживать изменения рыночной нестабильности. Обозначение периодов перелома трендов помогает выявлять точки смены рыночной динамики. Масштабируемость инвариантов обеспечивает множественность горизонтов ожиданий (микро-, медиа-, макроциклы). Численное выражение деловой активности косвенно отражает изменения рыночных оборотов.

#### Эмпирическая эффективность метода

Коэффициент корреляции динамических рядов (0,72) демонстрирует устойчивость подобия независимо от горизонта ожиданий. Микроциклы показывают более высокое подобие к макроциклам (0,74 против 0,72), что подтверждает масштабиру-

емость фрактальных структур. Медиациклы (7–9 дней) имеют умеренную корреляцию (0,69), что близко к порогу значимости (0,70). Семидневные циклы демонстрируют низкую сходимость (0,49), что указывает на ограничения метода на коротких временных отрезках.

#### Практическое применение фрактального анализа

Прогнозирование биржевой активности — метод позволяет уточнять прогнозы за счет выявления повторяющихся паттернов. Макроэкономическое планирование может служить дополнительным источником данных для стратегического прогнозирования. Управление рисками помогает заполнять информационные промежутки между регулярными отчетами (еженедельными, ежемесячными). Синтез опережающих индикаторов дает возможность комбинировать данные с другими аналитическими инструментами.

#### Ограничения и условия применения

Фрактальный анализ не является универсальным — его эффективность зависит от контекста и типа данных. Метод требует сочетания с другими подходами (технический, фундаментальный анализ), так как не гарантирует 100%-ной точности. Существуют ограничения на коротких временных интервалах (на-

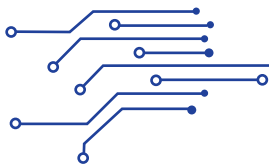




16. Колмогоров А. Н. О представлении непрерывных функций нескольких переменных суперпозициями непрерывных функций меньшего числа переменных. Доклады АН СССР. 1956;108:179-182. URL: <https://bigenc.ru/b/o-predstavlenii-nepneryvny-5e6285?ysclid=mlf1zpkbty280921913>

