

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
ВЫСШАЯ ШКОЛА ЭКОНОМИКИ

УДК: 004.8, 004.9, 519.8, 330.4

Рег. № НИОКТР АААА-А18-118051690084-7

Инв. №:

УТВЕРЖДАЮ
Проректор НИУ ВШЭ,
к.э.н.

М.М. Юдкевич

«___»_____2018 г.

ОТЧЕТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ
СОВРЕМЕННЫЙ КОНТЕКСТ МЕТОДОВ ПРИНЯТИЯ РЕШЕНИЙ И АНАЛИЗА
ДАННЫХ: ЧЕЛОВЕЧЕСКИЙ ФАКТОР, НЕОПРЕДЕЛЕННОСТЬ, РИСКИ,
СЕТЕВЫЕ МОДЕЛИ, БОЛЬШИЕ ДАННЫЕ
(заключительный)

Шифр: ТЗ-71

Руководители темы:

зав. МНУЛ анализа и выбора решений,
д.т.н.

Ф.Т. Алескеров

гл.н.с. МНУЛ анализа и выбора решений, PhD

Э. Маскин

Москва 2018

СПИСОК ИСПОЛНИТЕЛЕЙ

Соруководители темы:

Зав. лабораторией, д.т.н., рук. деп., проф.	_____	Ф.Т. Алескеров	(введение, заключение)
Гл. научн. сотр., PhD	_____	Э. Маскин	(пункты 1.1.5, 1.1.6)

Исполнители темы:

Вед. науч. сотр., д.т.н., проф.	_____	А.С. Беленький	(пункты 1.1.1 – 1.1.4)
Ст.-исследователь	_____	Ю.А. Веселова	(подготовка списка используемых источников)
Ст.-исследователь	_____	А.С. Власов	(пункты 4.5.4, 4.5.5)
Ст.-исследователь	_____	И.Е. Гавриленкова	(пункт 4.1.8)
Ст.-исследователь	_____	Ю.Э. Гурова	(пункт 4.1.9)
Ст.-исследователь, аспирант	_____	С.С. Демин	(пункты 5.1.1, 5.1.3)
Науч. сотр., PhD, доцент	_____	Э. Доган	(пункт 2.3.6 – 2.3.8)
Мл. науч. сотр., к.ф.-м.н., доцент	_____	Л.Г. Егорова	(пункты 4.2.1, 4.2.2)
Ст.-исследователь, аспирант	_____	М.Д. Ермолова	(пункты 3.1.1 -3.1.2)
Ст.-исследователь	_____	А.А. Иванов	(пункты 2.4.1 – 2.4.3)
Ст. науч. сотр., к.э.н., доцент	_____	Д.С. Карабекян	(пункт 2.4.4 – 2.4.5)
Ст. науч. сотр., к.э.н., доцент	_____	А.В. Карпов	(пункты 1.2.1, 1.2.3, 1.2.4)
Ст. науч. сотр., д.т.н.	_____	Н.А. Коргин	(подраздел 1.3)
Вед. науч. сотр., д.ф.-м.н., проф.	_____	А.А. Лазарев	(пункт 1.2.2)
Зам. зав. лаб., д.ф.-м.н., проф.	_____	А.Е. Лепский	(пункты 2.3.1 – 2.3.5)
Ст.-исследователь, аспирант	_____	Т.А. Ломоносов	(пункт 5.1.2)

Ст.-исследователь, студент _____	Н.Г. Мещерякова	(пункты 4.1.5, 4.1.10)
Вед. науч. сотр., д.т.н., проф. _____	Б.Г. Миркин	(пункты 4.5.1, 4.5.2)
Ст. науч. сотр., к.т.н., доцент _____	А.Л. Мячин	(пункты 4.2.3, 4.2.4, 4.3.3 - 4.3.5)
Ст.-исследователь, аспирант _____	А.С. Овчинников	(пункт 4.1.2)
Ст. науч. сотр., к.э.н., доцент _____	Г.И. Пеникас	(подраздел 3.2)
Вед. науч. сотр., д.т.н., проф. _____	В.В. Подиновский	(подраздел 2.1)
Ст.-исследователь, студент _____	А.Н. Резяпова	(пункт 4.1.6)
Эксперт, к.т.н., доцент _____	А.А. Рубчинский	(подраздел 4.4)
Ст.-исследователь _____	А.Э. Рявина	(пункты 4.3.1, 4.3.2)
Ст. науч. сотр., к.ф.-м.н., доц. _____	А.Н. Субочев	(подраздел 2.2)
Ст.-исследователь _____	М.А. Сурков	(пункты 3.1.3 -3.1.4)
Ст.-исследователь, аспирант _____	Д.Н. Тверской	(подраздел 5.2)
Ст.-исследователь _____	С.М. Тормышева	(пункт 4.1.1)
Ст.-исследователь _____	Д.С. Фролов	(пункт 4.5.3)
Мл. науч. сотр., аспирант _____	С.В. Швыдун	(пункты 4.1.3, 4.1.4)
Ст. науч. сотр., к.т.н. _____	В.И. Якуба	(пункт 4.1.7)

РЕФЕРАТ

Отчет 267 с., 1 ч., 28 рис., 20 табл., 293 источника.

ТЕОРИЯ ПРИНЯТИЯ РЕШЕНИЙ, ТЕОРИЯ КОЛЛЕКТИВНОГО ВЫБОРА, МЕТОДЫ ОПТИМИЗАЦИИ, ТЕОРИЯ ИГР, СЕТЕВОЙ И КЛАСТЕРНЫЙ АНАЛИЗ ДАННЫХ, ТЕОРИЯ ФУНКЦИЙ ДОВЕРИЯ, АНАЛИЗ РИСКА, АНАЛИЗ ПАТТЕРНОВ

Объектами исследования были модели и методы многокритериального принятия решений, теории коллективного выбора, теории функций доверия, сетевого анализа данных, анализа паттернов, модели в социально-экономических и банковских системах, математические модели специализаций в абстрактных системах и динамические модели распространения.

Целью проекта были разработка и исследование новых методов анализа данных и принятия решений с приложениями к различным задачам социально-экономической сферы. Особенности областей возможных применений диктуют необходимость работы с большими объемами сложно структурированных данных, учёт «человеческого фактора» (использование теоретико-игровых моделей, учёт предпочтений и влияний, и т.д.), неопределенности и/или неточности условий и данных, многокритериальность принимаемых решений и т.п.

В частности, целями исследований были:

- задача построения работоспособного инструментария для анализа выгоды для частных инвесторов капиталовложений (инвестиций) в проекты, связанные с развитием транспортной инфраструктуры;
- исследование задач теории расписания в турнирах;
- исследование конкурентной кооперации в Арктике;
- развития теории интервальных оценок замещений критериев, в частности анализ чувствительности таких оценок;
- эмпирическое исследование устойчивости агрегированных рейтингов;

- исследование мер конфликта, индуцированных правилами комбинирования свидетельств в рамках теории функций доверия;
- исследование структур мер конфликта в рамках теории функций доверия;
- исследования степени манипулируемости правил агрегирования предпочтений;
- анализ методов валидации моделей оценки вероятности дефолта;
- исследование критического уровня мошенничества в банковской системе с помощью агентно-ориентированного подхода;
- исследование по сетевому анализу данных, в частности анализ вычислительной сложности сетевых индексов ближнего и дальнего взаимодействия и построение быстрых алгоритмов их вычисления;
- оценка непрямого влияния в сетевых структурах;
- анализ международной миграции;
- анализ межстрановой торговли продовольствием;
- анализ террористических сообществ;
- анализ сообщества экономистов и построение рейтингов экономических журналов;
- развитие методов анализа паттернов;
- разработка новых алгоритмов частотного оценивания структурной сложности графов;
- исследования по кластерному анализу данных;
- исследование моделей распространения загрязнений на поверхности водоемов;
- исследование моделей специализации в абстрактных системах.

Методологической основой исследований были теория игр и теория многокритериального принятия решений, теория коллективного и индивидуального выбора, теория функций доверия, интеллектуальный и сетевой анализ данных. Основными методами исследований были методы оптимизации,

анализа данных, методы теории вероятностей, методы поиска решений на сетях, методы кластеризации и др.

Для прикладных исследований использовались данные и открытые базы данных из различных источников. Некоторые исследования были выполнены численно с использованием различного прикладного программного обеспечения. Другие исследования подразумевали и сопровождались разработкой комплексов прикладных программ.

В ходе выполнения проекта получены следующие новые теоретические и прикладные результаты:

- математически поставлены задачи об оценке окупаемости требуемых инвестиций в строительство транспортного узла в какой-либо конкретной точке региона и об оценке срока окупаемости этих инвестиций; показано, что обе задачи, которые формулируются как задачи отыскания максимума функции максимина, сводятся к задаче квадратичного программирования, в то время как отыскание наиболее неблагоприятного сценария формирования рыночных транспортных тарифов на перевозки грузов сводится к решению задач линейного программирования, образующих двойственную пару;
- сформулированы и доказаны проверяемые достаточные условия конкурентоспособности тарифов на услуги транспортного узла в условиях рынка, которые использованы при формулировке систем ограничений в обеих задачах;
- в рамках исследования теории расписаний в турнирах для случаев трех и четырех участников в одном матче построены обобщения стандартного расписания и расписания равных преимуществ; показано, что данные случаи являются наиболее значимыми с практической точки зрения;
- представлена первая версия концепции платформы для исследования проблем и возможностей по кооперации в Арктике на базе моделей конкурентной кооперации в идеологии многоуровневого подхода к декомпозиции сложных кибер-физических систем;

- в рамках теории многокритериального выбора предложен и исследован подход к анализу чувствительности недоминируемых альтернатив к изменению границ интервалов неопределенности замещения, где альтернативы выбираются на основе интервальных данных оценок замещения критериев;
- в рамках анализа способов выбора оптимальных вариантов и процедур ранжирования альтернатив выполнено эмпирическое исследование устойчивости агрегированных рейтингов, строящихся с помощью ординальных методов, основанных на парных сравнениях по правилу большинства;
- в рамках теории функций доверия исследованы меры конфликта, индуцированных конъюнктивными и дизъюнктивными правилами комбинирования свидетельств; рассмотрены условия согласования правил комбинирования с мерами конфликта;
- в рамках теории функций доверия показано, что в конъюнктивных мерах конфликта можно выделить метрическую и энтропийную составляющие; исследованы свойства изменений этих составляющих при комбинировании свидетельств;
- в рамках исследования степени манипулируемости правил агрегирования предпочтений предложена и исследована модель индивидуального и коалиционного манипулирования правилами агрегирования предпочтений в одномерной модели формирования предпочтений; для трех альтернатив выявлены минимально манипулируемые группы мажоритарных правил;
- в рамках анализа в банковской сфере рассмотрено применение модели биномиального теста коррелированных случайных бинарных величин для задачи проверки точности прогноза количества дефолтов;
- проведена оценка влияния мошеннических операций на уязвимость всей финансовой системы; предложен агентно-ориентированный подход к моделированию критических уровней мошеннической деятельности, которые делают финансовую систему уязвимой; этот подход применен для исследо-

- вания таких типов мошенничества, как финансовая пирамида и скрытый риск;
- на основе разработанных ранее моделей влияния в сетях предложен эффективный алгоритм вычисления индексов ближних и дальних взаимодействий; оценена вычислительная сложность этого алгоритма;
 - с помощью сетевых моделей проанализированы отношения между странами на различных уровнях взаимодействия, а именно, выполнен сетевой анализ: а) международной миграции с 1970 по 2014гг. на основе данных по потокам миграции между странами; б) торговли продовольствием; в) террористических сообществ;
 - проанализирована неоднородность сообщества российских экономистов по результатам анкетирования; предложены несколько методов составления рейтингов научных журналов по экономике с учетом выявленной неоднородности сообщества российских экономистов;
 - в рамках исследования паттернов сформулированы и доказаны свойства порядково-инвариантной паттерн-кластеризации; эти свойства показывают, что для практической реализации порядково-инвариантной и диффузионно-инвариантной паттерн-кластеризаций могут быть использованы быстрые алгоритмы сортировки данных;
 - в рамках исследования структурной сложности графа предложен новый, вычислительно более мощный вариант частотного алгоритма построения семейства дихотомий произвольного графа, что позволяет вычислить такой показатель сложности графа, как сцепленность; этот алгоритм применен к данным фондового рынка S&P-500 для анализа согласованности его функционирования;
 - проведено экспериментальное сравнение алгоритмов кластеризации, основанных на методе аффинного распространения (Affinity Propagation) с другими алгоритмами кластеризации с точки зрения правильности определения количества кластеров и оценки качества полученного

- разбиения по сравнению с истинным разбиением; численные эксперименты проведены на данных с различной степенью смешивания кластеров; определена иерархия качества алгоритмов кластеризации, основанных на методе аффинного распространения, для разных типов данных;
- разработан вычислительный комплекс, позволяющий моделировать процесс распространения нефти на поверхности Баренцева моря на основе агентно-сетевого подхода;
 - исследованы явные двухуровневые конечно-разностные схемы решения одномерных уравнений баротропной газовой динамики; линеаризованы схемы на постоянном решении и выведено необходимое условие типа фон Неймана и критерий типа CFL для слабой консервативности в L^2 для соответствующей задачи на всей прямой;
 - предложена и исследована новая фитнес-оптимизационная модель, описывающая влияние факторов окружающей среды на возможности возникновения репродуктивно-соматической специализации в колониях клеток.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	13
ОСНОВНАЯ ЧАСТЬ.....	23
1 Исследование оптимизационных моделей и механизмов	23
1.1 Исследование моделей по оптимальному размещению объектов и привлечению частных инвестиций	23
1.1.1 Введение.....	23
1.1.2 Достаточные условия конкурентоспособности тарифов на услуги транспортного узла по перевалке грузов.	32
1.1.3 Постановки задач и их математические формулировки	35
1.1.4 Краткий обзор публикаций по развитию региональных транспортных узлов	40
1.1.5 Обсуждение	45
1.1.6 Заключение по подразделу	47
1.2 Исследование задач теории расписания в турнирах	49
1.2.1 Введение.....	49
1.2.2 Описание модели расписаний.....	51
1.2.3 Теоремы о представлении	59
1.2.4 Заключение по подразделу	63
1.3 Исследование конкурентной кооперации в Арктике	63
2 Исследование моделей выбора и агрегирования.....	67
2.1 Развитие теории интервальных оценок замещений критериев.....	67
2.1.1 Введение в теорию интервальных оценок замещений критериев.....	67
2.1.2 Сведения из теории параметрической важности критериев и теории интервалов неопределенности замещений	71
2.1.3 Анализ чувствительности.....	75
2.1.4 Пример проведения анализа чувствительности	80
2.1.5 Заключение по подразделу	86
2.2 Эмпирическое исследование устойчивости агрегированных рейтингов.....	87
2.3 Исследование мер конфликта в рамках теории функций доверия, индуцированных правилами комбинирования свидетельств.....	94
2.3.1 Введение и постановка задачи	94
2.3.2 Основные понятия теории функций доверия	96
2.3.3 Аксиоматика мер конфликта.....	97
2.3.4 Меры конфликта индуцированные конъюнктивными и дизъюнктивными правилами	98
2.3.5 Меры конфликта, согласованные с правилами комбинирования	100
2.3.6 Метрическая и энтропийная составляющие меры конфликта	102
2.3.7 Изменение метрической и энтропийной частей меры конфликта	106
2.3.8 Заключение по подразделу	107
2.4 Построение и применение теоретических моделей для расчета степени манипулируемости правил агрегирования предпочтений	108
2.4.1 Модель манипулирования	108

2.4.2 Правила выбора	109
2.4.3 Индекс манипулируемости и методика расчета	112
2.4.4 Результаты расчетов	112
2.4.5 Некоторые выводы.....	115
3 Анализ моделей в финансовой и банковской сферах	116
3.1 Анализ методов валидации моделей оценки вероятности дефолта.....	116
3.1.1 Введение и постановка задачи	116
3.1.2 Модель дефолтов по Васичеку и на основе бинарного распределения.....	118
3.1.3 Эмпирическое распределение уровня частоты дефолтов	120
3.1.4 Заключение и выводы.....	125
3.2 Исследование критического уровня мошенничества в банковской системе: агентно-ориентированный подход	127
3.2.1 Введение постановка задачи	127
3.2.2 Обзор литературы	129
3.2.3 Методология исследования.....	132
3.2.4 Результаты	133
4 Разработка и применение методов анализа данных и прогнозирования	134
4.1 Исследование по сетевому анализу данных.....	134
4.1.1 Введение.....	134
4.1.2 Описание индексов взаимодействия	137
4.1.3 Вычислительная сложность индексов SRIC и LRIC	140
4.1.4 Вычислительные эксперименты	145
4.1.5 Оценка непрямого влияния в сетевых структурах.....	148
4.1.6 Анализ международной миграции	150
4.1.7 Анализ межстрановой торговли продовольствием.....	154
4.1.8 Анализ террористических сообществ	158
4.1.9 Индексы, учитывающие предпочтения индивидов по созданию коалиций.....	161
4.1.10 Некоторые выводы.....	166
4.2 Анализ сообщества экономистов и построение рейтингов экономических журналов	168
4.2.1 Введение и постановка задачи	168
4.2.2 Данные и методология.....	171
4.2.3 Результаты	177
4.2.4 Некоторые выводы.....	183
4.3 Развитие методов анализа паттернов.....	184
4.3.1 Введение.....	184
4.3.2 Анализ паттернов: общее описание	185
4.3.3 Анализ паттернов, основанный на парном сравнении при фиксированной последовательности показателей	187
4.3.4 Анализ паттернов с независимыми от выбора исходной последовательности показателей результатами.....	188
4.3.5 Некоторые выводы.....	190

4.4 Разработка нового частотного механизма оценивания структурной сложности графов.....	190
4.4.1 Введение и постановка задачи	190
4.4.2 Частотный алгоритм дихотомии.....	191
4.4.3 Сцепленность графа.....	195
4.4.4 Анализ фондовых рынков	199
4.4.5 Некоторые выводы.....	201
4.5 Исследования по кластерному анализу данных	202
4.5.1 Введение и постановка задачи	202
4.5.2 Краткий аналитический обзор алгоритмов кластеризации.....	203
4.5.3 Метод аффинного распространения.....	205
4.5.4 Описание экспериментов	213
4.5.5 Некоторые выводы.....	220
5 Исследование моделей распространения и развития.....	222
5.1 Исследование моделей распространения	222
5.1.1 Введение и постановки задач.....	222
5.1.2 Основная часть	223
5.1.3 Некоторые выводы.....	225
5.2 Исследование моделей специализации в абстрактных системах.....	226
5.2.1 Введение и постановка задачи моделирования специализации	226
5.2.2 Основная часть	229
5.2.3 Основные результаты	231
ЗАКЛЮЧЕНИЕ.....	234
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	238

ВВЕДЕНИЕ

В отчете представлены результаты работы в 2018 году сотрудников Международной научно-учебной лаборатории анализа и выбора решений (МЛАВР) НИУ ВШЭ над проектом «Современный контекст методов принятия решений и анализа данных: человеческий фактор, неопределенность, риски, сетевые модели, большие данные».

В соответствии с техническим заданием были проведены исследования в следующих научных направлениях:

- исследование оптимизационных моделей и механизмов;
- исследование моделей выбора и агрегирования;
- анализ моделей банковской сфере;
- разработка и применение методов анализа данных и прогнозирования;
- исследование моделей распространения и развития.

Работа над проектом этого года стала продолжением исследований, проводимых в Международной научно-учебной лаборатории анализа и выбора решений в 2017 г. в ходе работы над проектом «Анализ, выбор и принятие решений в социально-экономической, политической и финансовой сферах: новые модели, методы и алгоритмы» [1]. В частности, в 2017 г. были получены следующие новые теоретические и прикладные результаты:

- в рамках решения задачи размещения транспортных узлов с подъездными путями в некоторой региональной транспортной системе представлена классификация моделей используемых при решении подобных задач; предложена математическая модель, учитывающая наличие существующей транспортной системы региона; на основе представленной модели сформулированы оптимизационная (в случае известных цен на транспортировку и строительство) и робастная задачи (точные значения цен не известны); обе задачи протестированы на данных о транспортной системе части России;
- представлены результаты по модификации, разработанной в 2016 году математической модели, и на основе этой модификации сформулирована за-

- дача об оценке: а) требуемого объема инвестирования в проект по строительству заводов по переработке твердых бытовых отходов в биомассу, используемую как топливо в системах электроснабжения и отопления; б) начальных цен двух контрактов на строительство заводов по переработке отходов в топливную биомассу в рамках концессионного соглашения между частными инвесторами и городскими властями в виде игры трех лиц; доказана возможность отыскания точек равновесия Нэша этой игры из решения трех вспомогательных задач линейного программирования;
- разработана математическая модель и сформулирована оптимизационная задача оценки объема инвестиций, необходимых для разработки цепочки зарядных станций на автомобильных магистралях при линейных ограничениях; доказано, что оптимальное решение этой задачи может быть найдено из решения вспомогательной задачи смешанного математического программирования;
 - в рамках моделирования предпочтений при помощи функций ценности выделен и исследован новый интересный для приложений специальный вид функций, представляющих отношения полного порядка на конечных множествах – реперные функции; рассмотрены реперные функции полезности и реперные функции вредности;
 - предложена постановка задачи анализа чувствительности многокритериального выбора к изменению оценок параметрической важности критериев и, в частности, к изменению интервалов неопределенности замещений критериев; предложен общий метод проведения такого анализа для отдельных недоминируемых альтернатив и для множества альтернатив в целом;
 - в рамках сравнительного исследования функций коллективного выбора, зависящих от коллективных предпочтений, проведено исследование функций коллективного выбора, основанных на принципе устойчивости; для этих решений установлены некоторые свойства и их теоретико-множественные соотношения; установлены некоторые свойства процедур построения агрегированных ранжирований, основанных на рассмотренных правилах;

- исследованы индексы поляризованности мнений и разнородности предпочтений в группе; проведено сравнение существующих подходов к измерению разнообразия и выделены аксиоматические свойства индексов; построены новые аксиоматически обоснованные индексы измерения поляризованности и разнообразия предпочтений;
- в рамках исследования коалиционного манипулирования при неполной информации показано, что: а) для правил подсчета очков вероятность коалиционного манипулирования равна вероятности индивидуального манипулирования при публичной информации о победителях голосования; б) коалиционное манипулирование при неполной информации всегда безопасно для правила относительного большинства и правила вето и может быть небезопасно для правила Борда; проведена серия экспериментов для вычисления оценок вероятностей коалиционного манипулирования при неполной информации;
- на основе 24 200 гипотетических портфелей, отличающихся уровнем риска, рассчитан мультипликатор риск-веса, корректирующий требования Базельского комитета для покрытия рисков, возникающих вследствие концентрации портфеля и корреляции вероятности дефолта (PD) и доли потерь при дефолте (LGD);
- построена модель вероятности дефолта методом бинарной и категориальной логистической регрессии на выборке из 192 судоходных компаний 36 стран за период с 2001 по 2016;
- предложена модель анализа финансовых показателей с помощью процессов Хоукса с целью выявления кризисов на фондовом рынке и предупреждения их появления; модель протестирована на данных фондового индекса S&P500 и валютной пары USD/RUB;
- исследовано агрегирование рекомендаций финансовых аналитиков в рамках теории функций доверия на примере российского фондового рынка; показано, что правила комбинирования, которые широко развиты в теории

- функций доверия, позволяют агрегировать рекомендации аналитиков с учетом исторической надежности источников информации, характера принимаемых решений (пессимистичность-оптимистичность), конфликта между прогнозами и рекомендациями и т.д.; показано, что полученные агрегированные прогнозы в большинстве случаев оказываются точнее консенсус-прогнозов;
- в рамках анализа международной миграции разработана сетевая модель и проанализированы индексы центральности; исследована взаимосвязь численности международных мигрантов по странам и социально-экономических показателей стран (ВВП, уровень образования, интенсивность конфликтов);
 - предложены новые модели распределения конфликтных зон между заинтересованными участниками, которая была апробирована для Арктического региона; проведен анализ потенциальных конфликтных зон в Баренцевом море и их сравнение с распределением, предложенным в международных договорах;
 - изучены модели распространения нефтепродуктов в море, с помощью которых будет возможно прогнозировать данный процесс; построена динамическая модель распространения нефти в море с учётом различных природных факторов на основе предлагаемых нами моделей; разработан вычислительный комплекс, позволяющий визуализировать процесс распространения нефти во времени;
 - в рамках анализа сети продовольствия проанализированы сети международной торговли; выявлены страны, имеющие наибольшее влияние на других участников сети по каждой категории продуктов; исследовано изменение со временем ключевых стран в сетях продовольствия;
 - на основе предложенных ранее двух новых методов анализа паттернов разработаны численные методы расчета индексов неоднородности инноваци-

онного развития с использованием данных Глобального индекса инноваций;

- в рамках разработки методов анализа структурной сложности графов рассмотрены и исследованы новые модификации алгоритма анализа структурной сложности графа, основанные на нормализованном значении энтропии; эта модификация позволила ускорить и увеличить точность вычисления сложности данных фондового рынка в задаче прогнозирования кризисных периодов;
- исследованы подходы к операционализации полярности в международных исследованиях и изучена связь полярности с вероятностью возникновения конфликтов.

Некоторые исследования 2018 года, выполненные в рамках проекта «Современный контекст методов принятия решений и анализа данных: человеческий фактор, неопределенность, риски, сетевые модели, большие данные», являются продолжением исследований прошлого года. Прежде всего, речь идет о задаче размещения транспортных узлов, о задаче анализа чувствительности многокритериального выбора к изменению оценок параметрической важности критериев, о сетевом анализе международной миграции и международной торговли продовольствием, об исследовании модели распространения нефтепродуктов в море, о развитии методов и алгоритмов анализа структурной сложности графов. Вместе с тем ряд рассмотренных задач этого года являются новыми. Например, задача о составлении оптимальных расписаний в турнирах, об исследовании мер конфликта согласованных с правилами комбинирования свидетельств в рамках теории функций доверия, об анализе вычислительной сложности нахождения индексов ближнего и дальнего взаимодействий в сетях и построении быстрых алгоритмов их вычисления, об анализе структуры сообщества российских экономистов и т.д.

Предполагается, что некоторые исследования этого года будут продолжены в дальнейшем.

Методологической основой исследований являются теория игр и теория многокритериального принятия решений, теория коллективного и индивидуального выбора, теория функций доверия, интеллектуальный и сетевой анализ данных. Основными методами исследований были методы оптимизации, анализа данных, методы теории вероятностей, методы поиска решений на сетях, методы кластеризации и др.

Для прикладных исследований использовались данные и открытые базы данных из различных источников. Некоторые исследования были выполнены численно с использованием различного прикладного программного обеспечения. Другие исследования подразумевали и сопровождались разработкой комплексов прикладных программ.

Отчет состоит из 5 разделов, соответствующих основным направлениям исследований.

В первом разделе представлены результаты по исследованию оптимизационных моделей и механизмов. Эти исследования осуществлялись по нескольким направлениям.

В подразделе 1.1 приведены некоторые результаты решения проблемы построения работоспособного инструментария для анализа выгодности для частных инвесторов капиталовложений (инвестиций) в проекты, связанные с развитием транспортных инфраструктур. В частности, поставлены и проанализированы задачи об оценке окупаемости требуемых инвестиций в строительство транспортного узла и об оценке срока окупаемости этих инвестиций.

В подразделе 1.2 представлены результаты исследований теории расписаний в турнирах. В ходе этих исследований, в частности, построены обобщения стандартного расписания и расписания равных преимуществ для случаев трех и четырех участников в одном матче. Показано, что данные случаи являются наиболее значимыми с практической точки зрения.

В подразделе 1.3 представлена первая версия концепции платформы для исследования проблем и возможностей по кооперации в Арктике на базе мо-

делей конкурентной кооперации в идеологии многоуровневого подхода к декомпозиции сложных кибер-физических систем.

Второй раздел отчета содержит результаты исследований моделей выбора и агрегирования. Задачи выбора рассматривались в рамках развития теории многокритериального выбора, а также применения теории коллективного выбора. Задачи агрегирования рассматривались опять-таки в рамках теории коллективного выбора для построения агрегированных ранжирований альтернатив, а также в рамках теории функций доверия. Полученные результаты отражены в трех подразделах.

Так, в подразделе 2.1 в рамках теории многокритериального выбора приведены результаты исследований подхода к анализу чувствительности недоминируемых альтернатив к изменению границ интервалов неопределенности замещения.

В подразделе 2.2 в рамках анализа способов выбора оптимальных вариантов и процедур ранжирования альтернатив приведены результаты эмпирического исследования устойчивости агрегированных рейтингов, строящихся с помощью ординальных методов, основанных на парных сравнениях по правилу большинства.

В подразделе 2.3 приведены результаты исследования мер конфликта в рамках теории функций доверия. В частности, введено и исследовано понятие мер конфликта, индуцированных конъюнктивными и дизъюнктивными правилами комбинирования свидетельств. Кроме того, показано, что в конъюнктивных мерах конфликта можно выделить метрическую и энтропийную составляющие, которые имеют разный характер изменений при комбинировании свидетельств.

Подраздел 2.4 посвящен исследованию степени манипулируемости правил агрегирования предпочтений. Эта работа ведется на протяжении последних нескольких лет в МЛАВР. В этом году предложена и исследована модель индивидуального и коалиционного манипулирования правил агрегирования предпочтений в одномерной модели формирования предпочтений.

Третий раздел отчета посвящен результатам исследований моделей банковской сферы. В подразделе 3.1 рассмотрено применение модели биномиального теста коррелированных случайных бинарных величин для задачи проверки точности прогноза количества дефолтов.

В подразделе 3.2 проведена оценка влияния мошеннических операций на уязвимость всей финансовой системы. Предложен и частично исследован агентно-ориентированный подход к моделированию критических уровней мошеннической деятельности, которые делают финансовую систему уязвимой.

В четвертом разделе отчета приведены результаты по разработке и применению методов анализа данных и прогнозирования. Так в подразделе 4.1 отражены результаты исследований методов сетевого анализа данных и по применению этих методов к анализу ряда конкретных сетей. В частности, в этом подразделе приведены результаты разработки эффективного алгоритма вычисления индексов ближних и дальних взаимодействий. Кроме того, приведены результаты сетевого анализа: а) международной миграции с 1970 по 2014 гг.; б) торговли продовольствием; в) террористических сообществ.

В подразделе 4.2 приведены результаты анализа структуры сообщества экономистов России и построение рейтингов экономических журналов. В частности, проанализирована неоднородность сообщества российских экономистов по результатам проведенного анкетирования. Кроме того, предложены несколько методов составления рейтингов научных журналов по экономике с учетом выявленной неоднородности сообщества российских экономистов;

Подраздел 4.3 посвящен результатам исследования методов анализа паттернов. В частности, в этом подразделе сформулированы и доказаны свойства порядково-инвариантной паттерн-кластеризации. Эти свойства показывают, что для практической реализации порядково-инвариантной и диффузионно-инвариантной паттерн-кластеризаций могут быть использованы быстрые алгоритмы сортировки данных.

В подразделе 4.4 описана разработка нового, вычислительно более мощного варианта частотного механизма оценивания структурной сложности графов. Этот алгоритм позволяет вычислить такой показатель сложности графа, как сцепленность. Алгоритм применен к данным фондового рынка S&P-500 для анализа согласованности его функционирования.

Подраздел 4.5 посвящен результатам исследований по кластерному анализу данных. В частности, проведено экспериментальное сравнение алгоритмов кластеризации, основанных на методе аффинного распространения (Affinity Propagation) с другими алгоритмами кластеризации с точки зрения правильности определения количества кластеров и оценки качества полученного разбиения по сравнению с истинным разбиением.

Пятый раздел отчета посвящен результатам исследований моделей распространения и развития. Наряду с традиционным подходом динамического моделирования, основанного на построении систем дифференциальных уравнений, в этом разделе представлены результаты, в которых широко применяется агентное моделирование. Так в подразделе 5.1 представлены некоторые результаты моделирования процессов распространения нефти на поверхности Баренцева моря на основе агентно-сетевого подхода. На основе такого подхода разработан вычислительный комплекс по моделированию процессов распространения нефти на водной поверхности.

Кроме того, в этом же подразделе рассмотрены результаты анализа двухуровневых конечно-разностных схем решения одномерных уравнений баротропной газовой динамики.

В подразделе 5.2 приведены результаты исследований моделей возникновения специализаций в абстрактных системах. Предложена и исследована новая фитнес-оптимизационная модель, описывающая влияние факторов окружающей среды на возможности возникновения репродуктивно-соматической специализации в колониях клеток.

Все полученные результаты были своевременно апробированы на представительных международных научных конференциях, прошли апробацию

на ряде семинаров. Большинство описанных в отчете результатов опубликованы или представлены к публикации.

ОСНОВНАЯ ЧАСТЬ

В ходе выполнения проекта «Современный контекст методов принятия решений и анализа данных: человеческий фактор, неопределенность, риски, сетевые модели, большие данные» были проведены исследования в следующих научных направлениях:

- исследование оптимизационных моделей и механизмов;
- исследование моделей выбора и агрегирования;
- анализ моделей банковской сфере;
- разработка и применение методов анализа данных и прогнозирования;
- исследование моделей распространения и развития.

1 Исследование оптимизационных моделей и механизмов

В этом разделе отчета приведены результаты исследований оптимизационных моделей и механизмов. Работа проводилась в следующих направлениях:

- исследование моделей по оптимальному размещению объектов и привлечению частных инвестиций;
- исследование задач теории расписания в турнирах;
- исследование конкурентной кооперации в Арктике.

1.1 Исследование моделей по оптимальному размещению объектов и привлечению частных инвестиций¹

1.1.1 Введение

Развитие региональных инфраструктур является одной из важнейших задач в любой стране и требует от ее руководства принятия стратегических решений. Такие решения должны базироваться на количественном анализе

¹ Результаты этого подраздела отчета были выполнены совместно с Г. Фединым.

состояния и проблем развития региона и, как правило, подразумевают как строительство новых, так и модификацию успешно функционирующих объектов и систем, а также перепрофилирование, закрытие и даже ликвидацию неэффективно работающих объектов и систем. В свою очередь, такой анализ подразумевает разработку соответствующих математических моделей, формализующих закономерности функционирования существующих и планируемых к введению в эксплуатацию объектов и систем и формулировку на их основе математических задач, формализовано описывающих свойства и особенности предметов моделирования.

Транспортные инфраструктуры являются примером региональных инфраструктур, для которых такой анализ имеет особую важность, поскольку развитие этих инфраструктур осуществляется в рамках крупномасштабных проектов и сопряжено со значительными финансовыми затратами администрации региона. Более того, эти затраты, как правило, настолько велики, что не могут быть сделаны исключительно за счет бюджета региона и требуют привлечения частного капитала для их реализации. Такое привлечение, однако, возможно, только если администрации региона удастся убедить частных инвесторов в том, что их инвестиции в развитие транспортной инфраструктуры региона в какой-либо форме являются для них выгодными. Поэтому создание инструментария для оценки потенциальной прибыли частных инвесторов в виде систем поддержки принятия решений, доступного для использования, как администрацией региона, так и самими частными инвесторами, представляется чрезвычайно важным [2].

В тоже время именно для транспортных инфраструктур такие системы поддержки принятия решений, позволяющие заинтересовать частных инвесторов перспективными проектами, принадлежат к числу наиболее сложных, прежде всего с точки зрения моделирования функционирования этих инфраструктур. По этой причине исследования в части разработки математических моделей, которые могут быть использованы при построении указанного инструментария для транспортных инфраструктур, являются чрезвычайно акту-

альными и составляют предмет настоящего отчета. Эти модели базируются на математических моделях размещения транспортных узлов, которые ранее были рассмотрены авторами в отчете лаборатории за 2017 год, а также на новых моделях, предложенных авторами в [2].

Следует отметить, что, несмотря на экономическую важность проблемы построения работоспособного инструментария для анализа целесообразности капиталовложений в крупномасштабные проекты для частных инвесторов и прежде всего на региональном и федеральном уровне в научной литературе ей не уделяется должного внимания. В результате, многие проекты, в которых нуждаются сегодня регионы, оказываются не реализованными и даже не обсуждаются и, насколько известно авторам, развитие транспортных инфраструктур принадлежат к числу именно этих проектов. В то время как о частных инвестициях в эту отрасль говорят многие и в разных странах, ни самого указанного выше инструментария, ни математических моделей, на основе которых такой инструментарий можно было бы построить, пока что не было предложено.

В настоящем подразделе отчета рассматриваются именно математические модели, которые, могут служить основой для построения работоспособного инструментария для анализа целесообразности капиталовложений в крупномасштабные проекты для частных инвесторов, в том числе на транспорте.

Известно, что любое обсуждение экономического развития региона обычно связано с улучшением существующей в нем системы транспортировки грузов, особенно если это улучшение частично связано с преимуществами географического положения региона. Какие бы региональные услуги на перевозку грузов ни рассматривались – между пунктами в пределах региона, между пунктами в других регионах и пунктами в регионе, а также транзита через территорию региона – вышеупомянутые преимущества географического положения региона в конечном итоге обеспечивают более быструю и более дешевую доставку грузов.

Пусть, например, пункт назначения расположен в части области, близкой, скажем, к речному порту, в то время как все грузы в настоящее время доставляются в этот пункт на грузовиках. Тогда можно рассмотреть перевалку указанных грузов в пункт назначения через этот речной порт и оценить стоимость организации в таком порту грузового транспортного узла. Чтобы сделать перегрузку грузов в этом транспортном узле реальной, региональные власти должны либо построить новый речной порт на месте этого речного порта, либо построить новый порт в определенном, близлежащем к указанному речному порту географическом пункте. Вместо строительства нового порта власти также могут модернизировать речной порт, который уже функционирует. Для экономической оценки любого из этих вариантов, региональные власти должны исходить из: а) объемов грузовых потоков, которые должны быть перемещены через этот географический пункт во всех направлениях, а также транспортных тарифов на услуги, которые, как ожидается, будут предлагаться участниками рынка грузовых перевозок в последующие годы и б) стоимости строительства и эксплуатации транспортного узла в этом географическом пункте. Оценки всех этих чисел позволят администрации региона рассчитать годовую прибыль, которую этот проект может генерировать при всех возможных сценариях (разумеется, предполагая, что структура проекта принципиально позволяет генерировать прибыль) и решить, будет ли интересен этот проект для региона.

Объемы грузовых потоков. Все три набора чисел (объемы грузовых потоков, тарифы на транспортные услуги и расходы на услуги по строительству) имеют вероятностную природу, тогда как их оценки лежат в основе стратегических финансовых решений, которые администрация должна принимать. Таким образом, любые стратегии, которые могут гарантировать наилучшие экономические результаты для региона при любых возможных комбинациях упомянутых цифр должны интересовать региональные администрации больше всего.

Представляется почти невозможным надежно предсказать значения объемов потоков грузов, скажем, в следующем десятилетии на основе информации об этих потоках за предыдущие два десятилетия. Однако, нижние оценки этих объемов могут быть получены в результате анализа динамики а) развития экономики региона и соседних территорий, включая открытие новых и разработку уже обнаруженных полезных ископаемых, б) внутренней и международной торговли в рамках трех вышеупомянутых схем движения грузов по территории региона и в) миграционных процессов и рождаемости в регионе и в стране в целом.

Таким образом, если номенклатура грузов, которые считаются подлежащими перемещению через определенный географический пункт, состоит из N наименований, а ожидаемые объемы этих грузов образуют вектор $y \in R_+^N$, то оценки этих объемов в течение определенного периода времени можно описать векторным неравенством $y \geq a$, где $a \in R_+^N$ – вектор, компонентами которого являются нижние оценки соответствующих грузопотоков.

Тарифы на услуги по перевозке грузов на рынке транспортных услуг, в основном, зависят от: а) развития видов грузового транспорта; б) числа игроков на этом рынке (т. е., числа компаний, предлагающих транспортные услуги) и с) объемов грузопотоков, предлагаемых для перевозки грузоотправителями. Как правило, ожидаемые транспортные тарифы на перевозку груза на региональном рынке могут быть описаны системой линейных неравенств балансового вида [3, 4, 5, 6]. В векторно-матричной форме эта система может быть записана, например, как $Bb \leq \theta$, где B – матрица с вещественными элементами и b, θ – векторы соответствующих размеров с вещественными компонентами.

Стоимость строительных услуг, предоставляемые в любой сфере деятельности человека существенно зависит от размера комплекса, который должен быть построен. В случае грузового транспортного узла, который предполагается построить, этот размер зависит от а) объемов грузовых потоков, которые будут перемещаться через этот узел, б) типов грузов и техноло-

гических требований, связанных с присутствием этих грузов на территории узла и с) естественных территориальных ограничений на размер территории, которую этот узел может занимать. Существуют различные подходы к определению как оптимальной структуры всего грузового транспортного узла, так и оптимального размера отдельных частей этого узла и оптимальной последовательности всех строительных работ, которые должны быть выполнены в рамках строительства этого узла [7, 8, 9].

Ясно, что если, например, было бы известно ежегодное количество контейнеров всех размеров, которые, как ожидается, будут проходить через контейнерный терминал порта (контейнерные терминалы порта), то эти числа могли бы (грубо) оцениваться как числа, пропорциональные размеру этого терминала (размерам этих терминалов) с определенным коэффициентом (коэффициентами) пропорциональности. В свою очередь, этот размер терминала (терминалов) определил бы список строительных работ по созданию такого контейнерного терминала (таких терминалов) и объем каждой работы из списка. Так как, как правило, стоимость каждой работы из всего списка работ может быть известна только приближенно, в определенных числовых пределах, естественно предположить, что стоимость строительства терминала (терминалов) можно описать билинейной функцией двух векторных аргументов. Первый аргумент будет отражать объемы ожидаемых работ из списка работ, а второй аргумент будет описывать ожидаемую стоимость каждой из работ, включенных в вышеупомянутый список.

Поскольку аналогичные рассуждения применимы ко всем типам грузов, которые, как ожидается, будут перемещаться через транспортный узел, можно сделать вывод том, что стоимость строительных услуг для этого узла может быть описана билинейной функцией в форме скалярного произведения двух векторных аргументов. Первый из аргументов – это вектор с компонентами, значения которых являются объемами строительных работ, представленных в списке таких работ, подлежащих выполнению в отношении каждого объекта, с тем, чтобы этот объект стал частью транспортного узла. Эти

объемы зависят от объемов грузов, которые, как ожидается, будут обработаны (перегружены) в узле. Второй векторный аргумент – это вектор с компонентами, значения которых являются затратами на все эти строительные работы. Структура функции и векторов совершенно аналогична тем, которые были описаны выше для контейнерных терминалов.

Здесь набор допустимых значений первого векторного аргумента может быть описан системой линейных уравнений, связывающей векторы объемов всех строительных работ и векторы объемов всех транспортных потоков, которые, как ожидается, будет проходить (перегружаться) через узел (в узле). Эта система может иметь форму $E^p w - Hx = 0$, $w \in R_+^p$, $x \in R_+^N$, где p – число пунктов в номенклатуре всех строительных работ, подлежащих выполнению, E^p – единичная матрица размера p , H – $p \times N$ -матрица с неотрицательными элементами, а x – вектор с компонентами, равными объемам грузов, которые, как ожидается, будут перегружаться через узел. Предполагается, что для компонент вектора x выполнены наборы линейных неравенств, включая систему неравенств $E^N(x - y) \leq 0$, где E^N – единичная матрица размера N . (Как и выше, $y \in R_+^N$ – вектор ожидаемых объемов грузов в регионе в течение определенного периода времени.)

Набор допустимых значений второго аргумента – который является вектором с компонентами, равными ценам на строительные работы, которые должны быть выполнены в рамках строительства транспортного узла — может быть описана системой линейных уравнений и неравенств вида $Sc \geq \gamma$, где $c \in R_+^p$, и S – $r \times p$ матрица с вещественными элементами, а $\gamma \in R^r$.

Цель проведенной работы состояла в том, чтобы показать, что с учетом приведенных выше рассуждений об объемах грузовых потоков, тарифах на транспортные услуги и ценах на строительные работы, можно сформулировать две задачи, связанные с оценкой целесообразности строительства грузового транспортного узла в конкретно выбранной географической точке региона. Первая задача заключается в том, чтобы определить, сможет ли такой транспортный узел какой-либо (технически) допустимой мощности генери-

ровать годовой доход в размере, позволяющем покрыть инвестиции, необходимые для строительства этого узла в течение разумного периода времени, конкурируя с другими транспортными услугами, предоставляемыми на рынке транспортных услуг при любых ожидаемых транспортных тарифах. Если это возможно, тогда решается вторая задача, в рамках которой первая задача должна рассматриваться при наиболее неблагоприятных транспортных тарифах, которые могут существовать на рынке транспортных услуг в течение этого периода времени. Оказывается, что определение того, может ли такой транспортный узел быть построен в условиях любых ожидаемых транспортных тарифов (т.е., первая из двух задач) сводится к решению вспомогательной задачи квадратичного программирования.

Если удастся показать, что транспортный узел определенного размера и структуры может генерировать прибыль, переговоры региональной администрации с частными инвесторами могут заинтересовать последних. Действительно, при наличии указанного инструментария, позволяющего многократно и быстро оценивать прибыльность варианта инвестирования, вопрос здесь заключается в том, чтобы отыскать наиболее выгодные условия инвестирования, включая сроки инвестирования и его объем, а также схему возврата капиталовложений (с процентами). Решение указанных выше двух задач как раз и позволяет оценить прибыль (как годовую, так и за несколько лет и даже декад), которую может генерировать транспортный узел, если будет построен по проекту, предлагаемому администрацией региона. Именно распределение этой прибыли между всеми участниками проекта (после того, как из решения этих двух задач будет установлено, что построенный по предлагаемому для финансирования проекту транспортный узел действительно может генерировать прибыль и можно оценить какую и за какой период времени) и должно стать предметом переговоров региональной администрации и частных инвесторов.

Показано, что математическая формулировка этих задач включает в себя систему ограничений для векторных аргументов, содержащую специальную

систему линейных неравенств, выполнение которых обеспечивает конкурентоспособность перегрузочных тарифов на рынке транспортных услуг. Утверждение о том, что совместность этой системы неравенств является достаточным условием конкурентоспособности тарифов на перевалку грузов в рассматриваемом транспортном узле легко проверяется простыми арифметическими вычислениями.

Помимо Введения, настоящий подраздел отчета содержит четыре пункта. В пункте 1.1.2 сформулировано вышеупомянутое достаточное условие конкурентоспособности перегрузочных тарифов в транспортном узле. Пункт 1.1.3 содержит утверждения и математические формулировки обеих сформулированных выше задач в содержательной постановке. Задача, связанная с оценкой целесообразности строительства грузового транспортного узла в географическом пункте региона формулируется как оптимизационная задача в робастной постановке, решение которой сводится к решению задачи квадратичного программирования. Задача отыскания наиболее неблагоприятных перегрузочных тарифов в транспортном узле математически формулируется как задача билинейного программирования, решение которой сводится к решению задач линейного программирования, образующих двойственную пару. Пункт 1.1.4 содержит краткий обзор проблем, близких к проблемам, рассматриваемым в настоящем отчете. В пункте 1.1.5 обсуждаются проблемы практического использования предложенного подхода к оценке целесообразности строительства грузового транспортного узла в географическом пункте региона, а также связь представленных в настоящем отчете результатов с результатами, полученными авторами ранее. В Заключение изложены наиболее насущные проблемы, связанные с проблемами, рассмотренными в настоящем отчете, и обсуждаются направления дальнейших исследований в изучаемой области.

1.1.2 Достаточные условия конкурентоспособности тарифов на услуги транспортного узла по перевалке грузов.

Рассмотрим числа $x_i, y_j, a_i, b_j \in R_+^1$, $i \in \overline{1, n}$, $j \in \overline{1, m}$, $n \leq m$.

Утверждение 1.1. Пусть выполняются неравенства

$$x_i \leq \sum_{j=1, j \neq i}^m y_j, a_i \leq \min_{j \in \overline{1, m} \setminus \{i\}} b_j, i \in \overline{1, n},$$

и равенство

$$\sum_{i=1}^n x_i = \sum_{i=1}^m y_i.$$

Тогда имеет место неравенство

$$\sum_{i=1}^n a_i x_i \leq \sum_{i=1}^m b_j y_j.$$

Доказательство содержится в [2].

Пусть теперь $\tilde{x}_i, \tilde{y}_j, \tilde{a}_i, \tilde{b}_j \in R_+^1$, $i \in \overline{1, I}$, $\tilde{x}_i = x_{k_i}^i$, $\tilde{y}_i = y_{j_i}^i$, $\tilde{a}_i = a_{k_i}^i$, $\tilde{b}_i = b_{j_i}^i$, $k_i \in \overline{1, n_i}$, $j_i \in \overline{1, m_i}$, and let $n_i \leq m_i, \forall i \in \overline{1, I}$.

Пусть далее:

$y_{j_i}^i$ является объемом груза $j_i \in \overline{1, m_i}$, который, как ожидается, будет предложен отправителем i для перевалки через транспортный узел $i \in \overline{1, I}$,

$x_{k_i}^i$ является объемом груза $k_i \in \overline{1, n_i}$, который, как ожидается, будет перемещен от отправителя i через транспортный узел посредством перевалки через транспортный узел $i \in \overline{1, I}$,

$b_{j_i}^i$ является рыночным тарифом на перевозку единицы объема груза $j_i \in \overline{1, m_i}$ от отправителя i в пункт назначения без использования услуг, предоставляемых транспортным узлом $i \in \overline{1, I}$,

$a_{k_i}^i$ является тарифом на перевалку единицы объема груза $k_i \in \overline{1, n_i}$ от отправителя i в пункт назначения через транспортный узел, $i \in \overline{1, I}$.

Тогда Утверждение 1.1 принимает следующую форму.

Утверждение 1.2. Пусть выполняются неравенства

$$a_{k_i}^i \leq \min_{v \in \overline{1, m_i} \setminus \{k_i\}} b_v^i, \quad \forall k_i \in \overline{1, n_i}, \quad (1.1)$$

$$x_{k_i}^i \leq \sum_{v=1, v \neq k_i}^{m_i} y_v^i, \quad \forall k_i \in \overline{1, n_i},$$

и равенство

$$\sum_{\mu=1}^{n_i} x_{\mu}^i = \sum_{v=1}^{m_i} y_v^i. \quad (1.2)$$

Тогда выполняется неравенство

$$\sum_{\mu=1}^{n_i} a_{\mu}^i x_{\mu}^i \leq \sum_{v=1}^{m_i} b_v^i y_v^i. \quad (1.3)$$

Система линейных неравенств (1.1) и (1.2) может, по существу, рассматриваться как достаточное условие конкурентоспособности перевалочных услуг, предоставляемых транспортным узлом по отношению к услугам, предлагаемым на рынке транспортных услуг. Здесь конкурентоспособность транспортного узла, которая описывается линейным неравенством (1.3), понимается по отношению к грузам, предложенным грузоотправителем $i \in I$ к перевалке в случае, когда им предложено не менее двух видов грузов.