## REFERENCES

1. Mandelbrot B., Hudson R. L. (Not)compliant markets. Fractal revolution in finance. Moscow: Williams; 2006. 408 p. URL: [https://archive.org/details/20250615\\_20250615\\_1132](https://archive.org/details/20250615_20250615_1132) (In Russ.).
2. Peters E. Fractal analysis of financial markets. Application of chaos theory in investments and economics. Moscow: Internet Trading; 2004. 304 p. URL: <https://archive.org/details/fractalmarketana0000pete> (In Russ.).
3. Shannon B. Technical Analysis Using Multiple Timeframes. 2008. URL: <https://oceanofpdf.com/authors/brian-shannon/pdf-technical-analysis-using-multiple-timeframes-download/>
4. Zhukov D. S., Lyamin S. K. Methodology of fractal modeling in historical, political science and other social research: features of the conceptual and analytical apparatus. *Bulletin of Tambov State University. Series: Humanities*. 2010;7(87):223-234. URL: <https://www.elibrary.ru/msvjtf> (In Russ.).
5. Serikov A. E. Fractal analysis of time series. *Sociology: methodology, methods, mathematical models*. 2006;22:162-183. URL: <https://www.elibrary.ru/kvkiup> (In Russ.).
6. Dubovikov M. M., Kryanev A. V., Starchenko N. V. Index of variation and fractal analysis of time series. In collection: Scientific session of MEPhI-2004. Vol. 7: Astrophysics and Cosmophysics. Problems of modern mathematics. Beam physics and accelerator technology. 2004. URL: <https://scipeople.ru/publication/61054/?ysclid=mlf24pgbdq688239033> (In Russ.).
7. Vinberg E. B. On non-Euclidean geometry. *Mathematical education*. 2021;3(27):29-40. URL: <https://www.mathnet.ru/links/e9bf08d0c0633401ef58d725e327b182/mp994.pdf> (In Russ.).
8. Perepelitsa V. A., Tambieva D. A., Komissarova K. A. Investigation of the R/S trajectory of one insurance time series. *Researched in Russia*. 2004;7:2663-2672. URL: <http://zhurnal.apelarn.ru/articles/2004/248.pdf> (In Russ.).
9. Gregory-Williams J., Williams B. Trading Chaos: Applying Expert Techniques to Maximize Your Profits. 1958. URL: <https://archive.org/details/tradingchaosapp10000will/page/n5/mode/2up>
10. Almazov A. A. Fractal theory of the Forex market. Moscow: Admiral Markets; 2009. 296 p. URL: [https://rusneb.ru/catalog/004796\\_000040\\_TVERS-RU%7C%7C%7CTOUNB%7C%7C%7CBIBL%7C%7C%7C96508/?ysclid=mlf0yswonc151853516](https://rusneb.ru/catalog/004796_000040_TVERS-RU%7C%7C%7CTOUNB%7C%7C%7CBIBL%7C%7C%7C96508/?ysclid=mlf0yswonc151853516) (In Russ.).
11. Davydov A. A. Modular theory of society. Problems of theoretical sociology. Issue. 2. Collection of articles. Saint Petersburg: St. Petersburg University Press; 1996. URL: <https://spbib.ru/en/catalog/-/books/12280856-modul-naya-teoriya-sotsiuma?ysclid=mlf2ehtuyo838806930> (In Russ.).
12. Varyash I. Y. Research of the model of leading indicators of investment conditions. *Banking business*. 2022;21:13-21. URL: <https://www.elibrary.ru/kiywxr> (In Russ.).
13. Burova T. F., Varyash I. Yu., Panasenko K. K. Research of models of leading indicators of financing conditions. *CHRONOS*. 2019;11(98):100-114. URL: <https://www.elibrary.ru/njzzlv> (In Russ.).
14. Elokhova O. I. Fractality of virtual reality. *Bulletin of the Bashkir State University*. 2014;19(1):210-214. URL: <https://www.elibrary.ru/xhoqmr> (In Russ.).
15. Makarov S. V. Great Russian sociomaterial matryoshka. *History and historical memory*. 2024;28-29:11-24. URL: <https://www.elibrary.ru/olfazg> (In Russ.).
16. Kolmogorov A. N. On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables. Reports of the USSR Academy of Sciences. 1956;108:179-182. URL: <https://bigenc.ru/b/o-predstavlenii-nepneryvny-5e6285?ysclid=mlf1zpkbty280921913> (In Russ.).



## ИНФОРМАЦИЯ ОБ АВТОРЕ / ABOUT THE AUTHOR

**Игорь Юрьевич Варьяш** — доктор экономических наук, кандидат философских наук, профессор, руководитель Аналитического центра финансовых исследований, Научно-исследовательский финансовый институт (НИФИ), Москва, Российская Федерация

**Igor Yu. Varjas** — Dr. Sci. (Econ.), Cand. Sci. (Philos.), Honorary Prof., Head of the Analytical Center for Financial Research, Research Financial Institute (NIFI), Moscow, Russian Federation

<https://orcid.org/0000-0002-4816-8086>

[varjas@nifi.ru](mailto:varjas@nifi.ru)

[igorvarjas21@gmail.com](mailto:igorvarjas21@gmail.com)

*Конфликт интересов: автор заявляет об отсутствии конфликта интересов*

*Conflicts of Interest Statement: The author has no conflicts of interest to declare.*

*Статья поступила 19.12.2025; после рецензирования 15.01.2026; принята к публикации 20.01.2026.*

*Автор прочитал и одобрил окончательный вариант рукописи.*

*The article was submitted on 19.12.2025; revised on 15.01.2026 and accepted for publication on 20.01.2026.*

*The author read and approved the final version of the manuscript.*