Выполнение неравенств (1.1), (1.2) означает, что стоимость доставки грузов, предлагаемых к перевозке отправителем i путем их перевалки через транспортный узел, не превышает этой стоимости способами доставки, не включающими перевалку грузов через этот транспортный узел. Этот предлагаемый объем подразумевает все виды грузов, для которых транспортный узел предлагает свои услуги по их перевалке, т.е., $\sum_{\mu=1}^{n_i} x_{\mu}^i$ (который равен

$\sum_{v=1}^{m_i} y_v^i$). Легко убедиться, что эта особенность имеет место, даже если некоторые (но не все) рыночные тарифы на перевозку конкретных видов грузов из числа этих видов могут быть ниже, чем соответствующие тарифы на перевалку этих грузов, предлагаемые транспортным узлом.

Здесь первый набор неравенств, определяемый системой (1.1), означает, что стоимость перевалки для доставки единицы груза с номером k_i — которое входит в число грузов, предоставленных для транспортировки отправителем i — не превышает минимальные рыночные тарифные ставки для всех других грузов, предоставленных для перевозки отправителем i , т. е., грузов с номерами $\overline{1, m_i} \setminus \{k_i\}$.

Второй набор неравенств из (1.1) означает, что объем груза с любым номером k_i из $\overline{1, n_i}$, который перегружается через транспортный узел, не превышает общий объем всех других типов грузов, которые предоставляются отправителем i , т. е. грузов с номерами из набора $\overline{1, m_i} \setminus \{k_i\}$.

Ясно, что каждый отправитель i может быть заинтересован в использовании перевалочных услуг, предоставляемых транспортным узлом, если выполняются неравенства из первого набора неравенств из (1.1), и, вероятно, будет интересоваться этими услугами, если эти неравенства выполняются для чисел ωb_v^i , где $\omega < 1$.

В то же время следует иметь в виду, что взаимодействие администрации транспортного узла с отправителями идут следующим образом. Предлагая более дешевые тарифы на перевалку или более быструю доставку грузов по тарифам на перевалку, не превышающим, однако, рыночных или и то и другое, транспортные операторы, работающие с администрацией транспортного узла, могут выбирать такие объемы грузов от каждого отправителя i , которые сделают услуги по перевалке грузов через транспортный узел наиболее выгодными для этих операторов. Таким образом, тарифы по перевалке грузов через транспортный узел оказываются конкурентоспособными с рыночными тарифами для отправителя i только в объемах, определяемых неравенствами,

образующими второй набор неравенств из системы неравенств (1.1). Если I отправителей заинтересованы в перевалке своих грузов через транспортный узел, то выполнение I групп неравенств из системы неравенств (1.3) означает конкурентоспособность (в указанном смысле) услуг по перевалке грузов, предоставляемых транспортным узлом для всех этих I отправителей.

1.1.3 Постановки задач и их математические формулировки

Как упоминалось ранее, услуги по перевалке грузов, которые может предложить транспортный узел, могут стать конкурентоспособными с теми, которые доступны на рынке транспортных услуг для всех отправителей. Предполагая, что это требование конкурентоспособности удовлетворяется, при решении вопроса о том, следует ли строить транспортный узел в конкретном пункте географического региона, требуется команда управленцев транспортным узлом для определения оптимальной мощности этого узла и оценки уровня доходов, который может обеспечить функционирование транспортного узла этой оптимальной мощности.

Чтобы формализовать соответствующую задачу, будем считать, что:

$y^i \in R_+^{m_i}$ является вектором с компонентами $y_{j_i}^i, j_i \in \overline{1, m_i}$, равными объему груза, который, как ожидается, будет заявлен отправителем i на предмет предоставления ему перевалочных услуг через транспортный узел, $i \in \overline{1, I}$,
 $x^i \in R_+^{n_i}$ является вектором с компонентами $x_{k_i}^i, k_i \in \overline{1, n_i}$, равными частям объемов грузов, которые, как ожидается, будут перегружены через транспортный узел как часть объемов грузов, которые, как ожидается, будут предоставлены отправителем i для перевалки через транспортный узел, $i \in \overline{1, I}$,
 $b^i \in R_+^{m_i}$ является вектором с компонентами $b_{j_i}^i, j_i \in \overline{1, m_i}$, равными рыночным тарифам на транспортировку единицы объема груза $j_i \in \overline{1, m_i}$, который, как ожидается, будет заявлен отправителем i для перевозки в пункт

назначения, если, однако, такая транспортировка будет осуществлена без использования услуг, предоставляемых транспортным узлом, $i \in \overline{1, I}$,

$z^i \in R_+^{n_i}$ является вектором с компонентами $z_{k_i}^i, k_i \in \overline{1, n_i}$, равными тарифам на перевалку единичного объема груза $k_i \in \overline{1, n_i}$ до пункта назначения через транспортный узел который, как ожидается, будет заявлен отправителем i для перевозки в пункт назначения $i \in \overline{1, I}$,

$\Omega^i = \{y^i \in R_+^{m_i}: y^i D^i \leq d^i\}$ является множеством допустимых значений векторов $y^i, i \in \overline{1, I}$,

$M^i = \{z^i \in R_+^{n_i}: z^i P^i \leq \beta^i\}$ является множеством допустимых значений векторов $z^i, i \in \overline{1, I}$,

$\Lambda^i = \{b^i: b^i \leq \theta^i, i \in \overline{1, I}\}$ является множеством допустимых значений векторов $b^i, i \in \overline{1, I}$.

Пусть далее

$$\begin{aligned} Q^i(y^i) = \{x^i \in R_+^{n_i}: & \quad E^i x^i \leq T^i y^i, \quad \langle \epsilon^{n_i}, x^i \rangle - \langle \epsilon^{m_i}, y^i \rangle = 0, \\ & \epsilon^{n_i} \in R_+^{n_i}, \quad \epsilon^{m_i} \in R_+^{m_i}\}, \quad i \in \overline{1, I}, \end{aligned} \quad (1.4)$$

и

$$Q(y^1, \dots, y^I) = \bigcap_{i=1}^I Q^i(y^i) \cap \left\{ (x^1, \dots, x^I): \sum_{i=1}^I A^i x^i \leq q \right\}, \quad (1.5)$$

где:

T^i это $n_i \times m_i$ матрица с действительными элементами, каждый элемент которой в строке $k_i, k_i \in \overline{1, n_i}$ равен 1 кроме элемента в столбце k_i , равного 0, $i \in \overline{1, I}$,

$\epsilon^{n_i} = (1, 1, \dots, 1)$ это вектор, все компоненты которого равны 1, $i \in \overline{1, I}$,

$\epsilon^{m_i} = (1, 1, \dots, 1)$ это вектор, все компоненты которого равны 1, $i \in \overline{1, I}$,

E^i это $n_i \times n_i$ единичная матрица, $i \in \overline{1, I}$,

L это число всех видов грузов, которые, как ожидается, будут заявлены всеми I отправителями для перевалки через транспортный узел,

$q \in R^L$ это вектор с компонентами q_l , равными верхним границам объемов грузов с номерами l , которые, как ожидается, будут заявлены для перевалки через транспортный узел, $l \in \overline{1, L}$,

A^i это $L \times m_i$ матрица, соответствующая набору грузов, которые, как ожидается, будут заявлены для перевалки через транспортный узел отправителем $i, i \in \overline{1, I}$,

$w \in W \subset R_+^p$ это вектор, компонентами которого являются объемы строительных работ из списка всех p строительных работ, связанных со строительством транспортного узла.

Региональная администрация заинтересована в рассмотрении следующих задач:

Задача 1. Как выяснить, существуют ли размер и структура транспортного узла, работа которого позволила бы администрации полностью вернуть инвестиции, необходимые для строительства узла и его эксплуатации в течение ψ лет, при условии, что: а) транспортный узел может предлагать такие значения тарифов на перевалку, которые как ожидается, будут действовать на рынке транспортных услуг для перевозки соответствующие видов грузов в пункты назначения в течение следующих ψ лет, и б) стоимости всех требуемых в узле строительных работ совпадают со стоимостями всех строительных работ, контракты на выполнение которых в узле могут быть подписаны сегодня?

С учетом ранее введенных обозначений, задача 1 может быть математически сформулирована следующим образом

$$\max_{w \in W} \min_{c \in C} \left\{ \sum_{i=1}^I \langle z, E^{\psi} x \rangle - \langle c, w \rangle \right\} \rightarrow \max_{z \in Z, x \in X, w = Hx}. \quad (1.6)$$

Если значение этого максимума неотрицательно, то необходимо рассмотреть следующую задачу.

Задача 2. Как выяснить, существуют ли размер и структура транспортного узла, работа которого позволила бы администрации полностью вернуть инвестиции, необходимые для строительства узла и его эксплуатации в течение ψ лет, при условии, что узел сможет предложить значения тарифов на перевалку грузов (для перемещения этих грузов в соответствующие пункты назначения), которые меньше чем значения соответствующих рыночных транспортных тарифов, которые, как ожидается, будут в силе в течение следующих ψ лет в худшем случае значений этих тарифов для транспортного узла?

Прежде чем формулировать эту проблему математически, нужно определить: а) каким будет наихудший для транспортного узла сценарий по значениям тарифов на перевалку грузов через транспортный узел и б) как этот наихудший сценарий должен отражаться в математической постановке задачи 2?

Очевидно, что наихудшим сценарием тарифных значений для транспортного узла, который может иметь место на рынке транспортных услуг, будет тот, при котором прибыль транспортного узла будут достигать своего минимального значения, даже при условии, что тарифы на перевалку для соответствующих грузов (и их последующую доставку в пункты назначения) будут равными рыночным транспортным тарифам на перевозку этих грузов в пункты их назначения.

Пусть векторы рыночных транспортных тарифов на грузы, предложенные отправителем i , удовлетворяют системе линейных неравенств $B^i b^i \leq \theta^i, i \in \overline{1, I}$. Тарифы, соответствующие вышеупомянутому наихудшему сценарию тарифных значений могут быть найдены путем решения задачи

$$\min_{\prod_{i=1}^I (b^i \geq 0: B^i b^i \leq \theta^i)} \sum_{i=1}^I \langle b^i, x^i \rangle$$

$$\rightarrow \max_{(x^1, \dots, x^I; y^1, \dots, y^I): y^i \in \Omega^i; (x^1, \dots, x^I) \in Q(y^1, \dots, y^I), i \in \overline{1, I}} \quad (1.7)$$

которая является задачей нелинейного программирования, и эта проблема сводится к решению задач линейного программирования, образующих двойственную пару [10].

Пусть $b_i^*, i \in \overline{1, I}$ являются этими тарифами. С учетом введенных обозначений, Задача 2 может быть математически сформулирована следующим образом:

$$\begin{aligned} & \max_{w \in W} \min_{c \in C} \left(\sum_{i=1}^I \langle z^i, E^\psi x^i \rangle - \langle c, w \rangle \right) \\ & \rightarrow \max_{(z^1, \dots, z^I; x^1, \dots, x^I; y^1, \dots, y^I): z^i \in M^i \cap \{z^i \leq b_i^*(1-\eta)\}; y^i \in \Omega^i; (x^1, \dots, x^I) \in Q(y^1, \dots, y^I), i \in \overline{1, I}} \end{aligned} \quad (1.8)$$

где η – небольшое положительное число и $Z(\eta) = \prod_{i=1}^I (M^i \cap \{z^i \leq b_i^*(1-\eta)\})$.

Задачи (1.6) и (1.8) – нелинейные задачи оптимизации поиска максимума максимина функции, являющейся разностью между двумя квадратичными функциями. Первая из этих квадратичных функций является функцией двух векторных аргументов (z, x) , а вторая функция – это функция двух других векторных аргументов (c, w) . Два из этих четырех векторных аргументов связаны системой линейных ограничений $w = Hx$, и эта особенность позволяет свести эти задачи к задачам квадратичного программирования.

Для определенности рассмотрим задачу (1.6), которая является математической формулировкой задачи 1. Эта задача может быть записана как задача отыскания

$$\begin{aligned} & \max_{z \in Z} \max_{x \in X} \max_{w \in W} \min_{c \in C} \{ \langle z, E^\psi x \rangle - \langle c, w \rangle \} = \max_{z \in Z} \max_{x \in X} \min_{c \in C} \{ \langle z, E^\psi x \rangle - \langle c, Hx \rangle \} = \\ & = \max_{z \in Z} \max_{x \in X} \min_{c \in C} \{ \langle z, E^\psi x \rangle - \langle cH, x \rangle \} = \max_{z \in Z} \max_{x \in X} \min_{s \in S} \{ \langle z, E^\psi x \rangle - \langle s, x \rangle \}, \end{aligned}$$

где $s = cH$, $s \in S$ и S представляет собой выпуклый многогранник, являющийся образом выпуклого многогранника C при линейном операторе H .

Утверждение 1.3. Задача отыскания

$$\max_{z \in Z} \max_{x \in X} \min_{s \in S} \{ \langle z, E^\psi x \rangle - \langle s, x \rangle \}$$

сводится к задаче квадратичного программирования.

Доказательство этого утверждения приведено в [2].

1.1.4 Краткий обзор публикаций по развитию региональных транспортных узлов

Задачи расположения конкурентоспособных объектов активно изучались в последние десятилетия. Эта задача была впервые сформулирована в [11], где автор изучил проблему отыскания расположения двух конкурентоспособных участников рынка на конечном сегменте с равномерным распределением потребителей. Автор предполагал, что оба участника рынка принимают решения о размещении и предлагают цены на свои услуги одновременно. С момента публикации этой статьи, большое число работ по этой теме было выполнено, и многие работы были опубликованы.

В частности, [2] был, по-видимому, одним из первых, кто сформулировал задачу размещения конкурентоспособных объектов для сетей произвольного типа. Автор рассматривал рынок лишь с одним лидером и одним последователем, каждый из которых пытается максимизировать сумму весов тех узлов сети, для которых его объекты являются ближайшими. Автором были сформулированы две задачи. В первой задаче лидер решает, где разместить p своих объектов, полагая, что затем последователь будет решать, где разместить r своих объектов. Во второй задаче, последователь решает, где разместить r объектов, зная, что лидер уже разместил p объектов. Обе эти задачи оказались NP -трудными. Обзор статических задач о размещении конкурентоспособных объектов содержится в [13].

[14] сосредоточились на задаче о размещении конкурентоспособных объектов, в которых размещение объектов сначала производится лидером рынка, а затем последователем. Далее, в [15] эта задача была обобщена. Авторы рассмотрели последовательную пошаговую модель, где на каждом ша-

ге, каждый из игроков принимает решение о размещении только одного объекта и игра продолжается до тех пор, пока оба игрока не разместят все свои объекты. Авторы работы [16] сделали шаг в направлении усложнения постановки задачи и предположили, что число новых объектов последователя не определено заранее, а также предложили робастный метод решения оптимизационной задачи лидера.

Среди последних тем, рассматриваемых в задачах о размещении конкурентоспособных объектов, есть такие, которые изучают различные аспекты проектирования объекта, такие, как, например, стоимость соответствующих проектов по размещению таких объектов, ассортимент продукции, выпускаемой на каждом объекте или размер объекта, наряду с проблемами его расположения. Такие модели рассмотрены в [17, 18, 19, 20, 26]. В [17], целью работы является определение игроком расположения объекта и привлекательность каждого объекта, который будет функционировать в регионе исходя из уже функционирующих там объектов, принадлежащих конкуренту. При этом конкурент может реагировать на действия этого игрока, корректируя привлекательность своих объектов. Каждый игрок стремится максимизировать свою прибыль, которая зависит от его доли на рынке. Эта рыночная доля рассчитывается с помощью правила гравитации, и она (эта доля) пропорциональна привлекательности объекта и обратно пропорциональна расстоянию между объектом и клиентом (покупателем). Авторы предложили двухуровневую модель смешанного программирования. Использование условий оптимальности Каруша-Куна-Таккера позволило авторам [17] преобразовать сформулированную двухуровневую задачу смешанного математического программирования в эквивалентную одноуровневую задачу смешанного математического программирования. Они решили эту одноуровневую задачу для случаев с 30 клиентами и с несколькими (до пяти) транспортными узлами с использованием алгоритма GMIN-aBB, который реализует метод глобальной оптимизации, предложенный в [25].

В [26] автор рассмотрел проблему игрока-последователя, состоящую в выборе оптимального местоположения уже предопределенного числа магазинов и в нахождении такой подходящей цены для однородного продукта, продаваемого в конкурентной среде, которые максимизируют прибыль этого игрока. Здесь предполагается, что математическая формулировка этой задачи принимает во внимание местоположение уже существующих магазинов конкурента, который продает в них тот же самый продукт, а также цену на этот продукт. Тем не менее, в отличие от постановки задачи в [17], здесь не предполагается, что конкурент реагирует на действия участника путем изменения своей собственной цены на этот продукт. Другое предположение аналогично предположению [17] и исходит из того, что полезность клиента весьма зависит от цены (привлекательности) и стоимости поездки и описывает поток клиентов, отправляющихся в магазины, используя многомиллионную логит-модель (логистическую регрессию).

Для решения таким образом сформулированной задачи смешанного целочисленного нелинейного программирования, автор предложил разделить эту задачу на две подзадачи. Первая («разместительная») из этих подзадач связана с определением местоположения размещаемого объекта, а вторая («стоимостная») подзадача связана со стоимостными аспектами проблемы. Автор применил несколько эвристик для решения обеих подзадач в отдельности. Для части подзадачи, связанной с расположением объекта, автор применил жадный алгоритм, процедуру поиска табу и генетический алгоритм. Для решения второй подзадачи, связанной со стоимостным аспектом проекта, было предложено использовать метод градиентного спуска, метод градиентного спуска с несколькими случайными отправными точками и метод следования по пути. Эвристики были протестированы на сгенерированных задачах с 400 узлами спроса и несколькими (до 40) новыми магазинами.

Другие исследователи сосредоточиваются на играх по ценообразованию как части проблемы размещения. Например, [21] изучили игру, в которой лидер и последователи выбирают цены и местоположение одновременно. Ав-

торы доказали, что в последовательной игре последователь всегда зарабатывает больше, чем лидер. Аналогичные ценовые игры были изучены в [22, 23, 24].

Хотя конкуренция в проблеме размещения была детально изучена, конкуренция между уже функционирующим оператором и потенциально новым участником в сети с парами пункт отправления-пункт назначения изучена гораздо меньше. Возможно, [27] является первой статьей, посвященная этому типу конкуренции, в которой авторы предложили математическую формулировку проблемы только для последователя, когда существующая транспортная компания работает в сети, организованной по принципу "hub-and-spoke," а новая компания хочет зайти на этот рынок. В предложенной модели, новая компания получает 0,5, 0,75 или 1 потока грузов, проходящих через пары пункт отправления-пункт назначения, в зависимости от разности в тарифах, предлагаемых клиентам новой компанией и компанией, уже функционирующей на этом рынке. Авторы разработали эвристику для решения предлагаемой проблемы и протестировали модель на AP с 20-, 25-, 40- и 50-узловыми сетями. Хотя сама эвристика была сконструирована ими корректно, корректность ее реализации была оспорена в [28].

Проблема, аналогичная [27], изучалась в [29]. Там авторы сформулировали задачу только для последователя применительно к задаче выбора расположения аэропорта. Главное отличие их формулировки от формулировки на основе модели, предложенной в [27], состоит в способе моделирования предпочтений клиентов. Именно, здесь предпочтения постоянного клиента моделируются гравитационно-подобными функциями полезности, причем полезность (с точки зрения пользователя) зависит от некоторой базовой привлекательности, тарифов и времени в пути. Сформулированная нелинейная задача оптимизации была решена эвристическим алгоритмом, предложенным авторами, и он был протестирован на базе данных AP с 25 пунктами и несколькими (до 6) новыми транспортными узлами.

В отличие от авторов статьи [27], в статье [30], автор рассматривает случай, когда обе фирмы-игрока решают, где расположить транспортные узлы и какие пары пункт отправления-пункт назначения следует обслуживать. При этом автор предполагает, что лидер принимает решение первым и с этой целью лидер решает двухуровневую задачу, в которой на втором уровне им решается проблема последователя. Достижимая рыночная доля определяется на основе логит-функции с предопределенной бесполезностью, значения которой может быть истолковано как тарифы. В этой модели автор включает пороговые ограничения потока, которые используются для определения того, какие пары пункт отправления-пункт назначения обслуживать. Сформулированная автором задача была решена методом полного перебора, а также предложенной автором жадной эвристикой. Как и автор статьи [30], авторы статьи [31] изучают задачу, в которой две фирмы-участники рынка определяют, где им расположить транспортные узлы последовательно, а затем клиенты выбирают одну из фирм на основе сравнения наборов предлагаемых ими услуг или тарифов.

Фирма с лучшим уровнем обслуживания захватывает весь поток между определенной парой узлов. Авторы предлагают две модели и формулируют две проблемы на их основе. Для лидера сформулирована двухуровневая задача, а задача о максимальном покрытии формулируется для последователя, предполагая, что транспортные узлы лидера уже размещены. В обеих задачах каждая фирма стремится максимизировать свою долю на рынке. Доказано, что обе задачи являются *NP*-полными, поэтому авторы предложили эвристические алгоритмы для их решения. Модели и предложенные алгоритмы были проверены в рамках вычислительного эксперимента и в ходе анализа его результатов, проведенных на наборах данных CAB и TR.

В [32] авторы рассматривают задачу конкурентного размещения транспортного узла в условиях ценовой конкуренции. Эта задача описывает ситуацию, в которой новая компания: а) строит собственную "hub-and-spoke" сеть с возможными дополнительными прямыми связями между клиентами, и

б) устанавливает тарифы на свои услуги. Logit-функция используется для моделирования предпочтений клиентов, как в отношении существующей компании, так и в отношении новой компании, только лишь планирующей выйти на рынок транспортных услуг, на основе ее стратегии ценообразования. Рассматриваются две целевые функции новой компании – ее прибыль и ее доля рынка. Эти целевые функции подлежат максимизации, которая проводится а) путем выбора наилучшего расположения транспортных узлов и топологии сети, и б) путем применения оптимальных тарифов. Авторы сформулировали нелинейную задачу смешанного математического программирования, которую они решили с помощью генетического алгоритма. Сообщается об испытаниях и вычислительных экспериментах, основанных на использовании данных из набора данных CAB с 25 узлами.

1.1.5 Обсуждение

1. Хотя предложенное достаточное условие и задачи 1 и 2 сформулированы в рамках оценки того, следует ли строить новый центр грузовых перевозок или модернизировать уже существующий в конкретном географическом регионе, другое применение этого условия кажется очевидным. Именно, рассмотрим уже сформировавшийся и в настоящее время функционирующий региональный рынок транспортных услуг. В рамках имеющей там место конкуренции за клиентов, компания, управляющая грузовым транспортным узлом, может оценить экономическую эффективность предоставления новых услуг по перевалке грузов для набора клиентов.

Как и в пункте 1.1.3, пусть I отправителей в настоящее время перемещают свои грузы по территории региона — между пунктами в регионе, между пунктами в других регионах и пунктами в регионе и транзитом через территорию региона — с использованием одного из конкретных видов транспорта, скажем, грузовиков или водных путей. Предположим, что они (отправители) могут перемещать свои грузы по тарифам, существующим на рынке транспортных услуг для каждого вида транспорта.

Как указывалось ранее, управляющая компания грузового транспортного узла может переключить часть грузового потока, обслуживаемого в настоящее время другими транспортными компаниями, если она может предложить более быструю доставку и более дешевые тарифы для некоторых грузов, по крайней мере, некоторым из I отправителей и при этом оставаться прибыльной. Таким образом, возникают два естественных вопроса.

Вопрос 1. Существуют ли такие конкурентоспособные тарифы при известной стоимости предоставления этих услуг по перевалке грузов?

Вопрос 2. Предполагая, что конкурентоспособные тарифы, упомянутые в Вопросе 1, существуют, как найти наиболее прибыльные для транспортного узла тарифы?

Пусть $s^i \in R_+^{n_i}$ – вектор с компонентами $s_{k_i}^i, k_i \in \overline{1, n_i}$ равными стоимости предоставления перевалочных услуг в транспортном узле за единицу объема груза, который будет отправлен через узел из набора грузов, которые, как ожидается, будут заявлены отправителем i для перевалки через транспортный узел, $i \in \overline{1, I}$.

С использованием обозначений, введенных в пункте 1.1.3, можно переформулировать Задачу 1 и Задачу 2 следующим образом.

Задача 1'. Как выяснить, существуют ли значения тарифов на перевалку грузов для перемещения соответствующих видов грузов в пункты назначения, которые будут выгодны для транспортного узла при затратах, связанных с предоставлением услуг грузоотправителям по перевалке грузов через узел I .

Принимая во внимание уже введенные обозначения, задача 1 может математически формулироваться следующим образом:

$$\sum_{i=1}^I \{ \langle z^i, x^i \rangle - \langle s, x^i \rangle \} \rightarrow \max_{(z^1, \dots, z^I; x^1, \dots, x^I; y^1, \dots, y^I): z^i \in M^i; y^i \in \Omega^i; (x^1, \dots, x^I) \in Q(y^1, \dots, y^I), i \in \overline{1, I}}.$$

Если значение максимина неотрицательно, то следует решить Задачу 2'.

Задача 2'. Как выяснить, существуют ли значения тарифов на перевалку грузов для перемещения соответствующих видов грузов в пункты назначения, которые будут выгодны для транспортного узла по стоимости предоставления услуг перевалки для I отправителей при условии, что узел может предложить значения тарифов на перегрузку (для перемещения соответствующих видов грузов в их пункты назначения), которые меньше, чем значения тарифов на соответствующем рынке транспортных услуг при наихудшем сценарии (для транспортного узла), который может сложиться в ближайшем будущем.

Как и ранее, (см. пункт 1.1.3), наихудший сценарий значений тарифов может быть найден из решения задачи

$$\min_{\Pi_{i=1}^I \{b^i \geq 0: B^i b^i \leq \theta^i\}} \sum_{i=1}^I \langle b^i, x^i \rangle \rightarrow \max_{(x^1, \dots, x^I; y^1, \dots, y^I): y^i \in \Omega^i; (x^1, \dots, x^I) \in Q(y^1, \dots, y^I), i \in \overline{1, I}}.$$

Пусть $b_i^*, i \in \overline{1, I}$ являются тарифами, соответствующими этому наихудшему сценарию. С учетом введенных обозначений Задача 2' может быть математически сформулирована следующим образом

$$\sum_{i=1}^I \{ \langle z^i, x^i \rangle - \langle c, w \rangle \} \rightarrow \max_{(z^1, \dots, z^I; x^1, \dots, x^I; y^1, \dots, y^I): z^i \in M^i \cap \{z^i \leq b_i^* (1-\eta)\}; y^i \in \Omega^i; (x^1, \dots, x^I) \in Q(y^1, \dots, y^I), i \in \overline{1, I}},$$

где, как и прежде, η – небольшое положительное число.

Пример формулировки и решения задач 1 и 2 на модельных данных приведен в [2].

1.1.6 Заключение по подразделу

В этом подразделе отчета приведены некоторые результаты решения проблемы построения работоспособного инструментария для анализа выгодности для частных инвесторов капиталовложений (инвестиций) в проекты,

связанные с развитием транспортных инфраструктур. Предложена формулировка задач а) об оценке окупаемости требуемых инвестиций в строительство транспортного узла в какой-либо конкретной точке региона при любых ожидаемых тарифах на транспортные услуги, которые, могут предложить уже работающие в регионе транспортные структуры в течение последующих нескольких лет и (если эта задача разрешима) б) об оценке срока окупаемости этих инвестиций при наиболее неблагоприятном сценарии формирования рыночных транспортных тарифов на перевозки грузов. При этом в формулировке второй задачи предполагается, что транспортный узел сможет предлагать тарифы, которые, ниже, чем рыночные, даже при указанном наиболее неблагоприятном сценарии. Кроме того, в обеих задачах предполагается, что объемы строительных работ в этом узле линейно зависят от ожидаемых объемов грузов, которые могут пройти через него в течение какого-либо промежутка времени длиной в несколько лет.

Показано, что обе задачи, которые формулируются как задачи отыскания максимума функции максимина, сводятся к задаче квадратичного программирования, в то время как отыскание наиболее неблагоприятного сценария формирования рыночных транспортных тарифов на перевозки грузов сводится к решению задач линейного программирования, образующих двойственную пару. Таким образом, программную реализацию части инструментария, связанную с решением обеих указанных задач удастся осуществить конечными методами оптимизации, для которых имеются стандартные пакеты прикладных программ. Эта особенность предложенных в работе математических моделей позволяет решать задачи большого размера, какими и являются обе сформулированные задачи на практике, быстро и потому многократно, что очень важно как на практике подготовки к переговорам всех участвующих сторон, так и в ходе самих переговоров.

Также сформулированы и доказаны проверяемые достаточные условия конкурентоспособности тарифов на услуги транспортного узла в условиях рынка, которые использованы при формулировке систем ограничений в обе-

их задачах. Хотя обе вышеуказанные задачи сформулированы и решены при определенных упрощающих предположениях, результаты их решения могут быть использованы администрациями регионов при проведении ими переговоров с частными инвесторами о капиталовложениях в развитие транспортной инфраструктуры региона.

Полученные результаты, тем не менее, следует рассматривать лишь как первый шаг в направлении создания работоспособного инструментария для анализа выгодности для частных инвесторов капиталовложений в проекты, связанные с развитием транспортных инфраструктур. Необходимо установить характер нелинейностей в зависимостях объема строительных работ в транспортном узле от ожидаемого объема проходящих через него грузов и сформулировать задачи, аналогичные рассмотренным в настоящей работе, с учетом нелинейных зависимостей, которые будут установлены. При этом полученные в работе результаты в случае указанной линейной зависимости могут служить ориентиром при анализе нелинейного случая и его особенностей.

1.2 Исследование задач теории расписания в турнирах

1.2.1 Введение

Существует два основных вида спортивных турниров: турниры с круговой системой, и турниры на выбывание (олимпийская система). В турнирах на выбывание все проигравшие выбывают из турнира, все победители проходят в следующий раунд. В турнирах с круговой системой каждый участник играет со всеми оставшимися участниками. Основным преимуществом турниров на выбывание является существенно меньшее количество матчей и раундов и, как следствие, увеличивается зрительский интерес от раунда к раунду. Например, если число участников равно $N = 2^n$, тогда число игр (с двумя участниками в одном матче) в турнире на выбывание равно $2^n - 1$, в то время как турнире по круговой системе оно равно $2^{n-1}(2^n - 1)$. Соответствен-

но, число раундов в турнире на выбывание равно n , а в круговой системе равно $2^n - 1$. Более низкие требования к времени проведения, спортивному оборудованию, и возрастающий зрительский интерес являются главными причинами популярности турниров на выбывание.

Существует множество игр с одним победителем, и числом игроков одного матча больше двух (например, карточные игры покер и преферанс). Футбольные команды обычно делятся на группы по четыре команды в группе. Круговой подтурнир внутри группы может рассматриваться как матч с четырьмя командами.

Беговые дорожки, бассейны, дорожки для боулинга и другое спортивное оборудование имеют ограниченные мощности. Таким образом, невозможно организовать один забег (заплыв и т.д.) для всех спортсменов одновременно. Обычно организуют квалификационный раунд, полуфинал, финал. Для большого числа участников применяется система турнира на выбывание. В данной работе позиция (номер) дорожки не имеет значения, только множество участников одного матча играет важную роль. Реальные спортивные турниры имеют свои собственные дополнительные правила, например, влияют не только относительные результаты, но абсолютные значения, но в данной работе мы развиваем общую теорию спортивных турниров, которая может быть применена для всех соревнований.

В данном подразделе отчета представлены результаты, которые обобщают модель расписаний спортивных турниров, разработанную для случая двух участников в одном матче [33], через рассмотрение матчей с числом участников больше двух. Существует большое количество вариантов построения расписаний турниров на выбывание. Расписаниями мы будем называть распределение участников по матчам в каждом туре соревнования. Участники изначально упорядочены по силе (эта информация обычно доступна из результатов предыдущих соревнований). В вычислительной теории коллективного выбора турнир на выбывание соответствует дереву голосования или повестке голосования [34].

Существует несколько подходов к теоретическому обоснованию расписания турнира, но результаты неоднозначны. Соответствующая экономическая литература рассматривает издержки и выгоды различного дизайна турнира (расписаний) для неоднородных участников [35, 36, 37, 38, 39]. В экономическом анализе имеют значение структура призов, функции усилий и другие предпосылки. Строго обоснованные решения есть только для турниров с четырьмя участниками.

Подход комбинаторной оптимизации разработан в [40, 33]. В этих работах предложено обоснование трех различных расписаний, основными из которых являются стандартное расписание и расписание равных преимуществ.

Данная работа развивает аксиоматический подход к построению расписаний турнира для числа участников в матче больше двух. Разработаны несколько новых свойств расписаний и найдено количество расписаний, им удовлетворяющих. Предложены обобщения стандартного расписания и расписания равных преимуществ. Организаторы спортивных турниров могут применить эти расписания для любых соревнований.

Работа содержит комбинаторные результаты и теоремы о представлении. Так как расписания это новый комбинаторный объект, то все формулы являются новыми. Теоремы о представлении являются основным результатом работы. Они основаны и на комбинаторных результатах.

Структура этого подраздела следующая. В пункте 1.2.2 описаны расписания обобщенных турниров, их свойства и комбинаторные результаты. Пункт 1.2.3 содержит теоремы о представлении. В пункте 1.2.4 сделаны некоторые выводы по работе.

1.2.2 Описание модели расписаний

Пусть k будет являться числом участников в одном матче, n будет являться числом раундов и $X = \{1, 2, \dots, k^n\}$ – множеством участников турнира на выбывание (далее турнира). Индексы участников отражают силу участника, где участник 1 является сильнейшим, а участник k^n слабейшим.

Расписание турнира (далее расписание) это гиперграф с k^n вершинами проименованными с 1 до k^n , который описывается следующей системой множеств (вложенной системой множеств). k^{n-1} непересекающихся множеств с k вершинами (каждое множество это отдельный матч), k^{n-2} непересекающихся множеств с k^2 вершинами (каждое такое множество это подтурнир с двумя раундами), k^{n-3} непересекающихся множеств с k^3 вершинами (каждое такое множество это подтурнир с тремя раундами), и т.д.

Например, расписание турнира с $k = 2$ участниками в каждом матче и $n = 3$ раундами описывается следующей системой множеств $\{1,4\}, \{2,3\}, \{5,8\}, \{6,7\}, \{1,2,3,4\}, \{5,6,7,8\}, \{1,2,3,4,5,6,7,8\}$, но более удобно представлять расписание как систему вложенных множеств $\{\{\{1,4\}, \{2,3\}\}, \{\{5,8\}, \{6,7\}\}\}$. В данном турнире два подтурнира $\{\{1,4\}, \{2,3\}\}$ и $\{\{5,8\}, \{6,7\}\}$, каждый из которых также содержит два подтурнира. В каждом последующей раунде встречаются победители предыдущего. Порядок множеств внутри подтурнира не имеет значения. $\{\{\{1,4\}, \{2,3\}\}, \{\{5,8\}, \{6,7\}\}\}$ и $\{\{\{1,4\}, \{3,2\}\}, \{\{6,7\}, \{8,5\}\}\}$ отражают одно и то же расписание.

Для $k = 2$, наиболее популярным расписанием является стандартное расписание, которое самого сильного участника ставит в пару с самым слабым, второго – с вторым с конца и т.д. Матчи в последующих раундах определяются таким образом, чтобы не допустить матча между двумя сильнейшими участниками до финала, между четырьмя сильнейшими участниками о полуфинала и т.д. Сильные участники выигрывают от применения стандартного расписания. Для $k = 2$, $n = 3$, мы имеем $T_{2,3}^{standard} = \{\{\{1,8\}, \{4,5\}\}, \{\{2,7\}, \{3,6\}\}\}$.

Каждый турнир с $n \geq 2$ раундами содержит k подтурниров. Каждый подтурнир с $n \geq 2$ раундами также содержит k подтурниров. $T_{k,n}^{m,i}$ это подтурнир i с m раундами. Этот подтурнир принадлежит турниру с n раундами и k участниками в каждом матче. Для удобства обозначений примем $T_{k,n}^n =$

$T_{k,n}$ и $T_{k,n}^{0,i} = \{i\}$. Подтурниры $T_{k,n}^{m,i}, T_{k,n}^{m,j}$ являются непересекающимися если нет ни одного участника принадлежащего двум подтурнирам. Турнир с n раундами это множество $T_{k,n} = \bigcup_{i=1}^k T_{k,n}^{n-1,i}$, где все подтурниры являются непересекающимися.

Пусть $\mathbb{T}_{k,n}$ будет множеством всех возможных расписаний с k участниками в одном матче и n раундами. Мощность этого множества обозначается как $\#\mathbb{T}_{k,n}$. Пусть $\mathbb{T}_{k,n}^{property}$ будет множеством всех расписаний с k участниками в одном матче и n раундами, которые удовлетворяют данному свойству; мощность этого множества обозначается как $\#\mathbb{T}_{k,n}^{property}$.

Утверждение 1.4. Количество расписаний равно

$$\#\mathbb{T}_{k,n} = (k!)^{\frac{1-k^n}{k-1}} (k^n!).$$

Доказательство утверждения 1.4 и последующих утверждений представлены в [41]. Рассматривая турнир как объединение k подтурниров, мы получим рекурсивное представление числа расписаний:

$$\#\mathbb{T}_{k,n} = \frac{k^n!}{k!(k^{n-1}!)^k} (\#\mathbb{T}_{k,n-1})^k.$$

Все комбинаторные формулы являются новыми. Некоторые из них для случая $k = 2$ или $n = 2$ добавлены в онлайн энциклопедию целочисленных последовательностей (the On-line Encyclopaedia of Integer Sequences, далее OEIS [42]).

Для $k = 2$, $\#\mathbb{T}_{k,n}$ это последовательность A067667 из OEIS. Для $n = 2$, $\#\mathbb{T}_{k,n}$ это последовательность A057599 из OEIS. Даже для небольших значений k , n число расписаний быстро растет, например, $\#\mathbb{T}_{3,3} = 833'712'928'048'000'000$.

Расписание турнира является чисто комбинаторным объектом, без каких-либо предположений о поведении участников. Для исследования свойств расписаний мы предполагаем, что более сильный участник всегда побеждает более слабого участника.

Первое свойство делает расписание инвариантным относительно обращения изначального упорядочения участников. Нет специальных правил для сильных и слабых участников. Расписание, построенное для участников, упорядоченных по силе, тождественно расписанию, построенному для участников, упорядоченных по слабости.

Симметричность. Расписание инвариантно относительно отображения $i \rightarrow k^n + 1 - i$.

$\{\{1,4\}, \{2,3\}\}, \{\{5,8\}, \{6,7\}\}$ и $\{\{1,8\}, \{2,7\}\}, \{\{3,4\}, \{5,6\}\}$ являются примерами симметричных расписаний.

Утверждение 1.5. Количество расписаний, удовлетворяющих условию симметричности, равно для нечетных k

$$\#T_{k,n}^S = \frac{2^{\frac{k^n - nk + n - 1}{2}} \left(\frac{k^n - 1}{2}\right)!}{(k!)^{\frac{1}{2} \left(\frac{k^n - 1}{k - 1} - n\right)} \left[\left(\frac{k - 1}{2}\right)!\right]^n},$$

для четных k

$$\#T_{k,n}^S = \left(\frac{k^n}{2}\right)! \sum_{i=0}^{k/2} \frac{(\#T_{k,n-1}^S)^{2i}}{\left(\left(\frac{k^{n-1}}{2}\right)!\right)^{2i} i! \left(\frac{k-i}{2}\right)!} \left(\frac{2^{k^{n-1}-1}}{(k!)^{\frac{k^{n-1}-1}{k-1}}}\right)^{\frac{k}{2}-i}.$$

Для $k = 2$, $\#T_{k,n}^S$ это последовательность A261187 в (OEIS). Для этого случая существует рекурсивное представление:

$$\#T_{2,n}^S = 2^{n-1}! y_n,$$

где $y_n = 0.5(y_{n-1})^2 + 1$, при $y_1 = 1$.

Для нечетных k , существует другая рекурсия. Для каждого расписания, существует только один матч симметричный самому себе и $\frac{k^{n-1}-1}{2}$ симметричных пар матчей. Всего существует 2^{k-1} способов определить симметричную пару матчей. Рассматривая турнир как объединение k^{n-1} матчей, мы получим

$$\#T_{k,n}^S = \frac{\left(\frac{k^{n-1}}{2}\right)!}{\left(\frac{k-1}{2}\right)!(k!)^{\frac{k^{n-1}-1}{2}}\left(\frac{k^{n-1}-1}{2}\right)!} (2^{k-1})^{\frac{k^{n-1}-1}{2}} \#T_{k,n-1}^S.$$

Два представления расписания: как объединение k подтурниров и как объединение k^{n-1} матчей применяются для вывода последующих комбинаторных формул.

Следуя Райту [43] интенсивность конкуренции является ключевым свойством для дизайна спортивных турниров. Чем ближе сила участников, тем выше интенсивность конкуренции. В подходе с числом участников больше двух два сильнейших игрока одного матча являются главными соперниками. От раунда к раунду два сильнейших соперника должны быть ближе по уровню сил, что должно повышать зрительский интерес. В финальном матче два сильнейших соперника сыграют против друг друга.

Возрастающая интенсивность конкуренции. В каждом последующем раунде победитель встречается с более сильным соперником, чем в предыдущем раунде.

Утверждение 1.6. Количество расписаний, удовлетворяющих условию возрастающей интенсивности конкуренции, равно

$$\#T_{k,n}^{ICI} = ((k-2)!)^{\frac{1-k^{n-1}}{k-1}} \prod_{i=2}^n \left[\frac{(k^i-2)!}{((k^{i-1}-1)!)^2 ((k^{i-1})!)^{k-2}} \right]^{k^{n-i}}.$$

Для $k = 2$, $\#T_{k,n}^{ICI}$ это также число двоичных куч (последовательность A056972 в (OEIS)). Возрастающая интенсивность конкуренции является слабым ограничением, с $\lim_{k \rightarrow \infty} \frac{\#T_{k,2}^{ICI}}{\#T_{k,2}} = 1$ и $\lim_{k \rightarrow \infty} \frac{\#T_{k,3}^{ICI}}{\#T_{k,3}} = e^{-1}$. Следующее свойство усиливает свойство возрастающей интенсивности конкуренции, гарантируя сильнейший финал, сильнейший полуфинал и т.д.

Отложенная конфронтация [44]. Сильнейшие k^j участников не должны встретиться до тех пор пока число участников не уменьшится до k^j или меньше.

Это ключевое свойство для дизайна турнира. Данное свойство поддерживает зрительский интерес. Матчи с и между сильнейшими участниками имеют повышенный спрос. Сильнейшие участники по возможности не должны исключаться в начале турнира. Данное свойство распределяет сильнейших участников равномерно между подтурнирами. Таким образом, нет подтурниров с только сильными игроками или только со слабыми игроками.

Утверждение 1.7. Количество расписаний, удовлетворяющих условию отложенной конфронтации, равно

$$\#T_{k,n}^{DC} = ((k-1)!)^{\frac{k-k^n}{k-1}} \prod_{i=2}^n (k^i - k^{i-1})!.$$

Для $k = 2$, $\#T_{k,n}^{DC}$ это последовательность A261125 в (OEIS). Свойство отложенной конфронтации не требует предпосылки о детерминистическом исходе каждого матча. Сильные участники не играют друг против друга при любом исходе матчей. Далее мы введем несколько свойств, базирующихся на свойстве отложенной конфронтации: преимущество сильного игрока, равные преимущества, равные суммы, баланс, равное разделение проигравших.

Свойство преимущество сильного игрока введено Швенком [44]. Мы даем преимущество сильным игрокам, чтобы они не имели стимулов проиграть в предтурнирных матчах с целью показаться более слабым игроком и получить более слабого соперника. Модель, описывающая данные стимулы построена в [45].

Преимущество сильного игрока. При выполнении свойства отложенной конфронтации, в каждом раунде r , абсолютная разница рангов сильнейшего участника матча и его сильнейшего соперника по матчу строго возрастает по силе сильнейшего участника.

Стандартное расписание удовлетворяет этому свойству. В этом расписании сильнейший участник играет против слабейшего участника, гарантируя себе наибольшее преимущество.

Свойство равных преимуществ отражает идею минимального фаворитизма из [44]. Мы обобщили меру интенсивности конкуренции, предложенную в [40] для числа участников матча больше двух. Интенсивность конкуренции это абсолютная разность между рангами сильнейшего и второго сильнейшего участника матча. Уравнивая интенсивности конкуренции всех матчей одного раунда, мы получим свойство равных преимуществ.

Равные преимущества. При выполнении свойства отложенной конфронтации, все матчи одного раунда должны иметь равную абсолютную разность между рангами сильнейшего и второго сильнейшего участника.

Утверждение 1.8. Количество расписаний, удовлетворяющих условию равных преимуществ, равно

$$\#T_{k,n}^{ED} = ((k-2)!)^{\frac{k-k^n}{k-1}} \prod_{i=2}^n (k^i - 2k^{i-1})!.$$

Следующее свойство уравнивает качество матчей [40] и поддерживает интерес ко всем матчам. Под качеством матча понимается сумма рангов участников.

Равные суммы. При выполнении свойства отложенной конфронтации, все матчи одного раунда должны иметь равную сумму рангов участников.

Следующее свойство упрощает свойство симметричности при выполнении свойства отложенной конфронтации.

Баланс. При выполнении свойства отложенной конфронтации, все матчи одного раунда должны быть инвариантны относительно отображения $i \rightarrow k^{n-r+1} + 1 - i$, где r это номер раунда.

Утверждение 1.9. Количество расписаний, удовлетворяющих условию баланса, равно для нечетных k

$$\#\mathbb{T}_{k,n}^B = 0,$$

для четных k

$$\#\mathbb{T}_{k,n}^B = \left(\left(\frac{k}{2} - 1\right)!\right)^{\frac{k-k^n}{k-1}} \prod_{i=2}^n \left(\frac{k^i - 2k^{i-1}}{2}\right)!.$$

Следующее свойство уравнивает матчи по присутствию слабейших игроков. Таким образом, мы исключим преимущество от игр со слабыми соперниками.

Равное разделение проигрывающих При выполнении свойства отложенной конфронтации, во всех матчах одного раунда должен быть один участник из множества $\{k^{n-r+1} - k^{n-r} + 1, \dots, k^{n-r+1}\}$, где r это номер раунда.

Утверждение 1.10. Количество расписаний, удовлетворяющих условию равного разделения проигрывающих, равно

$$\#\mathbb{T}_{k,n}^{EPL} = ((k-2)!)^{\frac{k-k^n}{k-1}} \prod_{i=2}^n k^{i-1}! (k^i - 2k^{i-1})!.$$

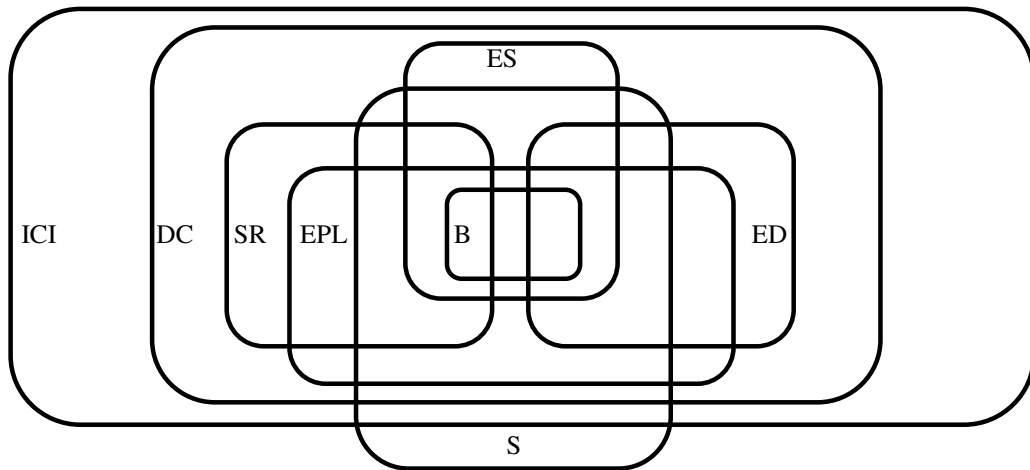


Рисунок 1.1 – Диаграмма Венна для расписаний турниров на выбывание: ICI – возрастающая интенсивность конкуренции; DC – отложенная конфронтация; SR – преимущество сильного игрока; EPL – равное разделение проигрывающих; B – баланс; ED – равные преимущества; ES – равные суммы; S – симметричность; все метки на-

ходятся внутри соответствующих областей и поставлены близко к границе

Рисунок 1.1 представляет диаграмму Венна для свойств расписаний. Если область X принадлежит области Y , тогда все расписания, удовлетворяющие свойству X также удовлетворяют свойству Y . Пустое пересечение означает невозможность совмещения двух свойств. Диаграмма описывает общий случай. Для конкретных параметров n и k , диаграмма может быть проще, некоторые пересечения могут быть пустыми.

Количество расписаний, удовлетворяющих каждому свойству достаточно велико, но мощность пересечение значительно меньше. На этом основываются теоремы о представлении в следующем подразделе.

1.2.3 Теоремы о представлении

В данном подразделе мы построим обобщение стандартного расписания и расписания равных преимуществ из [33].

1.2.3.1 Стандартное расписание

Для $k = 2$, наиболее популярным расписанием является стандартное расписание. Оно определяется рекурсивно. Для любого m от 1 до n , имеем

$$T_{2,n}^{m,i} = \{T_{2,n}^{m-1,i}, T_{2,n}^{m-1,2^{n-m+1}-i+1}\}, i = \overline{1, 2^{n-m}}.$$

Таким образом, для $n = 3$, получим

$$T_{2,3}^{standard} = \{\{\{1,8\}, \{4,5\}\}, \{\{2,7\}, \{3,6\}\}\}.$$

Существует несколько обоснований стандартного расписания.

Утверждение 1.11. [33] Для $k = 2$ стандартное расписание является единственным расписанием, которое удовлетворяет свойству равных сумм.

Утверждение 1.12. Для $k = 2$ стандартное расписание является единственным расписанием, которое удовлетворяет свойству преимущество сильного игрока.

Утверждение 1.13. Для $k = 2$ стандартное расписание является единственным расписанием, которое удовлетворяет свойству баланса.

Стандартное расписание также удовлетворяет свойству равного разделения проигрывающих. Не существует явного обобщения стандартного расписания на случай произвольного k . Для $k = 3$ и $n = 2$, $\{\{1,6,8\}, \{2,4,9\}, \{3,5,7\}\}$ и $\{\{1,5,9\}, \{2,6,7\}, \{3,4,8\}\}$ удовлетворяют свойствам симметричности и равных сумм, $\{\{1,6,7\}, \{2,5,8\}, \{3,4,9\}\}$ удовлетворяют свойствам симметричности и преимуществу сильного игрока, но не свойству равных сумм. В работе разработаны два расписания для $k = 3$ и $k = 4$, которые удовлетворяют ключевым свойствам стандартного расписания и представлено соответствующее аксиоматическое обоснование.

Для $k = 3$, модифицированное стандартное расписание определяется рекурсивно. Для любого m от 1 до n , имеем

$$T_{3,n}^{m,i} = \{T_{3,n}^{m-1,i}, T_{3,n}^{m-1, 2 \cdot 3^{n-m} - i + 1}, T_{3,n}^{m-1, 2 \cdot 3^{n-m} + i}\}, i = \overline{1, 3^{n-m}}.$$

Таким образом, для $k = 3$ и $n = 3$ получим

$$T_{3,3}^{MS} = \left\{ \begin{array}{l} \{\{1,18,19\}, \{6,13,24\}, \{7,12,25\}\}, \{\{2,17,20\}, \{5,14,23\}, \{8,11,26\}\}, \\ \{\{3,16,21\}, \{4,15,22\}, \{9,10,27\}\} \end{array} \right\}.$$

Утверждение 1.14. Для $k = 3$ модифицированное стандартное расписание является единственным расписанием, которое удовлетворяет свойствам преимущества сильного игрока и симметричности.

Для $k = 4$, модифицированное стандартное расписание определяется рекурсивно. Для любого m от 1 до n , имеем

$$T_{4,n}^{m,i} = \left\{ T_{4,n}^{m-1,i}, T_{4,n}^{m-1, \frac{4^{n-m+1}}{2} - i + 1}, T_{4,n}^{m-1, \frac{4^{n-m+1}}{2} + i}, T_{4,n}^{m-1, 4^{n-m+1} - i + 1} \right\}, i = \overline{1, 4^{n-m}}.$$

Таким образом, для $k = 4$ и $n = 2$ получим

$$T_{4,2}^{MS} = \{\{1,8,9,16\}, \{2,7,10,15\}, \{3,6,11,14\}, \{4,5,12,13\}\}.$$

Утверждение 1.15. Для $k = 4$, модифицированное стандартное расписание является единственным расписанием, которое удовлетворяет свойствам преимущество сильного игрока и баланса.

Даже для большего k не существует расписания, удовлетворяющего всем свойствам стандартного расписания.

Утверждение 1.16. Для $k = 5$ не существует расписания, удовлетворяющего свойствам равных сумм, симметричности и преимущество сильного игрока.

1.2.3.2 Расписание равных преимуществ

Для $k = 2$ расписание равных преимуществ разработано в [33]. Приведем ниже обобщение данного расписания. Расписание равных преимуществ определяется рекурсивно. Для любого m от 1 до n , имеем

$$T_{k,n}^{m,i} = \bigcup_{j=0}^{k-1} T_{k,n}^{1,i+jk^{n-m}}, i = \overline{1, k^{n-m}}.$$

Таким образом, для $k = 2$ и $n = 4$, получим

$$T_{2,3}^{EG} = \left\{ \left\{ \{1,9\}, \{5,13\} \right\}, \left\{ \{3,12\}, \{7,15\} \right\} \right\}, \left\{ \left\{ \{2,10\}, \{6,14\} \right\}, \left\{ \{4,13\}, \{8,16\} \right\} \right\};$$

для $k = 3$ и $n = 3$, получим

$$T_{3,3}^{EG} = \left\{ \left\{ \{1,10,19\}, \{4,13,22\}, \{7,16,25\} \right\}, \left\{ \{2,11,20\}, \{5,14,23\}, \{8,17,26\} \right\}, \right. \\ \left. \left\{ \{3,12,21\}, \{6,15,24\}, \{9,18,27\} \right\} \right\};$$

для $k = 4$ и $n = 2$, получим

$$T_{4,2}^{EG} = \{ \{1,5,9,13\}, \{2,6,10,14\}, \{3,7,11,15\}, \{4,8,12,16\} \}.$$

Ниже приведены аксиоматические обоснования расписания равных преимуществ.

Утверждение 1.17 [33]. Для $k = 2$ расписание равных преимуществ является единственным расписанием, которое удовлетворяет свойству равных преимуществ.

Для $k = 2$ расписание равных преимуществ также удовлетворяет свойству симметричности.

Утверждение 1.18. Для $k = 3$, расписание равных преимуществ является единственным расписанием, которое удовлетворяет свойствам равных преимуществ и симметричности.

Модифицированное расписание равных преимуществ определяется рекурсивно. Для любого m от 1 до n , имеем

$$T_{4,n}^{m,i} = \{T_{4,n}^{m-1,i}, T_{4,n}^{m-1,i+4^{n-m}}, T_{4,n}^{m-1,4^n-i+1-4^{n-m}}, T_{4,n}^{m-1,4^n-i+1}\}, i = \overline{1, 4^{n-m}}.$$

Таким образом, для $k = 4$ и $n = 2$, получим

$$T_{4,2}^{modified\ equal\ gap} = \{\{1,5,12,16\}, \{2,6,11,15\}, \{3,7,10,14\}, \{4,8,9,13\}\}.$$

Модифицированное расписание равных преимуществ удовлетворяет свойствам равных сумм и равных разностей, объединяя свойства стандартного расписания и расписания равных преимуществ.

Утверждение 1.19. Для $k = 4$, расписание равных преимуществ является единственным расписанием, которое удовлетворяет свойствам равных преимуществ и баланса.

Для $k = 7$, существует расписание, удовлетворяющее свойствам равных преимуществ, симметричности, равных сумм и равного разделения проигрывающих:

$$T_{7,2} = 1,8,23,29,35,36,43,2,9,18,31,34,37,44,3,10,20,26,33,38,45,4,11,22,24,28,39,46,5,14,17,24,30,40,47,6,13,16,19,32,41,48,7,14,15,29,35,42,49.$$

Четвертый матч является симметричным самому себе. Первый и седьмой матчи, второй и шестой, третий и пятый являются симметричными парами матчей. Данный пример не является единственным.

1.2.4 Заключение по подразделу

Для случаев трех и четырех участников в одном матче построены обобщения стандартного расписания и расписания равных преимуществ. Данные случаи являются наиболее значимыми с практической точки зрения. Для большего количества участников необходимо определить дополнительные свойства, чтобы однозначно построить искомое расписание.

Результаты данного подраздела отчета отражены в [41].

1.3 Исследование конкурентной кооперации в Арктике

Явление, когда между фирмами есть конкуренция в одних направлениях и кооперация в других получила название конкурентного сотрудничества (coopetition) – см. [46]. При этом подобное явление может проявляться и быть изучено как при горизонтальном, так и при вертикальном взаимодействии по цепочке добавленной стоимости. Хотя первоначально модель конкурентного сотрудничества была разработана как полезная и гибкая концептуальная модель для изучения взаимодействия бизнеса с бизнесом, она также может быть принята для моделирования взаимодействия на других уровнях агрегирования сообщества и среди других типов агентов, например между различными отраслями, в разных ГЧП конфигурациях и даже между странами.

Арктический регион и его развитие являются хорошим примером для построения моделей конкурентного сотрудничества. Несмотря на то, что он занимает всего лишь 6 процентов земного шара, согласно последним исследованиям, более пятой части неразведанных запасов нефти и газа в мире находится в недрах выше Полярного круга и значительное количество видов рыбы обитает в арктических водах. Из-за экстремальных климатических условий и, как следствие, высокой стоимости, разведка и разработка природных ресурсов в Арктике выглядела непривлекательно несколько десятилетий назад. Однако, текущие изменения климата, приводящие к таянию льда, открыли новые возможности для региона. Арктические воды стали более дос-

тупными для разведки и разработки природных ресурсов. Арктические морские маршруты доставки становятся все более доступными, что потенциально дает значительные экономические преимущества для многих стран. Возможности для рыбной промышленности также значительно увеличиваются за счет увеличения продолжительности периода, когда поверхность Ледовитого океана свободна ото льда.

Сегодня проводится большое количество исследований по вопросу возможностей и потенциальных выигрышей от кооперации в Арктическом регионе между различными заинтересованными сторонами: между странами, причисленными к Арктическому кругу в целом и по отдельным вопросам [47, 48, 49], между странами из более отдаленных регионов, чтобы координировать свои действия в Арктике [50, 51], а также между различными отраслями промышленности в арктических странах [52, 53]. Совершенно очевидно, что, хотя многие проблемы в Арктике мотивируют заинтересованные стороны к сотрудничеству, их собственные интересы и возможности могут вызывать и уже вызывают расхождения между целями общей политики и результатами, которые могут быть получены при согласованном взаимодействии между ними [54]. Поэтому исследования конкурентного сотрудничества крайне актуальны для разработки инструментов разрешения конфликтов в рамках экономического развития Арктики, в частности для дележа спорных территорий – см. [55].

Конкурентное сотрудничество в конкретном контексте сотрудничества в области НИОКР и конкуренции на товарном рынке для рынков олигополии, на основе классической модели первоначально разработанной [56], было проанализировано с помощью экспериментального подхода в [57]. Данные наработки мы предполагаем взять за основу построения концептуальной платформы для исследования проблем и возможностей по кооперации в Арктике, уделяя больше внимания вопросам разработки и анализа механизмов взаимодействия участников конкурентного сотрудничества.

Формальные модели этого исследования служат основой для платформы бизнес-игр, которую мы реализуем, следуя концепции, описанной в [58] для анализа управленческих механизмов в сложных кибер-физических системах, предполагающего трех уровневый подход. Он охватывает нижний кибер-физический слой, см., например [59], на котором находятся объединенные модели природной и технической системы, экономический слой, моделирующий аспекты взаимодействия между заинтересованными сторонами и верхний слой – на котором моделируется индивидуальное принятие решений каждой из заинтересованных сторон. Одним из наиболее очевидных примеров, который в дальнейшем планируется рассмотреть с помощью развиваемого подхода, является рыбная отрасль, в которой инновационные подходы к разведению аквакультур [60, 61], предлагающие эффективные технические решения для экологических проблем могут оказать воздействие на взаимодействие между различными отраслями промышленности внутри одной страны [62, 63, 53], что, в свою очередь, может воздействовать на принятие политических решений на межгосударственном уровне – см. [48, 52, 64].

Такой многоуровневый подход позволит проводить исследования по следующим актуальным вопросам: будут ли стороны, которые конкурируют в некоторых аспектах своего взаимодействия, иметь стимулы к сотрудничеству в других или, наоборот, – будут ли стороны, которые сотрудничают в некоторых аспектах, конкурировать в других из-за конкретных аспектов функционирования конкретных киберфизических систем. Также, будут ли предполагаемые к внедрению технические инновации полезны для повышения уровня кооперации или нет – см., например [65].

Важным аспектом данного подхода является декомпозиция сложной управленческой проблемы на набор элементарных, «базовых» проблем, каждая из которых может быть проанализирована и решена в отдельности. Декомпозиционный подход показал себя достаточно успешным в области задач управления техническими системами [66], теории управления организационными [67] и эколого-экономическими [68] системами. Разработка упрощен-

ных и обобщенных «игр» в рамках данного подхода позволяет выявить самую суть конфликтной ситуации, отделив ее от других, порой крайне сложных аспектов рассматриваемой проблемы, тем самым сделав доступными ее понимание и анализ даже для неспециалистов, что в дальнейшем позволяет для каждой из элементарных управленческих проблем предложить и проиллюстрировать эффективные управленческие решения.

Реализовать данный подход мы предполагаем на основе одной из новейших систем компьютерного моделирования – Расчет Динамических Систем [69], развиваемой в ИПУ РАН, которая позволяет комбинировать блочный подход к имитационному моделированию механизмов управления в социально-экономических системах – см., например [70] с экспериментальными исследованиями этих механизмов в форме деловых игр [65], которые могут быть адаптированы и проведены даже с неподготовленной аудиторией, например – для детей, см. [71]. Деловые игры и мультиагентное имитационное моделирование могут быть также полезны на стадии их апробации при интеграции данных решений в контекст конкретной киберфизической системы для проверки их применимости, настройки и анализа эффективности их функционирования.

Таким образом, разрабатываемая концептуальная платформа для исследования проблем и возможностей по кооперации в Арктике на базе моделей конкурентной кооперации в идеологии многоуровневого подхода к декомпозиции сложных кибер-физических систем – путем выстраивания иерархии взаимосвязанных моделей для многостороннего анализа конфликтных ситуаций лежащих в основе многих управленческих задач, позволит моделировать одновременную кооперацию в одних направлениях и конкуренцию в других между различными агентами влияния. В 2018 г. была представлена первая версия концепции данной системы [72], в 2019 году планируется разработать прототип на базе конкретного примера, иллюстрирующего ситуацию конкурентной кооперации в Арктическом регионе.

2 Исследование моделей выбора и агрегирования

Этот раздел отчета содержит результаты исследований моделей выбора и агрегирования. Задачи выбора рассматривались в рамках развития теории многокритериального выбора, а также применения теории коллективного выбора. Задачи агрегирования рассматривались опять-таки в рамках теории коллективного выбора для построения агрегированных ранжирований альтернатив, а также в рамках теории функций доверия. Полученные результаты отражены в следующих подразделах:

- развития теории интервальных оценок замещений критериев;
- аксиоматический анализ способов выбора оптимальных вариантов и процедур ранжирования альтернатив;
- исследование мер конфликта в рамках теории функций доверия, индуцированных правилами комбинирования свидетельств;
- построение и применение теоретических моделей для расчета степени манипулируемости правил агрегирования предпочтений.

2.1 Развития теории интервальных оценок замещений критериев

2.1.1 Введение в теорию интервальных оценок замещений критериев

При анализе и выборе многокритериальных решений широко используются параметрические модели предпочтений. Самый известный пример – аддитивная функция ценности (в частности, линейная «свертка») с весами (коэффициентами важности) критериев. Так как назначение точных значений параметров по целому ряду причин (из-за наличия неопределенных факторов, неточности исходных данных, неполноты сведений о предпочтениях лица, принимающего решения, и т.п.) – непростая проблема, то такие значения оказываются недостаточно надежными и могут содержать ошибки. Поэтому важным этапом анализа прикладных задач является оценка чувствительности результатов выбора к изменению назначенных величин параметров, т.е. вы-

яснение пределов, в которых можно изменять исходные значения параметров так, чтобы результаты выбора оставались неизменными. Например, в руководстве по подаче предложений по политическому анализу (policy analysis) для Европейской Комиссии Европейского союза в контексте по оценке влияния [73] записано: «Если допущения, лежащие в основе базового сценария, могут изменяться в результате воздействия внешних факторов, необходимо провести анализ чувствительности, чтобы оценить, существенно ли различаются последствия вариантов политики для различных значений ключевых переменных».

С другой стороны, следует отметить, что результаты анализа чувствительности могут прояснить и требования к точности оценок параметров моделей предпочтений.

Методы анализа чувствительности хорошо развиты для однокритериальных задач оптимизации (см., например, [74]), и особенно для задач линейного программирования [75]. Для многокритериальных задач методы анализа чувствительности первоначально разрабатывались применительно к моделям предпочтений в виде аддитивной функции ценности (в том числе для методов анализа иерархий, TOPSIS и PROMETHEE), в которой параметрами являлись веса критериев, а затем и значения частных функций ценности [76, 77, 78, 79, 80, 81, 82, 83]. Однако разработанные для параметрических функций ценности подходы к анализу чувствительности в принципе не годятся для моделей с частичными бинарными отношениями, ибо при фиксированных значениях параметров функция ценности позволяет сравнивать по предпочтительности любые варианты, а частичное отношение предпочтения этого не позволяет. Поэтому здесь требуются иные методы, основанные на новых идеях.

Вначале методы анализа чувствительности для частичных отношений предпочтения были разработаны для моделей, базирующихся на отношениях превосходства (outranking relations) и использующихся в таких методах решения многокритериальных задач, как ELECTRE (разных видов) [84, 85]. В

[84] для метода ELECTRE I предложены аналитические методы расчета границ интервалов, в пределах которых изменение весов отдельных критериев, а также отдельных уровней согласия или несогласия не меняют множество недоминируемых альтернатив. В [85] для метода ELECTRE III, который используется для упорядочения альтернатив по предпочтительности на основании конструируемого нечеткого отношения превосходства (outranking relation), предложен метод оценки «устойчивости» полученного упорядочения при помощи трех показателей, имеющих вероятностный смысл и определяемых для каждой пары альтернатив: индекс приемлемости рангов (альтернатив в их итоговой упорядоченности), индекс парного выигрыша (превосходства в предпочтительности) и индекс несравнимости. При этом предполагается, что для параметров отношения превосходства (весов критериев и пороговых уровней), а также и для значений критериев заданы области (обычно интервалы), в которых истинные значения этих параметров распределены равномерно. Значения указанных индексов рассчитываются методом Монте-Карло. Следует иметь в виду, что отношения превосходства не являются транзитивными.

Для транзитивных отношений предпочтения методы анализа чувствительности начали создаваться лишь в последнее десятилетие. Во множестве многокритериальных моделей предпочтений, основанных на частичных бинарных отношениях, можно выделить модели с однородными (обычно безразмерными) параметрами и модели с неоднородными (имеющими разную «физическую» размерность) параметрами. Примерами моделей первого из указанных классов являются модели из теории (симметрической, или истинной) важности (однородных, т.е. имеющих общую шкалу, а потому и единую область значений) критериев [86, 87, 88]. Примерами моделей второго класса служат модели из теории параметрической важности неоднородных критериев и, в частности, из теории интервальных замещений критериев [89, 90].

Чувствительность выбора к изменению коэффициентов важности однородных критериев с порядковой шкалой или шкалой первой порядковой мет-

рики и конечной областью значений изучена в [91, 92, 93, 94]. В [91, 92] рассматривается случай точечной оценки важности, когда задаются точные значения коэффициентов важности. Оценивание чувствительности множества недоминируемых альтернатив (или отдельных альтернатив) к изменению значений коэффициентов важности сводится к решению задач линейного программирования. В [93, 94] рассматривается случай интервальной информации о важности критериев, состоящей из интервальных оценок степеней превосходства в важности одних критериев над другими. Здесь задача оценивания чувствительности множества недоминируемых альтернатив (или отдельных альтернатив) к изменению границ интервальных оценок значений коэффициентов важности оказывается достаточно сложной задачей нелинейного программирования, которая не сводится к решению задач линейного программирования. Поэтому был предложен аналитический метод оценки чувствительности результатов сравнения по предпочтительности альтернатив к изменению границ указанных интервальных оценок.

В [95] рассматривались параметрические модели предпочтений, возникающие при использовании методов параметрической (пропорциональной) важности (неоднородных) критериев. Предпочтения представляются в форме интервалов неопределенности замещения (ИНЗ) одних критериев другими. Задача оценивания чувствительности множества альтернатив к изменению границ ИНЗ оказалась нелинейной и достаточно сложной. Поэтому была поставлена проблема разработки эффективных методов анализа чувствительности для моделей с ИНЗ. В данном подразделе отчета приведены результаты решения указанной актуальной проблемы, имеющей как теоретическую, так и прикладную значимость.

2.1.2 Сведения из теории параметрической важности критериев и теории интервалов неопределенности замещений

Дальнейшее изложение опирается на следующую математическую модель ситуации индивидуального выбора в условиях определенности при многих критериях:

$$\mathcal{M} = \langle X, f_1, \dots, f_m, Z_1, \dots, Z_m, P \rangle, \quad (2.1)$$

где X – множество альтернатив, или вариантов, планов, стратегий, ... (число их не менее двух), f_1, \dots, f_m – критерии ($m \geq 2$), т.е. функции $f_i: X \rightarrow Z_i$, где $Z_i \subseteq \mathbb{R} = (-\infty, +\infty)$ – область значений i -го критерия (множество его шкальных оценок), P – отношение (строгого) предпочтения ЛПР. Критерии f_i образуют векторный критерий $f = (f_1, \dots, f_m)$. Предполагается, что каждый из критериев независим по предпочтению от остальных и его бóльшие значения предпочтительнее меньших. Каждая альтернатива x из множества X характеризуется её *векторной оценкой* $y(x) = f(x) = (f_1(x), \dots, f_m(x))$. Множество всех векторных оценок (как достижимых, которые соответствуют альтернативам, так и гипотетических) есть $Z = Z_1 \times \dots \times Z_m$.

Предпочтения ЛПР моделируются при помощи отношения предпочтения P на Z : запись yPz означает, что векторная оценка y более предпочтительна, чем z . Принимается допущение, что отношение P – строгий (частичный) порядок, т.е. оно иррефлексивно (для любой векторной оценки y неверно yPy) и транзитивно (для любых векторных оценок y, z, u из yPz и zPu следует yPu). Предполагается, что альтернатива полностью характеризуется её векторной оценкой, так что сравнение вариантов по предпочтительности сводится к сопоставлению их векторных оценок. Поэтому задача выбора наилучшей альтернативы из множества X сводится к выбору наиболее предпочтительной векторной оценки из множества достижимых векторных оценок $Y = f(X) = \{y \in Z \mid y = f(x), x \in X\}$.

Отношение P неизвестно и строится (восстанавливается) на основе информации о предпочтениях, получаемой от ЛПР и/или экспертов. Правило,

задающее отношение P на основе такой информации, называется решающим. Оно может быть аналитическим (задаваться формулами) или алгоритмическим (в частности, оптимизационным). Далее рассматривается случай, когда в состав информации о предпочтениях ЛППР входят сведения об относительной важности критериев, или компенсациях ухудшения значений одних критериев улучшением значений других критериев (как принято говорить, о замещениях одних критериев другими).

Поскольку бóльшие значения каждого критерия предпочтительнее его меньших значений, то полагается, что отношение предпочтения P согласовано с отношением Парето, или Эджворта-Парето P^0 (т.е. верно $P^0 \subset P$), которое определяется так:

$$yP^0z \Leftrightarrow y_i \geq z_i, i \in M = \{1, \dots, m\}, x \neq y. \quad (2.2)$$

Приведем сведения из теории параметрической важности и интервалов неопределенности замещений критериев [90, 89, 96, 97, 98], необходимые для дальнейшего изложения. Через $(y \parallel y_i + a, y_j - b)$ обозначим вектор, полученный из вектора y заменой его компоненты y_i на $y_i + a$ и компоненты y_j на $y_j - b$.

Определение 2.1. Критерий f_i важнее критерия f_j с параметрами (или в отношении) $h_i: h_j$, где h_i и h_j – положительные числа, если

$$(y \parallel y_i + h_i t, y_j - h_j t)Py \text{ для любых } y \in Z, t > 0, y_i + h_i t \in Z, y_j - h_j t \in Z. \quad (2.3)$$

Сообщение о том, что критерий f_i важнее критерия f_j с параметрами $h_i: h_j$, обозначается “словом” $i \succ^{h_i: h_j} j$. Отношение, задаваемое (2.3), обозначается $P(i \succ^{h_i: h_j} j)$.

Определение 2.2. Интервалом неопределенности замещения (ИНЗ) критерия f_i критерием f_j называется интервал $\lambda_{ij} = (\lambda_{ij}^-, \lambda_{ij}^+)$, где $0 < \lambda_{ij}^- < \lambda_{ij}^+$, обладающий следующими свойствами:

$$(y \parallel y_i - t, y_j + \lambda_{ij}^+ t)Py \text{ для любых } y \in Z, t > 0, y_i - t \in Z, y_j + \lambda_{ij}^+ t \in Z; \quad (2.4)$$

$$yP(y \parallel y_i - t, y_j + \lambda_{ij}^- t) \text{ для любых } y \in Z, t > 0, y_i - t \in Z, y_j + \lambda_{ij}^- t \in Z. \quad (2.5)$$

Отношение, задаваемое (2.4) и (2.5), обозначается $P(\lambda_{ij})$. Сравнение (2.3) с (2.4) – (2.5) показывает, что одному ИНЗ $\lambda_{ij} = (\lambda_{ij}^-, \lambda_{ij}^+)$ соответствуют два сообщения о важности: $i \succ^{1:\lambda_{ij}^-} j$ и $j \succ^{\lambda_{ij}^+:1} i$. Поскольку практически сведения о предпочтениях ЛПР удобнее получать в виде интервальных оценок замещений [99], то далее будем считать, что информация о предпочтениях ЛПР представляется в виде множества Λ , состоящего из ИНЗ λ_{ij} . Поскольку отношение P транзитивно, то отношение предпочтения $P(\Lambda)$, порождаемое на Z информацией Λ , определяется как наименьшее транзитивное отношение, включающее объединение отношения P^0 и всех отношений $P(\lambda_{ij})$. В соответствии с данным определением $yP(\Lambda)z$ выполнено в том и только в том случае, если существует цепочка вида

$$yP^1 z^1, z^1 P^2 z^2, \dots, z^{n-1} P^n z,$$

где все $z^k \in Z$, а P^k есть P^0 или $P(\lambda_{ij})$ для некоторого $\lambda_{ij} \in \Lambda$ при каждом $k = 1, \dots, n$ (число n , разумеется, зависит от y, z).

Для произвольного состава множества Λ разработаны оптимизационные решающие правила, позволяющие свести проверку справедливости соотношения $yP(\Lambda)z$ к решению задачи линейного программирования. Для ряда специальных случаев удалось разработать и аналитические решающие правила. Далее будем рассматривать тот важный для практических приложений случай, когда Λ состоит из ИНЗ для каждого из критериев f_2, \dots, f_m и критерия f_1 (его называют базовым): $\Lambda = \{\lambda_{21}, \dots, \lambda_{m1}\}$. Для такого состава Λ в предположении, что все критерии непрерывны и не ограничены, в [96] было предложено простое аналитическое решающее правило. Затем в [90] оно было переформулировано в более удобной и наглядной форме и было показано, что оно остается при определенных условиях верным даже тогда, когда критерии ограничены или дискретны.

Пусть y и z – произвольные фиксированные векторные оценки и $\delta = y - z$. Выделим из множества всех критериев без базового $M' = \{2, \dots, m\}$ два непересекающиеся подмножества: $M^+(\delta) = \{i \in M' \mid \delta_i > 0\}$, $M^-(\delta) = \{i \in M' \mid \delta_i < 0\}$. При $\delta_i \neq 0$ для некоторого $i > 1$ по крайней мере одно из этих подмножеств не пусто. Положим:

$$l = \delta_1 + \sum_{i \in M^+(\delta)} \delta_i \lambda_{i1}^- + \sum_{i \in M^-(\delta)} \delta_i \lambda_{i1}^+; \quad r = \delta_1 + \sum_{i \in M^+(\delta)} \delta_i \lambda_{i1}^+ + \sum_{i \in M^-(\delta)} \delta_i \lambda_{i1}^-. \quad (2.6)$$

Согласно известной договоренности, если одно из множеств $M^+(\delta)$ или $M^-(\delta)$ пусто, то соответствующая сумма в (2.6) равна нулю. Заметим, что $l \leq r$, причем равенство достигается только при $\delta = 0$, т.е. когда $y = z$. Полезно также иметь в виду, что для $\delta' = -\delta$ верны равенства $l' = -r$ и $r' = -l$.

Теорема 2.1. Пусть в задаче с $\Lambda = \{\lambda_{21}, \dots, \lambda_{m1}\}$ выполнено одно из следующих условий: а) все критерии континуальны; б) базовый континуальный критерий неограниченный или ограничен только с одной стороны. Тогда для векторных оценок y, z таких, что $y_i \neq z_i$ для некоторого $i > 1$, справедливо решающее правило:

$$yP(\Lambda)z \Leftrightarrow l \geq 0; \quad zP(\Lambda)y \Leftrightarrow r \leq 0; \quad yN(\Lambda)z \Leftrightarrow l < 0 < r. \quad (2.7)$$

В (2.7) через $N(\Lambda)$ обозначено отношение несравнимости: $yN(\Lambda)z$ означает, что $y \neq z$ и неверно ни $yP(\Lambda)z$, ни $zP(\Lambda)y$.

Пусть $\hat{\Lambda} = \{\hat{\lambda}_{21}, \dots, \hat{\lambda}_{m1}\}$ – множество ИНЗ, которые являются сужением ИНЗ из Λ , т.е. справедливы включения $(\hat{\lambda}_{i1}^-, \hat{\lambda}_{i1}^+) \subseteq (\lambda_{i1}^-, \lambda_{i1}^+)$, $i = 2, \dots, m$, причем хотя бы одно из них строгое. Тогда отношение $P(\hat{\Lambda})$ не уже отношения $P(\Lambda)$, т.е. $P(\hat{\Lambda}) \supseteq P(\Lambda)$, причем это включение может быть строгим (например, когда все критерии континуальны и не ограничены).

Примечание 2.1. Выше предполагалось, что все критерии желательно максимизировать. При наличии в задаче критериев, которые нужно минимизировать, следует внести в определения (2.3) – (2.5) соответствующие очевидные изменения. Приведем здесь лишь формулы (2.6.) для случая возмож-

ного наличия и минимизируемых критериев. Введем переменную s_i , положив $s_i = 1$, если критерий f_i желательно максимизировать, и $s_i = -1$, если критерий f_i желательно минимизировать, и пусть $M^+(s, \delta) = \{i \in M' \mid s_i \delta_i > 0\}$, $M^-(s, \delta) = \{i \in M' \mid s_i \delta_i < 0\}$. Формулы (2.6.) переписываются так:

$$l = s_1 \delta_1 + \sum_{i \in M^+(s, \delta)} s_i \lambda_{i1}^- \delta_i + \sum_{i \in M^-(s, \delta)} s_i \lambda_{i1}^+ \delta_i, \quad r = s_1 \delta_1 + \sum_{i \in M^+(s, \delta)} s_i \lambda_{i1}^+ \delta_i + \sum_{i \in M^-(s, \delta)} s_i \lambda_{i1}^- \delta_i.$$

Формулировка теоремы 2.1 не изменяется.

2.1.3 Анализ чувствительности

Пусть ситуация многокритериального выбора описывается математической моделью (2.1), а информация о предпочтениях ЛПР представлена в виде множества ИНЗ $\Lambda = \{\lambda_{21}, \dots, \lambda_{m1}\}$, порождающего на множестве векторных оценок Z отношение предпочтения $P(\Lambda)$. Будем полагать, что условие теоремы 2.1 выполнено, так что справедливо решающее правило (2.7). Множество альтернатив X считается конечным: $X = \{x^1, \dots, x^j, \dots, x^n\}$.

Векторная оценка $y^j = f(x^j)$ называется недоминируемой (по $P(\Lambda)$), если не существует векторной оценки $y^k = f(x^k)$ такой, что верно $y^k P(\Lambda) y^j$, $k = 1, \dots, n$; альтернатива x^j также называется недоминируемой. В противном случае альтернатива x^j и ее векторная оценка y^j называются доминируемыми (по $P(\Lambda)$). Множество недоминируемых векторных оценок внешне устойчиво (см., например, [100]): если y^k – доминируемая векторная оценка, то найдется недоминируемая векторная оценка y^j такая, что верно $y^j P(\Lambda) y^k$. Поэтому оптимальную альтернативу следует выбрать среди недоминируемых.

Целью анализа чувствительности отдельной недоминируемой альтернативы или множества таких альтернатив к изменению (границ) ИНЗ из Λ является получение оценок максимально возможных изменений, при которых недоминируемость сохраняется. Далее будем полагать, что уменьшение меньших из изменений не компенсируется увеличением больших из них. Поскольку после расширения ИНЗ отношение предпочтения не расширится, а

может лишь сузиться, то недоминируемость сохранится. Поэтому можно ограничиться рассмотрением случая сужения исходных ИНЗ из Λ до ИНЗ, составляющих $\hat{\Lambda} = \{\hat{\lambda}_{21}, \dots, \hat{\lambda}_{m1}\}$.

Векторную оценку y^k назовем потенциально доминирующей над векторной оценкой y^j , если найдутся такие ИНЗ $\hat{\lambda}_{21}, \dots, \hat{\lambda}_{m1}$, являющиеся сужением соответствующих ИНЗ из Λ , что будет верно $y^k P(\hat{\Lambda}) y^j$.

Пусть $\hat{\Lambda} = \{\hat{\lambda}_{21}, \dots, \hat{\lambda}_{m1}\}$ – множество ИНЗ, которые являются сужением ИНЗ из Λ , и верно $y^k P(\hat{\Lambda}) y^j$. Если y^k доминируема по $P(\Lambda)$, то, в силу внешней устойчивости множества недоминируемых по $P(\Lambda)$ векторных оценок среди них найдется y^s такая, что верно $y^s P(\Lambda) y^k$. Но $P(\Lambda) \subseteq P(\hat{\Lambda})$ и отношение $P(\hat{\Lambda})$ транзитивно. Поэтому верно и $y^s P(\hat{\Lambda}) y^j$. Следовательно, при анализе чувствительности можно учитывать только те потенциально доминирующие векторные оценки, которые сами являются недоминируемыми по $P(\Lambda)$.

В многокритериальных прикладных задачах критерии могут иметь разную «физическую» размерность, так что и величины-границы ИНЗ тоже могут иметь разную размерность. Для соизмерения степеней сужения разных ИНЗ поступим так. Положим

$$d_i = \lambda_{i1}^+ - \lambda_{i1}^-, \hat{\lambda}_{i1}^- = \lambda_{i1}^- + t_i^- d_i, \hat{\lambda}_{i1}^+ = \lambda_{i1}^+ - t_i^+ d_i, t_i^- \geq 0, t_i^+ \geq 0, i = \{2, \dots, m\}.$$

Поскольку должны выполняться неравенства $\hat{\lambda}_{i1}^- < \hat{\lambda}_{i1}^+$, то $t_i^- + t_i^+ < 1, i = \{2, \dots, m\}$. Таким образом, система ограничений для переменных t_i^- и t_i^+ такова:

$$t_i^- \geq 0, t_i^+ \geq 0, t_i^- + t_i^+ < 1, i \in M' = \{2, \dots, m\}. \quad (2.8)$$

Относительная длина суженного ИНЗ $\hat{\lambda}_{i1} = (\hat{\lambda}_{i1}^-, \hat{\lambda}_{i1}^+)$ равна

$$\frac{\hat{\lambda}_{i1}^+ - \hat{\lambda}_{i1}^-}{d_i} = 1 - (t_i^+ + t_i^-) > 0, i \in M'.$$

Поэтому степени сужения ИНЗ можно оценивать величинами t_i^- и t_i^+ : чем они больше, тем сильнее сужение соответствующего ИНЗ. В итоге приходим к многокритериальной задаче максимизации с $2(m - 1)$ равноважными критериями – величинами t_i^- и t_i^+ (точное определение равноважности критериев см. в [101]). Известно, что информация о том, что все критерии равноважны, позволяет лишь сузить множество выбора, но не обеспечивает его однозначность. Для окончательного выбора следует привлечь дополнительную информацию, касающуюся возможности компенсации потерь по одним критериям выигрышами по другим. Крайним является тот случай, когда такая компенсация считается недопустимой. Возникающие в таком случае задачи называются симметрически-лексикографическими, или, кратко, SL-задачами оптимизации [101, 102]. Будем полагать, что увеличение больших значений одних величин t_i^- и t_i^+ не компенсирует уменьшение меньших значений других таких величин. Иначе говоря, вначале желательно максимизировать наименьшую из величин t_i^- и t_i^+ , $i \in M'$, затем – следующую в порядке их возрастания (точнее, неубывания), и т.д. SL-задачи такого рода известны под названием лексиминных [103]. Решение поставленной лексиминной задачи сводится к решению последовательности задач максимизации вначале наименьшего из указанных $2(m - 1)$ величин t_i^- и t_i^+ при соответствующих ограничениях на переменные t_i^- и t_i^+ , $i = 2, \dots, m$, затем – следующего по величине (в порядке возрастания) критерия, и т.д.

Векторная оценка y^k будет потенциально доминирующей над векторной оценкой y^j , если найдутся числа t_i^- и t_i^+ , $i \in M'$, такие, что будет верно $y^k P(\hat{\Lambda}) y^j$, т.е., согласно (2.6), справедливо неравенство $l_{kj} \geq 0$. Его можно раскрыть и записать в таком виде:

$$-\sum_{i \in M^-(\delta)} d_i \delta_i^{kj} t_i^+ + \sum_{i \in M^+(\delta)} d_i \delta_i^{kj} t_i^- \geq -l_{kj}, \quad (2.9)$$

где l_{kj} вычисляется для $\delta = \delta^{kj} = y^k - y^j$. Отметим, что фактически переменные t_i^- и t_i^+ входят в левую часть неравенства (2.9) с положительными коэффициентами; если векторная оценка y^j недоминируема по $P(\Lambda)$, то в правой части неравенства (2.9) стоит положительная величина. Проверку справедливости сформулированного условия потенциальной доминируемости можно провести с использованием компьютера, решая задачу линейного программирования с любой целевой функцией (например, $0 \cdot t_{21}^-$) и ограничениями (2.8) – (2.9).

Пусть $N(j)$ – множество отличных от y^j недоминируемых по $P(\Lambda)$ векторных оценок, потенциально доминирующих над y^j . Это множество будем полагать не пустым: в противном случае анализ чувствительности теряет смысл, ибо y^j будет оптимальной по $P(\Lambda)$ и останется таковой при любом сужении ИНЗ из Λ . Условие, что неверно $y^k P(\hat{\Lambda}) y^j$, выполнено, когда верно $y^k N(\hat{\Lambda}) y^j$ или $y^j P(\hat{\Lambda}) y^k$, т.е., согласно (2.7), когда справедливо неравенство $\hat{l}_{kj} < 0$. Поэтому система ограничений, обеспечивающих недоминируемость y^j , включает неравенства $\hat{l}_{kj} < 0$, $k \in N(j)$. Используя (2.6), раскроем эти неравенства и получим неравенства

$$- \sum_{i \in M^-(\delta)} \delta_i^{kj} d_i t_i^+ + \sum_{i \in M^-(\delta)} \delta_i^{kj} d_i t_i^- < -l_{kj}, \quad k \in N(j). \quad (2.10)$$

Итак, первой оптимизационной задачей является задача максимизации целевой функции $\min_{i \in M'} \min\{t_i^-, t_i^+\}$ при ограничениях (2.8), (2.10). Для получения оптимизационной задачи в стандартной формулировке заменим строгие неравенства в (2.8) и (2.10) нестрогими (этот факт далее будет учтен тем, что полученные границы для ИНЗ будут полагаться недостижимыми). Используя известный стандартный прием замены нелинейной целевой функции линейной за счет введения дополнительной переменной t (ее можно считать неотрицательной), в итоге получим задачу линейного программирования

с ограниченной целевой функцией и непустым множеством допустимых решений:

$$t \rightarrow \max \text{ при ограничениях } t \leq t_i^-, t \leq t_i^+, i = \{2, \dots, m\}, \text{ и (3.1), (3.3).} \quad (2.11)$$

Далее поступаем следующим образом [104, 105]. Пусть $t^*, \{t_i^{-*}, t_i^{+*}\}$ – решение задачи (2.11), M_1 – совокупность номеров i таких, что $t^* = t_i^{-*}$ и соответствующие двойственные переменные равны нулю, и M_2 – совокупность номеров i таких, что $t^* = t_i^{+*}$ и соответствующие двойственные переменные равны нулю. Если оба множества M_1 и M_2 пусты, то решение SL-задачи найдено.

В противном случае переходим ко второй задаче линейного программирования: максимизировать t при ограничениях $t \leq t_i^-, i \in M_1; t \leq t_i^+ i \in M_2$, а также неравенствах (2.10) при фиксированных значениях $t_i^- = t_i^{-*}$ для $i \in M_1$ и $t_i^+ = t_i^{+*}$ для $i \in M_2$. (Понятно, что неравенства из (3.3) и $t_i^- + t_i^+ \leq 1$, в которых значения всех переменных оказались фиксированными, можно опустить). Если для решения второй задачи все неравенства $t \leq t_i^-$ и $t \leq t_i^+$ выполняются как равенства и соответствующие им двойственные переменные положительны, то решение SL-задачи найдено. В противном случае аналогичным путем формируем третью задачу линейного программирования, и т.д.

Пусть $\{\bar{t}_i^-, \bar{t}_i^+\}$ – решение лексиминной задачи, т.е. последней из описанной последовательности задач линейного программирования. Тогда искомыми минимальными ИНЗ, при которых альтернатива остается недоминируемой, будут ИНЗ с концами $\bar{\lambda}_{i1}^- = \lambda_{i1}^- + \bar{t}_i^- d_i$ и $\bar{\lambda}_{i1}^+ = \lambda_{i1}^+ - \bar{t}_i^+ d_i, i = \{2, \dots, m\}$. Эти концы, как указывалось выше, следует считать недостижимыми, так что (учитывая, что ограничений на расширение ИНЗ нет) левый конец $\bar{\lambda}_{i1}^-$ ИНЗ λ_{i1} может изменяться в пределах от $0 + \varepsilon$ до $\bar{\lambda}_{i1}^- + \bar{t}_i^- d_i - \varepsilon$, а правый – от $\bar{\lambda}_{i1}^+ - \bar{t}_i^+ d_i + \varepsilon$ до $+\infty$, где ε – сколь угодно малое положительное число.

Поскольку решением задачи выбора с использованием отношения предпочтения $P(\Lambda)$ является множество недоминируемых по $P(\Lambda)$ альтернатив, то для практических приложений основной интерес представляет анализ устойчивости этого множества в целом. С учетом изложенного выше, приходим к выводу, что для проведения этого анализа следует решить SL-задачу, отличающуюся от рассмотренной выше тем, что вместо множества $N(j)$ для фиксированного номера j нужно учесть множества $N(j)$ номеров j всех недоминируемых по $P(\Lambda)$ альтернатив x^j . То есть вместо ограничений (3.3) здесь нужно использовать ограничения

$$-\sum_{i \in M^-(\delta)} \delta_i^{kj} d_i t_i^+ + \sum_{i \in M^-(\delta)} \delta_i^{kj} d_i t_i^- < -l_{kj}, \quad k \in N(j), j \in N(\Lambda), \quad (2.12)$$

где $N(\Lambda)$ – множество номеров альтернатив, недоминируемых по $P(\Lambda)$.

Примечание 2.2. Для грубой оценки чувствительности можно ограничиться решением лишь одной – первой задачи линейного программирования.

Примечание 2.3. Выше рассмотрен анализ чувствительности к изменению всех ИНЗ. Если интересоваться только частью ИНЗ (в частности, чувствительностью к изменению границ одного из ИНЗ), то в описанный порядок проведения анализа нужно внести очевидные изменения.

2.1.4 Пример проведения анализа чувствительности

Для пояснения и иллюстрации изложенных выше теоретических положений рассмотрим числовой пример. В трехкритериальной задаче выбора множество X образовано тремя альтернативами: $X = \{x^1, x^2, x^3\}$. Все критерии являются континуальными неограниченными. Значения векторного критерия $f = (f_1, f_2, f_3)$ для альтернатив таковы:

$$f(x^1) = y^1 = (8, 3, 7), f(x^2) = y^2 = (7, 10, 5), f(x^3) = y^3 = (7, 2, 9).$$

Множество Λ состоит из двух ИНЗ $\lambda_{21} = (0, 2; 2)$ и $\lambda_{31} = (0, 3; 1)$. Здесь $m = 3$, $M' = \{2, 3\}$, $d_2 = 1, 8$ и $d_3 = 0, 7$. Условия теоремы 2.1. выполнены.

Для сравнения по предпочтительности альтернатив воспользуемся решающим правилом (2.7). Согласно (2.6) для x^1 и x^2 имеем:

$$\delta^{12} = y^1 - y^2 = (1, -7, 2), M^+(\delta^{12}) = \{3\}, M^-(\delta^{12}) = \{2\};$$

$$l_{12} = 1 + 2 \cdot 0,3 + (-7) \cdot 2 = -12,4; r_{12} = 1 + 2 \cdot 1 + (-7) \cdot 0,2 = 1,6;$$

поскольку $l_{12} < 0 < r_{12}$, то $y^1 N^\Lambda y^2$.

Аналогично для x^1 и x^3 :

$$\delta^{13} = y^1 - y^3 = (1, 1, -2), M^+(\delta^{13}) = \{2\}, M^-(\delta^{13}) = \{3\};$$

$$l_{13} = 1 + 1 \cdot 0,2 - 2 \cdot 1 = -0,8; r_{13} = 1 + 1 \cdot 2 - 2 \cdot 0,3 = 2,4;$$

поскольку $l_{13} < 0 < r_{13}$, то $y^1 N^\Lambda y^3$.

Наконец, для x^2 и x^3 :

$$\delta^{23} = y^2 - y^3 = (-6, 8, -4), M^+(\delta^{23}) = \{2\}, M^-(\delta^{23}) = \{3\};$$

$$l_{23} = -6 + 8 \cdot 0,2 - 4 \cdot 1 = -8,4; r_{23} = -6 + 8 \cdot 2 - 4 \cdot 0,3 = 8,8;$$

поскольку $l_{23} < 0 < r_{23}$, то $y^2 N^\Lambda y^3$.

Таким образом, все три альтернативы недоминируемы по P^Λ .

Проведем анализ чувствительности для альтернативы x^1 . Вначале проверим, является ли альтернатива x^2 потенциально доминирующей над x^1 . Система неравенств (2.8) и (2.10) с учетом $\delta^{21} = y^2 - y^1 = (-1, 7, -2)$ и $-l_{21} = r_{12} = 1,6$ записывается так:

$$12,6 \cdot t_2^- + 1,4 \cdot t_3^+ \geq 1,6; t_2^- + t_2^+ \leq 1, t_3^- + t_3^+ \leq 1, t_2^- \geq 0, t_2^+ \geq 0, t_3^- \geq 0, t_3^+ \geq 0.$$

Легко видеть, что эта система совместна: подходят, например, значения $t_2^- = 0,2$ и $t_2^+ = 0, t_3^- = 0, t_3^+ = 0$. Поэтому x^2 является потенциально доминирующей над x^1 .

Проверим теперь, является ли альтернатива x^3 потенциально доминирующей над x^1 . Система неравенств (2.9) и (2.10) с учетом $\delta^{31} = -\delta^{13} = (-1, -1, 2)$ и $-l_{31} = r_{13} = 2,4$ записывается так:

$$1,8 \cdot t_2^+ + 1,4 \cdot t_3^- \geq 2,4; \quad t_2^- + t_2^+ \leq 1, \quad t_3^- + t_3^+ \leq 1, \quad t_2^- \geq 0, \quad t_2^+ \geq 0, \quad t_3^- \geq 0, \quad t_3^+ \geq 0.$$

Она совместна: подходят, например, значения $t_2^- = 0$, $t_2^+ = 0,8$, $t_3^- = 0,8$, $t_3^+ = 0$. Поэтому x^3 также является потенциально доминирующей над x^1 . Таким образом, $N(1) = \{2, 3\}$.

Теперь запишем задачу (2.11):

$$t \rightarrow \max \quad (2.13)$$

при ограничениях:

$$t \leq t_2^-, \quad t \leq t_2^+, \quad t \leq t_3^-, \quad t \leq t_3^+; \quad (2.14)$$

$$t_2^- + t_2^+ \leq 1, \quad t_3^- + t_3^+ \leq 1; \quad (2.15)$$

$$12,6 \cdot t_2^- + 1,4 \cdot t_3^+ \leq 1,6; \quad 1,8 \cdot t_2^+ + 1,4 \cdot t_3^- \leq 2,4; \quad (2.16)$$

$$t_2^- \geq 0, \quad t_2^+ \geq 0, \quad t_3^- \geq 0, \quad t_3^+ \geq 0. \quad (2.17)$$

Решение задачи (4.1) – (4.5): $t = 0,1142857$; $t_2^- = 0,1142857$; $t_2^+ = 0,1142857$; $t_3^- = 0,1142857$; $t_3^+ = 0,1142857$. Все неравенства из (2.14) выполняются как равенства, но двойственные переменные для второго и третьего равны нулю, а для первого и четвертого положительны. Поэтому фиксируем величины $t_2^- = 0,1142857$ и $t_3^+ = 0,1142857$ и убираем содержащие их неравенства из (2.14). Получаем вторую задачу:

$$t \rightarrow \max \quad (2.18)$$

при ограничениях:

$$t \leq t_2^+, \quad t \leq t_3^-; \quad (2.19)$$

$$t_2^- + t_2^+ \leq 1, \quad t_3^- + t_3^+ \leq 1; \quad (2.20)$$

$$3,6 \cdot t_2^+ + 0,7 \cdot t_3^- \leq 2,7; 10,8 \cdot t_2^+ + 0,7 \cdot t_3^+ \leq 7; \quad (2.21)$$

$$t_2^- = 0,1142857, t_2^+ \geq 0, t_3^- \geq 0, t_3^+ = 0,1142857. \quad (2.22)$$

Решение задачи (2.18) – (2.22): $t = 0,7500000$; $t_2^- = 0,1142857$; $t_2^+ = 0,7500000$; $t_3^- = 0,7500000$; $t_3^+ = 0,1142857$. Оба неравенства из (2.19) выполняются как равенства и двойственные переменные для них положительны. Поэтому решение SL-задачи получено. Таким образом, альтернатива x^1 остается недоминируемой, если границы ИНЗ изменяются в таких пределах (с точностью до двух знаков после запятой; округление для λ_{21}^- и λ_{31}^- в меньшую сторону, а для λ_{31}^+ – в большую сторону):

$$0 < \lambda_{21}^- \leq 0,40 \approx 0,2 + 1,8 \cdot 0,1142857; 2 - 1,8 \cdot 0,75 = 0,65 < \lambda_{21}^+ < +\infty;$$

$$0 < \lambda_{31}^- \leq 0,82 \approx 0,3 + 0,7 \cdot 0,75; 1 - 0,7 \cdot 0,1142857 \approx 0,93 \leq \lambda_{31}^+ < +\infty.$$

Проведем теперь анализ чувствительности к изменению ИНЗ множества недоминируемых альтернатив $\{x^1, x^2, x^3\}$ в целом, учитывая результаты, полученные выше при сравнении альтернатив по P^Λ и при анализе чувствительности альтернативы x^1 .

Вначале проверим, является ли альтернатива x^1 потенциально доминирующей над x^2 . Система неравенств (2.9) и (2.10) с учетом $\delta^{12} = y^1 - y^2 = (1, -7, 2)$ и $-l_{12} = 12,4$ записывается так:

$$12,6 \cdot t_2^+ + 1,4 \cdot t_3^- \geq 12,4; t_2^- + t_2^+ \leq 1, t_3^- + t_3^+ \leq 1, t_2^- \geq 0, t_2^+ \geq 0, t_3^- \geq 0, t_3^+ \geq 0.$$

Легко видеть, что эта система совместна: подходят, например, значения $t_2^- = 0,9$, $t_2^+ = 0$, $t_3^- = 0,9$, $t_3^+ = 0$. Поэтому x^1 является потенциально доминирующей над x^2 .

Теперь проверим, является ли альтернатива x^3 потенциально доминирующей над x^2 . Система неравенств (2.9) и (2.10) с учетом $\delta^{32} = -\delta^{23} = (6, -8, 4)$ и $-l_{32} = r_{23} = 8,8$ записывается так:

$$14,4 \cdot t_2^+ + 2,8 \cdot t_3^- \geq 8,8; t_2^- + t_2^+ \leq 1, t_3^- + t_3^+ \leq 1, t_2^- \geq 0, t_2^+ \geq 0, t_3^- \geq 0, t_3^+ \geq 0.$$

Она совместна: подходят, например, значения $t_2^- = 0, t_2^+ = 0,9; t_3^- = 0, t_3^+ = 0,9$. Поэтому x^3 также является потенциально доминирующей над x^2 . Таким образом, $N(2) = \{1, 3\}$.

Проверим, является ли альтернатива x^1 потенциально доминирующей над x^3 . Система неравенств (2.9) и (2.10) с учетом $\delta^{13} = (1, 1, -2)$ и $-l_{13} = 0,8$ записывается так:

$$1,8 \cdot t_2^- + 1,4 \cdot t_3^+ \geq 0,8; t_2^- + t_2^+ \leq 1, t_3^- + t_3^+ \leq 1, t_2^- \geq 0, t_2^+ \geq 0, t_3^- \geq 0, t_3^+ \geq 0.$$

Она совместна: подходят, например, значения $t_2^- = 0,5; t_2^+ = 0, t_3^- = 0, t_3^+ = 0$. Поэтому x^1 является потенциально доминирующей над x^3 .

Проверим, наконец, является ли альтернатива x^2 потенциально доминирующей над x^3 . Система неравенств (2.9) и (2.10) с учетом $\delta^{23} = (-6, 8, -4)$ и $l_{23} = 8,4$ записывается так:

$$14,4 \cdot t_2^- + 2,8 \cdot t_3^+ \geq 8,4; t_2^- + t_2^+ \leq 1, t_3^- + t_3^+ \leq 1, t_2^- \geq 0, t_2^+ \geq 0, t_3^- \geq 0, t_3^+ \geq 0.$$

Она совместна: подходят, например, значения $t_2^- = 0,9; t_2^+ = 0, t_3^- = 0, t_3^+ = 0$. Поэтому x^2 является потенциально доминирующей над x^3 . Таким образом, $N(3) = \{1, 2\}$

Теперь можно записать следующую задачу линейного программирования с использованием ограничений (2.8) и (2.12):

$$t \rightarrow \max \tag{2.23}$$

при ограничениях:

$$t \leq t_2^-, t \leq t_2^+, t \leq t_3^-, t \leq t_3^+; \tag{2.24}$$

$$t_2^- + t_2^+ \leq 1, t_3^- + t_3^+ \leq 1; \tag{2.25}$$

$$12,6 \cdot t_2^- + 1,4 \cdot t_3^+ \leq 1,6; 1,8 \cdot t_2^+ + 1,4 \cdot t_3^- \leq 2,4 \tag{2.26}$$

$$12,6 \cdot t_2^+ + 1,4 \cdot t_3^- \leq 12,4; 14,4 \cdot t_2^+ + 2,8 \cdot t_3^- \leq 8,8; \quad (2.27)$$

$$1,8 \cdot t_2^- + 1,4 \cdot t_3^+ \leq 0,8; 14,4 \cdot t_2^- + 2,8 \cdot t_3^+ \leq 8,4; \quad (2.28)$$

$$t_2^- \geq 0, t_2^+ \geq 0, t_3^- \geq 0, t_3^+ \geq 0. \quad (2.29)$$

Решение задачи (2.23) – (2.29): $t = 0,1142857$; $t_2^- = 0,1142857$; $t_2^+ = 0,1142857$; $t_3^- = 0,1142857$; $t_3^+ = 0,1142857$. Все нестрогие неравенства в (2.24) выполняются как равенства, двойственные переменные для первого и четвертого из них положительны, а для второго и третьего равны нулю. Поэтому фиксируем значения величин $t_2^- = 0,1142857$, $t_3^+ = 0,1142857$ и убираем содержащие их неравенства из (2.29). Получаем вторую задачу:

$$t \rightarrow \max \quad (2.30)$$

при ограничениях:

$$t \leq t_2^+, t \leq t_3^-; \quad (2.31)$$

$$t_2^- + t_2^+ \leq 1, t_3^- + t_3^+ \leq 1; \quad (2.32)$$

$$3,6 \cdot t_2^+ + 0,7 \cdot t_3^- \leq 2,7; 10,8 \cdot t_2^+ + 0,7 \cdot t_3^+ \leq 7; \quad (2.33)$$

$$3,6 \cdot t_2^- + 0,7 \cdot t_3^+ \leq 1,6; 7,2 \cdot t_2^+ + 1,4 \cdot t_3^+ \leq 5; \quad (2.34)$$

$$10,8 \cdot t_2^- + 0,7 \cdot t_3^- \leq 4,5; 7,2 \cdot t_2^- + 1,4 \cdot t_3^- \leq 3,6; \quad (2.35)$$

$$t_2^- = 0,1142857, t_2^+ \geq 0, t_3^- \geq 0, t_3^+ = 0,1142857. \quad (2.36)$$

Решение задачи (2.31) – (2.36): $t = 0,5116279$; $t_2^- = 0,1142857$; $t_2^+ = 0,5116279$; $t_3^- = 0,5116279$; $t_3^+ = 0,1142857$. Оба нестрогих неравенства в (2.31) выполняются как равенства, двойственные переменные для них положительны. Решение SL-задачи получено.

Таким образом, множество недоминируемых альтернатив остается неизменным, если границы ИНЗ изменяются в таких пределах (с точностью до двух знаков после запятой; округление для λ_{21}^- и λ_{31}^- в меньшую сторону, а для λ_{21}^+ и λ_{31}^+ – в большую сторону):

$$0 < \lambda_{21}^- \leq 0,40 \approx 0,2 + 1,8 \cdot 0,1142857; \quad 2 - 1,8 \cdot 0,5116279 \approx 1,08 \leq \lambda_{21}^+ < +\infty;$$

$$0 < \lambda_{31}^- \leq 0,65 \approx 0,3 + 0,7 \cdot 0,5116279; \quad 1 - 0,7 \cdot 0,1142857 \approx 0,93 \leq \lambda_{31}^+ < +\infty.$$

Сравнение этих результатов с полученными выше при анализе чувствительности одной альтернативы x^1 показывает, что интервалы допустимых изменений границ λ_{21}^- и λ_{31}^+ остались прежними, но интервалы допустимых изменений границ λ_{21}^+ и λ_{31}^- сузились.

2.1.5 Заключение по подразделу

Информация о предпочтениях ЛПР в виде интервалов неопределенности замещений (ИНЗ) критериев порождает на множестве альтернатив отношение предпочтения, являющее строгим частичным порядком. Решение задачи многокритериального выбора с использованием ИНЗ представляется множеством альтернатив, недоминируемых по этому отношению: наилучшая альтернатива должна быть выделена именно из этого множества. Анализ чувствительности указанного множества к изменению ИНЗ является необходимым этапом решения прикладной задачи выбора.

Анализ чувствительности как отдельной недоминируемой альтернативы, так и множества недоминируемых альтернатив в целом, к изменению ИНЗ сводится к решению одной или нескольких последовательно формируемых задач линейного программирования. Поэтому его нетрудно осуществить при решении практических задач многокритериального выбора с использованием ИНЗ, особенно при использовании компьютерных систем поддержки принятия решений.

Поставленные перед началом исследований задачи выполнены полностью. Результаты проведенных исследований нашли отражение в [106, 107, 108].

2.2 Эмпирическое исследование устойчивости агрегированных рейтингов

В настоящее время для объективной оценки качества научных журналов, количество которых непрерывно растет, используются различные библиометрические показатели, такие как импакт-фактор, индексы SNIP, SJR и другие. На основании расчета значений этих индексов строятся рейтинги, отражающие сравнительную значимость журнала как средства внутринаучной коммуникации. Однако множественность показателей приводит к несовпадению оценок влияния журналов, сделанных на их основе.

На предыдущем этапе данного исследования на основании массива эмпирических данных о 212 международных научных журналах по экономике, 93 научных журналах по менеджменту и 99 научных журналах по политологии была дана количественная оценка степени (не) согласованности различных библиометрических показателей, построены рейтинги журналов, агрегирующие информацию об их сравнительной значимости, которую дают ранжирования по отдельным показателям, и эмпирически решена задача и выборе наилучшего представителя, являются ли построенные рейтинги более эффективными инструментами оценки влияния журналов, чем ранжирования по значению одного индекса.

Используется новый подход к построению агрегированных рейтингов, основанный на использовании ординальных методов ранжирования, впервые предложенных в теории коллективного выбора, а затем примененных в задачах многокритериального оценивания. Различными критериями, по которым оцениваются журналы, в данном случае, являются такие библиометрические показатели, как двух- и пятилетний импакт-факторы, индекс оперативности, индекс влияния статьи, индекс Хирша, индексы SNIP и SJR. Критериальные оценки агрегируются на основании мажоритарного правила (правила большинства). Результатом агрегирования становится бинарное отношение, называемое мажоритарным. Это отношение содержит информацию о парных сравнениях журналов, то есть дает ответ на вопрос – какой из двух сравни-

ваемых журналов лучше по большинству критериев. С помощью мажоритарного отношения можно определить те журналы, которые следует считать наилучшими с точки зрения всей совокупности показателей. Для этого можно использовать либо основанные на мажоритарном отношении правила ранжирования, такие как правило Коупланда и марковский метод, либо сортировку с помощью турнирных решений - правил оптимального выбора, основанных на обобщении принципа выбора максимального (т.е. наилучшего) элемента отношения, например, с помощью отбора альтернатив, принадлежащих непокрытому множеству, или альтернатив, принадлежащих минимальному внешнеустойчивому множеству. Методы ранжирования подробно описаны в публикациях [109, 110].

Однако публикация результатов этой части исследования встретила критику части сообщества специалистов по наукометрии. Выражалось сомнение в теоретической возможности и практической целесообразности агрегирования рейтингов основанных на индикаторах различных типов – зависящих от объёма издания (size-dependent) и независящих от него (size-independent). Поскольку из семи индикаторов к первому типу относится только индекс Хирша, был предпринят пересчёт всех результатов для данного набора индикаторов без индекса Хирша на основе того же самого массива эмпирических данных. Также было проведено сравнение новых результатов со старыми.

Результаты опубликованы в [111, 110].

Поскольку номер позиции является ранговой переменной, для оценки согласованности двух различных рейтингов журналов нужно использовать ранговые меры корреляции. В настоящей работе, как и на предыдущем этапе, в качестве способов количественной оценки степени различия двух ранжирований используются схожие, но не идентичные величины: основанные на расстоянии Кендалла коэффициент ранговой корреляции Кендалла и доля совпадающих парных сравнений r .

Рассмотрим пару журналов и сравним их позиции в двух рейтингах. Если в одном рейтинге позиция первого журнала выше позиции второго журна-

ла, а в другом рейтинге – ниже, то имеет место инверсия. Расстояние Кендалла между двумя ранжированиями – это число инверсий N_- (число неупорядоченных пар, ранжированных противоположным образом), соответственно, чем больше между рейтингами инверсий, тем дальше они друг от друга. Коэффициент ранговой корреляции Кендалла τ_b зависит от расстояния Кендалла следующим образом $\tau_b = \frac{N_+ - N_-}{\sqrt{(N - n_1) \cdot (N - n_2)}}$. Здесь N_+ – это число пар, в которых позиция одного из журналов выше позиции другого журнала как в одном рейтинге, так и в другом, n_1 – число таких пар, в которых оба журнала имеют одинаковый ранг в первом рейтинге, а n_2 , соответственно, – это число таких пар во втором рейтинге. Результаты вычислений τ_b приведены в Таблице 2.1.

Таблица 2.1 – Коэффициент Кендалла τ_b

	импакт-фактор	5-л. импакт-фактор	и. оперативн. ости	и. влияния	SNIP	SJR	Коупланд (2)	Коупланд (3)	UC	MES	Марков
Экономика											
импакт-фактор	1,000	0,830	0,503	0,637	0,698	0,700	0,832	0,840	0,845	0,852	0,829
5-л. имп.-фактор	0,830	1,000	0,510	0,725	0,726	0,741	0,902	0,900	0,905	0,891	0,884
и. операт-ти	0,503	0,510	1,000	0,475	0,454	0,472	0,559	0,562	0,572	0,569	0,560
и. влияния	0,637	0,725	0,475	1,000	0,673	0,674	0,768	0,764	0,771	0,761	0,760
SNIP	0,698	0,726	0,454	0,673	1,000	0,638	0,773	0,773	0,780	0,779	0,771
SJR	0,700	0,741	0,472	0,674	0,638	1,000	0,785	0,782	0,790	0,785	0,775
Менеджмент											
импакт-фактор	1,000	0,790	0,520	0,641	0,679	0,626	0,809	0,800	0,808	0,799	0,797
5-л. имп.-фактор	0,790	1,000	0,475	0,743	0,798	0,702	0,904	0,897	0,901	0,904	0,881
и. операт-ти	0,520	0,475	1,000	0,456	0,399	0,391	0,510	0,512	0,516	0,511	0,511
и. влияния	0,641	0,743	0,456	1,000	0,695	0,728	0,803	0,809	0,815	0,811	0,799
SNIP	0,679	0,798	0,399	0,695	1,000	0,719	0,824	0,824	0,836	0,843	0,806
SJR	0,626	0,702	0,391	0,728	0,719	1,000	0,767	0,778	0,789	0,791	0,766
Политология											
импакт-фактор	1,000	0,773	0,422	0,671	0,653	0,673	0,812	0,806	0,810	0,815	0,795
5-л. имп.-фактор	0,773	1,000	0,374	0,835	0,705	0,717	0,896	0,901	0,905	0,896	0,871
инд. опер-ти	0,422	0,374	1,000	0,356	0,372	0,398	0,440	0,439	0,446	0,451	0,451
инд. влияния	0,671	0,835	0,356	1,000	0,671	0,653	0,820	0,823	0,829	0,830	0,789
SNIP	0,653	0,705	0,372	0,671	1,000	0,662	0,753	0,766	0,768	0,762	0,763
SJR	0,673	0,717	0,398	0,653	0,662	1,000	0,784	0,770	0,781	0,769	0,763

Если же позиция одного из журналов выше позиции другого журнала как в одном рейтинге, так и в другом, или если номера позиций обоих журналов совпадают как в одном рейтинге, так и в другом, то можно сказать, что данные рейтинги совпадают в оценке данной пары. Соответственно, в качестве меры согласия рейтингов можно использовать выраженную в процентах

долю (неупорядоченных) пар журналов, в оценке которых рейтинги совпадают друг с другом, $r=100 \cdot \frac{N_++N_0}{N}$, где N_0 – число пар, в которых номера позиций обоих журналов совпадают как в одном рейтинге, так и в другом, N – общее число неупорядоченных пар. Результаты вычислений r приведены в Таблице 2.2.

Таблица 2.2 – Доля совпадающих парных сравнений r (в %)

	импакт-фактор	5-л. импакт-фактор	и. оперативности	и. влияния	SNIP	SJR	Коупланд (2)	Коупланд (3)	UC	MES	Марков
Экономика											
импакт-фактор	100,00	91,46	74,70	81,77	84,80	83,38	91,23	91,64	89,75	89,77	91,37
5-л. имп.-фактор	91,46	100,00	75,08	86,22	86,26	85,45	94,74	94,65	92,69	91,70	94,15
и. оперативности	74,70	75,08	100,00	73,31	72,28	71,79	77,27	77,41	75,98	75,53	77,58
и. влияния	81,77	86,22	73,31	100,00	83,60	82,12	88,07	87,86	86,10	85,35	87,95
SNIP	84,80	86,26	72,28	83,60	100,00	80,32	88,31	88,33	86,54	86,19	88,50
SJR	83,38	85,45	71,79	82,12	80,32	100,00	87,33	87,24	85,72	85,33	87,11
Менеджмент											
импакт-фактор	100,00	89,43	75,83	81,95	83,87	79,64	89,60	89,32	87,70	86,84	89,76
5-л. имп.-фактор	89,43	100,00	73,59	87,10	89,86	83,43	94,37	94,20	92,26	91,98	94,04
и. оперативности	75,83	73,59	100,00	72,63	69,78	68,00	74,64	74,92	73,33	72,72	75,39
и. влияния	81,95	87,10	72,63	100,00	84,69	84,71	89,34	89,78	88,03	87,40	89,90
SNIP	83,87	89,86	69,78	84,69	100,00	84,27	90,39	90,58	89,08	89,01	90,25
SJR	79,64	83,43	68,00	84,71	84,27	100,00	86,07	86,70	85,58	85,13	86,61
Политология											
импакт-фактор	100,00	88,56	69,53	83,45	82,58	80,09	89,88	89,61	88,27	87,32	89,67
5-л. имп.-фактор	88,56	100,00	67,20	91,67	85,20	82,25	94,10	94,41	93,03	91,30	93,53
и. оперативности	69,53	67,20	100,00	66,27	67,12	65,70	69,96	69,96	69,10	68,19	71,02
и. влияния	83,45	91,67	66,27	100,00	83,47	79,12	90,27	90,48	89,26	88,08	89,36
SNIP	82,58	85,20	67,12	83,47	100,00	79,57	86,97	87,69	86,27	84,75	88,08
SJR	80,09	82,25	65,70	79,12	79,57	100,00	84,97	84,39	83,67	82,40	84,48

На основании значений мер согласованности ранжирований можно определить ранжирования, которые являются наилучшими представителями данной совокупности рейтингов (в данном случае семи исходных ранжирований журналов по значению соответствующего библиометрического индекса). Для этого воспользуемся техникой парных сравнений с помощью правила большинства. Будем считать, что рейтинг А лучше представляет данную совокупность рейтингов, чем рейтинг В, если он лучше коррелирует с большинством рейтингов совокупности, то есть больше похож на большинство рейтингов совокупности, которую нужно представить, то есть агрегировать. В итоге, на множестве рейтингов возникает мажоритарное отношение (свое

для своей меры схожести рейтингов). Подробно данный механизм построения мажоритарного отношения на множестве ранжирований описан в [109]. Турнирные матрицы этих отношений для случая шести библиометрических показателей приведены в Таблицах 2.3 и 2.4.

Таблица 2.3 – Турнирная матрица мажоритарного отношения на множестве ранжирований при оценке схожести с помощью коэффициента Кендалла τ_b

	импакт-фактор	5-л. импакт-фактор	и. оперативности	и. влияния	SNIP	SJR	Коупланд (2)	Коупланд (3)	UC	MES	Марков
Экономика											
импакт-фактор	0	0	1	1	1	1	0	0	0	0	0
5-л. имп.-фактор	1	0	1	1	1	1	0	0	0	0	0
и. оперативности	0	0	0	0	0	0	0	0	0	0	0
и. влияния	0	0	1	0	0	0	0	0	0	0	0
SNIP	0	0	1	0	0	0	0	0	0	0	0
SJR	0	0	1	0	1	0	0	0	0	0	0
Коупланд - 2	1	1	1	1	1	1	0	0	0	0	1
Коупланд - 3	1	1	1	1	1	1	0	0	0	0	1
непокр. множ. UC	1	1	1	1	1	1	1	1	0	1	1
мин. внешнеуст. множ. MES	1	1	1	1	1	1	1	1	0	0	1
марковское ранжирование	1	1	1	1	1	1	0	0	0	0	0
Менеджмент											
импакт-фактор	0	0	1	0	0	0	0	0	0	0	0
5-л. имп.-фактор	1	0	1	1	1	1	0	0	0	0	0
и. оперативности	0	0	0	0	0	0	0	0	0	0	0
и. влияния	0	0	1	0	0	1	0	0	0	0	0
SNIP	1	0	1	0	0	1	0	0	0	0	0
SJR	0	0	1	0	0	0	0	0	0	0	0
Коупланд - 2	1	1	1	1	1	1	0	0	0	0	1
Коупланд - 3	1	1	1	1	1	1	1	0	0	0	1
непокрытое множество UC	1	1	1	1	1	1	1	1	0	0	1
мин. внешнеуст. множ. MES	1	1	1	1	1	1	1	1	0	0	1
марковское ранжирование	1	1	1	1	1	1	0	0	0	0	0
Политология											
импакт-фактор	0	0	1	0	1	1	0	0	0	0	0
5-л. имп.-фактор	1	0	1	1	1	1	0	0	0	0	0
и. оперативности	0	0	0	0	0	0	0	0	0	0	0
и. влияния	0	0	1	0	0	0	0	0	0	0	0
SNIP	0	0	1	0	0	0	0	0	0	0	0
SJR	0	0	1	0	1	0	0	0	0	0	0
Коупланд - 2	1	1	1	1	1	1	0	0	0	0	1
Коупланд - 3	1	1	1	1	1	1	0	0	0	0	1
непокрытое множество UC	1	1	1	1	1	1	1	1	0	0	1
мин. внешнеуст. множ. MES	1	1	1	1	1	1	1	0	0	0	1
марковское ранжирование	1	1	1	1	1	1	0	0	0	0	0

Таблица 2.4 – Турнирная матрица мажоритарного отношения на множестве ранжирований при оценке похожести с помощью доли совпадающих парных сравнений r

	импакт-фактор	5-л. импакт-фактор	и. оперативности	и. влияния	SNIP	SJR	Коупланд (2)	Коупланд (3)	UC	MES	Марков
Экономика											
импакт-фактор	0	0	1	1	1	1	0	0	0	0	0
5-л. имп.-фактор	1	0	1	1	1	1	0	0	0	1	0
и. оперативности	0	0	0	0	0	0	0	0	0	0	0
и. влияния	0	0	1	0	0	1	0	0	0	0	0
SNIP	0	0	1	0	0	1	0	0	0	0	0
SJR	0	0	1	0	0	0	0	0	0	0	0
Коупланд – 2	1	1	1	1	1	1	0	0	1	1	0
Коупланд - 3	1	1	1	1	1	1	0	0	1	1	0
непокрытое множество UC	1	0	1	1	1	1	0	0	0	1	0
мин. внешнеуст. множ. MES	1	0	1	1	1	1	0	0	0	0	0
марковское ранжирование	1	1	1	1	1	1	0	0	1	1	0
Менеджмент											
импакт-фактор	0	0	1	0	0	0	0	0	0	0	0
5-л. имп.-фактор	1	0	1	1	1	1	0	0	1	1	0
и. оперативности	0	0	0	0	0	0	0	0	0	0	0
и. влияния	0	0	1	0	0	1	0	0	0	0	0
SNIP	1	0	1	0	0	1	0	0	0	0	0
SJR	0	0	1	0	0	0	0	0	0	0	0
Коупланд – 2	1	1	1	1	1	1	0	0	1	1	0
Коупланд - 3	1	1	1	1	1	1	1	0	1	1	0
непокрытое множество UC	1	0	1	1	1	1	0	0	0	1	0
мин. внешнеуст. множ. MES	1	0	1	1	1	1	0	0	0	0	0
марковское ранжирование	1	1	1	1	1	1	1	0	1	1	0
Политология											
импакт-фактор	0	0	1	0	1	1	0	0	0	0	0
5-л. имп.-фактор	1	0	1	1	1	1	0	0	0	1	0
и. оперативности	0	0	0	0	0	0	0	0	0	0	0
и. влияния	0	0	1	0	0	1	0	0	0	0	0
SNIP	0	0	1	0	0	1	0	0	0	0	0
SJR	0	0	1	0	0	0	0	0	0	0	0
Коупланд – 2	1	1	1	1	1	1	0	0	1	1	1
Коупланд - 3	1	1	1	1	1	1	1	0	1	1	0
непокрытое множество UC	1	0	1	1	1	1	0	0	0	1	0
мин. внешнеуст. множ. MES	1	0	1	1	1	1	0	0	0	0	0
марковское ранжирование	1	1	1	1	1	1	0	1	1	1	0

Применяя к этим отношениям правило Коупланда, получаем ранжирования исходных и агрегированных рейтингов по степени их «представительности». Результаты вычислений для двух мер и трех совокупностей их значений представлены в Таблице 2.5. Рейтинги упорядочены от наиболее согласованного (ранг 1) до наименее согласованного (ранг 11). Для сравнения в Таблице 2.6 приведены ранжирования ранжирований, полученные на преды-

дущем этапе исследования при агрегировании семи библиометрических показателей.

Таблица 2.5 – Ранжирование ранжирований (без индекса Хирша)

ранг	Рейтинги сравниваются на основании τ_b			Рейтинги сравниваются на основании r		
	экономика	менеджмент	политология	экономика	менеджмент	политология
1	<i>UC</i>	<i>UC / MES</i>	<i>UC</i>	Copeland2 /	Copeland3 /	Copeland2 /
2	<i>MES</i>		<i>MES</i>	Copeland3 /	Markov	Copeland3 /
3	Copeland2 /	Copeland 3	Copeland2 /	Markov	Copeland 2	Markov
4	Copeland3	Copeland 2	Copeland3	IF-5 / <i>UC</i>	IF-5	IF-5 / <i>UC</i>
5	Markov	Markov	Markov		<i>UC</i>	
6	5-y. impact	5-y. impact	5-y. impact	<i>MES</i>	<i>MES</i>	<i>MES</i>
7	Impact	SNIP	Impact	IF	SNIP	IF
8	SJR	AI	SJR	AI / SNIP	AI	AI / SNIP
9	AI / SNIP	Impact / SJR	AI / SNIP		IF / SJR	
10				SJR	SJR	
11	Immediacy	Immediacy	Immediacy	II	II	II

Таблица 2.6 – Ранжирование ранжирований (с индексом Хирша)

ранг	Рейтинги сравниваются на основании τ_b			Рейтинги сравниваются на основании r		
	экономика	менеджмент	политология	экономика	менеджмент	политология
1	<i>MES</i>	<i>MES</i>	<i>MES</i>	Copeland 3	Copeland 3	Copeland 3/
2	<i>UC</i>	<i>UC</i>	<i>UC</i>	Copeland 2	Copeland 2	Copeland 2/
3	Copeland 3	Copeland 2	Copeland 3	Markov	Markov	Markov
4	Copeland 2	Copeland 3	Copeland 2	<i>UC</i>	<i>UC</i>	<i>UC</i>
5	Markov	Markov	Markov	5-y. impact	5-y. impact/ <i>MES</i>	5-y. impact/ <i>MES</i>
6	5-y. impact	5-y. impact	5-y. impact	impact/ <i>MES</i>		
7	impact	SNIP	Hirsch		SNIP	AI
8	SJR	Hirsch	AI/impact/SJR	AI	AI	impact
9	AI	AI		SNIP	impact/Hirsch/SJR	SNIP
10	SNIP	SJR		SJR		SJR
11	Hirsch	impact		SNIP		Hirsch
12	immediacy	immediacy	immediacy	immediacy	immediacy	immediacy

Сравнительный анализ представленных в Таблицах 2.5 и 2.6 ранжирований показывает, что для трёх использовавшихся массивов данных (для трёх совокупностей научных журналов – по экономике, менеджменту и политологии) исключение индекса Хирша не приводит к изменению репрезентативности ранжирований, получаемых с помощью вышеперечисленных методов агрегирования. Величина коэффициента Кендалла для любого из пяти агрегированных ранжирований при их сопоставлении с ранжированиями по библиометрическим показателям всегда (а значение r почти всегда) превосходит значения, получающиеся при сравнении последних между собой, то есть переход от исходных рейтингов к агрегированным действительно является вы-

игрышным. Иначе говоря, агрегированные ранжирования лучше представляют совокупность рейтингов, основанных на семи выбранных библиометрических показателях, чем любой из исходных рейтингов, и могут служить в качестве интегральных рейтингов журналов. Расхождения в рейтингах, строящихся по разным мерам корреляции, как и «провал» агрегированных рейтингов, основанных на сортировке с помощью непокрытого множества и минимального внешнеустойчивого множества, можно объяснить тем, что мера r принимает более низкие значения у рейтингов с более грубой шкалой, то есть рейтингов, представляющих из себя объединение журналов в более крупные группы. Однако мы полагаем, что такое более «грубое» разбиение лучше соответствует интуитивным представлениям о различии в значимости журналов.

2.3 Исследование мер конфликта в рамках теории функций доверия, индуцированных правилами комбинирования свидетельств

2.3.1 Введение и постановка задачи

При принятии решения о комбинировании свидетельств в рамках теории свидетельств ([112, 113]) следует принимать во внимание различные факторы, важнейшим из которых является мера конфликта между комбинируемыми телами свидетельств. В последнее время исследование мер конфликта все больше оформляется в самостоятельное направление в рамках теории свидетельств. Так в [114, 115, 116] обсуждалась аксиоматика мер конфликта, в [117, 118, 119, 114, 120] рассматривались различные подходы к оцениванию конфликта свидетельств. В [121, 122, 123] исследовалось понятие внутреннего конфликта свидетельств.

Выбор той или иной меры для оценивания конфликта зависит от решаемой задачи. Так, если мы оцениваем конфликт между телами свидетельств с целью принятия решения о комбинировании свидетельств, то мера конфликта должна быть определенным образом согласована с правилом комбиниро-

вания свидетельств. Так, правило комбинирования Демпстера естественно согласуется с конъюнктивной мерой конфликта. Для других правил комбинирования условия согласования не так очевидны. В данной работе исследуются меры конфликта, индуцированные дизъюнктивными правилами комбинирования. Исследуется вопрос о связи условий согласования с аксиоматикой мер конфликта.

Кроме того, в работе исследуется структура конъюнктивных и дизъюнктивных мер конфликта. В частности, показано, что в таких мерах можно выделить метрическую и энтропийную части. Причем эти составляющие по-разному меняются при комбинировании тел свидетельств.

Основная цель статьи – исследовать некоторые факторы (выбор меры конфликта, условий согласования с правилом комбинирования, учет энтропии свидетельств и т.д.), которые следует учитывать (наряду, например, с мерами неопределенности свидетельств [124], изменением незнания после комбинирования свидетельств [120]) при принятии решения о комбинировании свидетельств.

Структура этого подраздела отчета следующая. В пункте 2.3.2 приведены основные понятия теории свидетельств. В пункте 2.3.3 рассмотрена аксиоматика мер конфликта. В пункте 2.3.4 рассмотрены меры конфликта, индуцированные дизъюнктивными правилами комбинирования. В пункте 2.3.5 введено понятия мер конфликта, согласованных с правилами комбинирования. Обсуждается, как условия согласования связаны с аксиоматикой мер конфликта. В пункте 2.3.6 показано, что в конъюнктивных и дизъюнктивных мерах конфликта можно выделить метрическую и энтропийную составляющие. А в пункте 2.3.7 обсуждаются, как меняются эти составляющие при комбинировании тел свидетельств. Наконец, в пункте 2.3.8 приведены некоторые выводы из работы.

2.3.2 Основные понятия теории функций доверия

Напомним основные определения из теории функций доверия (теории свидетельств, теории Демпстера-Шейфера) [112, 113]. Пусть X – конечное множество, 2^X – множество всех подмножеств из X . Функцией масс (или основным доверительным назначением, basic belief assignment, BBA) называется функция множеств $m: 2^X \rightarrow [0,1]$, удовлетворяющая условию нормировки $\sum_{A \subseteq X} m(A) = 1$. Функция масс m называется нормальной, если $m(\emptyset) = 0$. Она называется субнормальной, если $m(\emptyset) > 0$. Значение $m(A)$ характеризует относительную степень уверенности того, что актуальная альтернатива из X принадлежит множеству $A \in 2^X$.

Подмножество $A \in 2^X$ называется фокальным элементом, если $m(A) > 0$. Пусть $\mathcal{A} = \{A\}$ есть множество всех фокальных элементов свидетельства. Тогда пару $F = (\mathcal{A}, m)$ называют телом свидетельства.

Пусть $F_A = (A, 1)$ (т.е. $\mathcal{A} = \{A\}$ and $m(A) = 1$), $A \in 2^X$ и $\mathcal{F}(X)$ есть множество всех тел свидетельств на X . Тело свидетельств $F_A = (A, 1)$ называется категориальным. В частности, тело $F_X = (X, 1)$ называют бессодержательным (vacuous), поскольку оно абсолютно неинформативно.

Предположим, что мы имеем два тела свидетельств $F_1 = (\mathcal{A}_1, m_1)$ и $F_2 = (\mathcal{A}_2, m_2)$, которые получены из двух разных источников информации. В теории свидетельств рассматриваются различные правила комбинирования свидетельств: $R: \mathcal{F}(X) \times \mathcal{F}(X) \rightarrow \mathcal{F}(X)$. Например, рассматривается не-нормализованное конъюнктивное правило $D_0(F_1, F_2)$ [125] $m^{D_0}(A) = \sum_{B \cap C = A} m_1(B)m_2(C)$, $A \in 2^X$. Значение $K^D(F_1, F_2) = m^{D_0}(\emptyset)$ характеризует величину конфликта между двумя источниками информации, описываемыми телами свидетельств F_1 и F_2 . Назовем величину $K^D(F_1, F_2) = m^{D_0}(\emptyset)$ конъюнктивной мерой конфликта. Если $K^D \neq 1$, то определено классическое

правило Демпстера комбинирования двух свидетельств:

$$m^D(A) = \frac{1}{1 - K^D} \sum_{B \cap C = A} m_1(B)m_2(C), \quad A \neq \emptyset, \quad m^D(\emptyset) = 0.$$

Дизъюнктивное правило консенсуса Дюбуа и Прада (Dubois and Prade's disjunctive consensus rule) является двойственным к правилу Демпстера в некотором смысле. Оно определяется по формуле [126]: $m^{DP}(A) = \sum_{B \cup C = A} m_1(B)m_2(C)$, $A \in 2^X$. Некоторые общие схемы комбинирования были рассмотрены, в частности, в [127, 128].

Свидетельство $\bar{F} = (\bar{\mathcal{A}}, \bar{m})$ называют отрицанием (или дополнением) свидетельства $F = (\mathcal{A}, m)$, если $\bar{\mathcal{A}} = \{\bar{A} : A \in \mathcal{A}\}$ и $\bar{m}(A) = m(\bar{A})$ для всех $A \in \bar{\mathcal{A}}$, где \bar{A} – дополнение множества A [129]. Заметим, что если $F = \sum_{A \in \mathcal{A}} m(A)F_A$, то $\bar{F} = \sum_{A \in \mathcal{A}} m(A)F_{\bar{A}}$. Для конъюнктивного (ненормализованного) и дизъюнктивного правил комбинирования справедливо соотношение двойственности (аналог правила де Моргана) [129]:

$$\overline{D_0(F_1, F_2)} = DP(\bar{F}_1, \bar{F}_2). \quad (2.37)$$

Вместе с отрицанием (или дополнением) $\bar{F} = (\bar{\mathcal{A}}, \bar{m})$ свидетельства $F = (\mathcal{A}, m)$ будем рассматривать двойственное к $F = (\mathcal{A}, m)$ свидетельство $F^{(-)} = (\bar{\mathcal{A}}, m^{(-)})$, где $m^{(-)}(\bar{A}) = \frac{1}{N-1}(1 - m(A))$, $\forall A \in \mathcal{A}$, $N = |\mathcal{A}| > 1$.

2.3.3 Аксиоматика мер конфликта

В общем случае, мера конфликта $K(F_1, F_2)$ должна удовлетворять следующим условиям (аксиомам) [114, 115]:

- A1: $0 \leq K(F_1, F_2) \leq 1$ для всех $F_1, F_2 \in \mathcal{F}(X)$ (неотрицательность и нормализация);
- A2: $K(F_1, F_2) = K(F_2, F_1)$ для всех $F_1, F_2 \in \mathcal{F}(X)$ (симметричность);
- A3: $K(F, F) = 0$ для всех $F \in \mathcal{F}(X)$ (нильпотентность);

A4: $K(F', F) \geq K(F'', F)$, если $F' = (\mathcal{A}', m)$, $F'' = (\mathcal{A}'', m)$, где $\mathcal{A}' = \{A'_i\}$, $\mathcal{A}'' = \{A''_i\}$ и $A'_i \subseteq A''_i$ для всех i и $F \in \mathcal{F}(X)$ (антимонотонность относительно неточности свидетельств);

A5: $K(F_X, F) = 0$ для всех $F \in \mathcal{F}(X)$ (блаженство в незнании [115]);

A6: $K(F_A, F_B) = 1$, если $A \cap B = \emptyset$.

Кроме того, если считать, что пустое множество может быть фокальным элементом (значение $m(\emptyset)$ можно интерпретировать как степень доверия к тому, что истинная альтернатива $x \notin X$), то будем рассматривать и следующую аксиому:

A7: $K(F_\emptyset, F) = 1$ для всех $F \in \mathcal{F}(X) \setminus \{F_\emptyset\}$.

Но в этом случае будем считать, что аксиомы A3 и A5 выполняются для всех $F \in \mathcal{F}(X) \setminus \{F_\emptyset\}$.

Также рассматриваются и другие аксиомы (см., например, [115]). Заметим, что конъюнктивная мера конфликта K^D не удовлетворяет аксиоме A3.

2.3.4 Меры конфликта индуцированные конъюнктивными и дизъюнктивными правилами

Предположим, что информация от двух источников описывается с помощью двух тел свидетельств $F_1 = (\mathcal{A}_1, m_1)$ и $F_2 = (\mathcal{A}_2, m_2)$. Тогда меру конфликта $K^D(F_1, F_2)$ можно рассматривать как меру конфликта, индуцированную конъюнктивным правилом. Эта мера удовлетворяет аксиомам меры конфликта A1, A2, A4-A7. Заметим, что в случае, когда тела свидетельств определены на множестве непересекающихся фокальных элементов (так называемые непересекающаяся структура доверия, disjoint belief structure), конъюнктивная мера конфликта $K^D(F_1, F_2) = 1 - \sum_B m_1(B)m_2(B)$, где $F_1 = (\mathcal{A}_1, m_1)$ и $F_2 = (\mathcal{A}_2, m_2)$ и $A' \cap A'' = \emptyset \ \forall A', A'' \in \mathcal{A} \ (A' \neq A'')$. Кроме того, в этом случае мера K^D будет «близкой» к мере

$K^H(F_1, F_2) = 1 - \sum_B \sqrt{m_1(B)m_2(B)} = \left(\frac{1}{\sqrt{2}} \left\| \sqrt{\mathbf{m}_1} - \sqrt{\mathbf{m}_2} \right\| \right)^2$, где $\sqrt{\mathbf{m}_i} = (\sqrt{m_i(A)})_{A \in \mathcal{A}}$, $i = 1, 2$, $\|\mathbf{x}\| := \sqrt{\mathbf{x}^T \mathbf{x}}$ – евклидова норма. Здесь $\frac{1}{\sqrt{2}} \left\| \sqrt{\mathbf{m}_1} - \sqrt{\mathbf{m}_2} \right\|$ – так называемое расстояние Хеллингера (Hellinger distance [130]) между вероятностными распределениями \mathbf{m}_1 и \mathbf{m}_2 . Другой «близкой» к K^D мерой на непересекающейся структуре доверия будет [116, 131] $K^M(F_1, F_2) = 1 - \sum_B \min\{m_1(B), m_2(B)\}$.

Можно рассмотреть различные меры конфликта, индуцированные дизъюнктивным правилом и которые удовлетворяют тем или иным (но не обязательно всем) аксиомам мер конфликта. Ниже будем рассматривать следующие меры конфликта, индуцированные дизъюнктивным правилом (будем называть их дизъюнктивными):

$$K_1^{DP}(F_1, F_2) = \sum_{B \cup C = X} m_1(B)m_2(C) = m^{DP}(X), \quad K_2^{DP}(F_1, F_2) = 1 - K_1^{DP}(F_1, F_2).$$

Заметим, что мера $K_1^{DP}(F_1, F_2)$ удовлетворяет только аксиомам A1, A2 и A7 (и частному случаю условия A6: $K(F_A, F_{\bar{A}}) = 1$), а мера $K_2^{DP}(F_1, F_2)$ – аксиомам A1, A2, A4, A5. Справедливо следующее соотношение между конъюнктивным и дизъюнктивным мерами конфликта, отражающих соотношение двойственности (2.37).

Утверждение 2.1. $K_1^{DP}(F_1, F_2) = K^D(\bar{F}_1, \bar{F}_2)$.

На непересекающейся структуре доверия конъюнктивные и дизъюнктивные меры конфликта связаны с помощью простых соотношений.

Утверждение 2.2. Если $F_1 = (\mathcal{A}, m_1)$ и $F_2 = (\mathcal{A}, m_2)$, где $A' \cap A'' = \emptyset$ $\forall A', A'' \in \mathcal{A}$ ($A' \neq A''$), то: 1) $K_1^{DP}(F_1, F_2^{(-)}) = \frac{1}{N-1} K^D(F_1, F_2)$; 2) $K_2^{DP}(F_1, \bar{F}_2) = K^D(F_1, F_2)$.

Связь между различными мерами конфликта можно описать с помощью «множеств уровня» тех тел свидетельств, для которых мера конфликта $K(F_1, F_2)$ постоянна: $U_c(K) = \{(F_1, F_2) : K(F_1, F_2) = c\}$, $c \in [0, 1]$.

Пример 2.1. Пусть $F_1 = \alpha_1 F_{\{x_1\}} + \alpha_2 F_{\{x_2\}} + (1 - \alpha_1 - \alpha_2) F_{\{x_1, x_2\}}$,
 $F_2 = \beta_1 F_{\{x_1\}} + \beta_2 F_{\{x_2\}} + (1 - \beta_1 - \beta_2) F_{\{x_1, x_2\}}$ – два свидетельства на $X = \{x_1, x_2\}$, где
 $\alpha = (\alpha_1, \alpha_2)$, $\beta = (\beta_1, \beta_2) \in V_2 = \{(x_1, x_2) : x_1 + x_2 \leq 1, x_1 \geq 0, x_2 \geq 0\}$. Тогда
 $K^D(F_1, F_2) = \alpha_1 \beta_2 + \alpha_2 \beta_1 = |[\alpha, \beta_-]|$, где $\beta_- = (\beta_1, -\beta_2)$ и $[\cdot, \cdot]$ – векторное произ-
ведение векторов. Аналогично, $K_1^{DP}(F_1, F_2) = 1 - \alpha_1 \beta_1 - \alpha_2 \beta_2 = 1 - (\alpha, \beta)$, где
 (\cdot, \cdot) – скалярное произведение векторов, и $K_2^{DP}(F_1, F_2) = (\alpha, \beta)$. Нетрудно ви-
деть, что в этом случае $K^D(F_1, F_2) + K_2^{DP}(F_1, F_2) = (\alpha_1 + \alpha_2)(\beta_1 + \beta_2) \leq 1$ и, сле-
довательно, $K^D(F_1, F_2) \leq K_1^{DP}(F_1, F_2)$. Поэтому верны вложения
 $U_1(K^D) \subseteq U_1(K_1^{DP})$ и $U_0(K_1^{DP}) \subseteq U_0(K^D)$.

В частности, $K^D(F_1, F_2) = 1 \Leftrightarrow \alpha = (1, 0), \beta = (0, 1)$ или наоборот. Другими
словами, максимальное значение конъюнктивной меры конфликта будет
у свидетельств, соответствующих различным сингулярным распределениям.
В то же время $K_1^{DP}(F_1, F_2) = 1 \Leftrightarrow (\alpha, \beta) = 0 \Leftrightarrow \alpha \perp \beta \Leftrightarrow \alpha = (\alpha, 0) \wedge \beta = (0, \beta)$ или
наоборот, где $\alpha, \beta \in [0, 1]$. Но, $K_2^{DP}(F_1, F_2) = (\alpha, \beta) = 1 \Leftrightarrow F_1, F_2$ – тела свиде-
тельств, соответствующие сингулярным одинаковым распределениям. Это
показывает, что большие значения $K_2^{DP}(F_1, F_2)$ не соответствует интуитив-
ным представлениям о большом конфликте между свидетельствами.

Утверждение 2.3. Если $F_1 = (\mathcal{A}_1, m_1)$, $F_2 = (\mathcal{A}_2, m_2)$ и $\mathcal{A}_1 = \bar{\mathcal{A}}_2$, то из
 $K^D(F_1, F_2) = 1$ следует, что $K_1^{DP}(F_1, F_2) = 1$.

2.3.5 Меры конфликта, согласованные с правилами комбинирования

Значение меры конфликта является важным фактором при принятии
решения о применении правила комбинирования для агрегирования инфор-
мации из нескольких источников. В этом случае мера конфликта выступает
как априорная характеристика применимости правила комбинирования.
Большая величина меры конфликта означает, что такие свидетельства не сле-

дует агрегировать (или следует осторожно относиться к результату комбинирования). В тоже время понятно, что в таких задачах выбор правила комбинирования и меры конфликта должен быть в определенной мере согласованным.

Будем рассматривать следующие условия согласования.

Определение 2.3. Правило комбинирования R и меру конфликта K назовем положительно согласованными, если $K(F_1, F_2) \leq K(R(F_1, F_2), F_i)$, $i = 1, 2$ для всех $F_1, F_2 \in \mathcal{F}(X)$. Если же выполняется противоположное неравенство, то такую пару R и K будем называть отрицательно согласованной.

Положительная (отрицательная) согласованность означает, что при комбинировании двух тел свидетельств мера конфликта результирующего свидетельства с любым операндом не уменьшится (не увеличится) по сравнению с мерой конфликта между операндами.

Утверждение 2.4. 1) Конъюнктивное (ненормализованное) правило комбинирования D_0 и мера конфликта K^D являются положительно согласованными;

2) дизъюнктивное правило комбинирования DP и мера конфликта K_1^{DP} являются положительно согласованными;

3) дизъюнктивное правило комбинирования DP и мера конфликта K_2^{DP} являются отрицательно согласованными.

Нетрудно видеть, что наличие той или иной согласованности мер конфликта с правилами комбинирования делает некоторые аксиомы зависимыми или противоречивыми. Например,

1) если мера конфликта положительно согласована с конъюнктивным правилом или дизъюнктивным правилом, то из аксиомы А3 следует аксиома А5, поскольку $K(F_x, F) \leq K(D(F_x, F), F) = K(F, F) = 0$ и $K(F_x, F) \leq K(DP(F_x, F), F_x) = K(F_x, F_x) = 0$;

2) если мера конфликта положительно согласована с конъюнктивным правилом, то из аксиомы А6 следует, что $K(F_\emptyset, F_A) = 1$ для любого

$A \neq \emptyset$ (частный случай аксиомы A7), поскольку

$$1 = K(F_A, F_{\bar{A}}) \leq K(D(F_A, F_{\bar{A}}), F_A) = K(F_{\emptyset}, F_A);$$

3) если мера конфликта положительно согласована с дизъюнктивным правилом, то аксиомы A5 и A6 противоречивы, также как аксиомы A3 и A7, поскольку $1 = K(F_A, F_{\bar{A}}) \leq K(DP(F_A, F_{\bar{A}}), F_A) = K(F_X, F_A) = 0$ и $1 = K(F_{\emptyset}, F) \leq K(DP(F_{\emptyset}, F), F) = K(F, F) = 0$;

4) если мера конфликта отрицательно согласована с конъюнктивным правилом, то из аксиомы A5 следует аксиома A3, поскольку $0 = K(F_X, F) \geq K(D(F_X, F), F) = K(F, F)$;

5) если мера конфликта отрицательно согласована с конъюнктивным правилом, то из аксиомы A7 следует, что $K(F_A, F_{\bar{A}}) = 1$ для любого $A \neq \emptyset$ (частный случай аксиомы A6), поскольку $K(F_A, F_{\bar{A}}) \geq K(D(F_A, F_{\bar{A}}), F_A) = K(F_{\emptyset}, F_A) = 1$.

2.3.6 Метрическая и энтропийная составляющие меры конфликта

При принятии решения о комбинировании свидетельств мы обращаем внимание не только на меру конфликта. Рассмотрим следующий пример.

Пример 2.2. Предположим, что есть три кандидата $X = \{x_1, x_2, x_3\}$ на некоторую должность и три эксперта высказали свои предпочтения по этим кандидатурам в виде трех тел свидетельств F_1 : $m_1(\{x_1\}) = m_1(\{x_2\}) = m_1(\{x_3\}) = 1/3$; F_2 : $m_2(\{x_1, x_2\}) = 1/3$, $m_2(\{x_3\}) = 2/3$; F_3 : $m_3(\{x_2\}) = 7/8$, $m_3(\{x_2, x_3\}) = 1/8$. Конъюнктивные меры конфликта будут равны $K^D(F_1, F_2) = 5/9$, $K^D(F_1, F_3) = 5/8$ и $K^D(F_2, F_3) = 7/12$, т.е. $K^D(F_1, F_2) < K^D(F_2, F_3) < K^D(F_1, F_3)$. Выберем для комбинирования пару свидетельств F_1 и F_2 с наименьшим конфликтом. После комбинирования по правилу Демпстера этих свидетельств, получим $m_{1,2}^D(\{x_1\}) = m_{1,2}^D(\{x_2\}) = 1/4$ и $m_{1,2}^D(\{x_3\}) = 1/2$, т.е. предпочтение в этом случае отдается третьему кандида-

ту. Вместе с тем, свидетельство F_1 малосодержательно, поскольку первый эксперт не отдал предпочтение ни одному из кандидатов. Если же найти комбинацию второго и третьего тел свидетельств, то получим $m_{2,3}^D(\{x_2\}) = 4/5$ и $m_{2,3}^D(\{x_3\}) = 1/5$, т.е. предпочтение отдается второму свидетельству. Аналогичная картина наблюдается и при использовании меры дизъюнктивного конфликта и комбинирования с помощью дизъюнктивного правила: $K_1^{DP}(F_1, F_2) = 1/9 < K_1^{DP}(F_1, F_3) = K_1^{DP}(F_2, F_3) = 2/9$; $m_{1,2}^{DP}(\{x_1, x_2\}) = m_{1,2}^{DP}(\{x_1, x_3\}) = m_{1,2}^{DP}(\{x_2, x_3\}) = m_{1,2}^{DP}(\{x_3\}) = 2/9$, $m_{1,2}^{DP}(\{x_1, x_2, x_3\}) = 1/9$; $m_{2,3}^{DP}(\{x_1, x_2\}) = 7/24$, $m_{2,3}^{DP}(\{x_2, x_3\}) = 2/3$, $m_{2,3}^{DP}(\{x_1, x_2, x_3\}) = 1/24$. При комбинировании первых двух свидетельств мы получим одинаковые значения функции масс для всех пар кандидатов. При комбинировании двух последних свидетельств, предпочтение будет отдано второму или третьему кандидатам. Объяснить этот пример можно тем, что первое тело свидетельство является равномерным вероятностным распределением. Оно имеет высокую энтропию Шеннона и его лучше не использовать для комбинирования.

Вместе с тем в самой мере конфликта можно выделить метрическую и энтропийную части.

Конъюнктивную меру конфликта двух тел свидетельств $F_1 = (\mathcal{A}, m_1)$ и $F_2 = (\mathcal{A}_2, m_2)$ можно записать следующим образом (для краткости будем опускать запись $B \in \mathcal{A}$, $C \in \mathcal{A}_2$, что не должно привести к недоразумениям, если считать $m_1(B) = 0$ и $m_2(C) = 0$ в случае $B \notin \mathcal{A}$, $C \notin \mathcal{A}_2$):

$$K^D(F_1, F_2) = \sum_{B \cap C = \emptyset} m_1(B)m_2(C) = 1 - \sum_{B \cap C \neq \emptyset} m_1(B)m_2(C) = \frac{1}{2} \left(2 - 2 \sum_{B, C} q_{B, C} m_1(B)m_2(C) \right) - \sum_{B, C} (t_{B, C} - q_{B, C}) m_1(B)m_2(C), \quad (2.38)$$

где $Q = (q_{B, C})$ – симметричная положительно определенная матрица, удовлетворяющая условиям: 1) $q_{B, C} \in [0, 1] \quad \forall B, C \in 2^X$; 2) $q_{B, C} = 0$, если $B \cap C = \emptyset$;

3) $q_{B,B} = 1 \quad \forall B \in 2^X; \quad T = (t_{B,C}), \quad t_{B,C} = \begin{cases} 1, & B \cap C \neq \emptyset, \\ 0, & B \cap C = \emptyset. \end{cases}$ Пусть $R = (r_{B,C})$,

$r_{B,C} = t_{B,C} - q_{B,C} = \begin{cases} 0, & B \cap C = \emptyset \vee B = C, \\ 1 - q_{B,C} & \text{иначе.} \end{cases}$ Примерами коэффициентов $q_{B,C}$ яв-

ляются индекс (подобность) Жаккарда (Jaccard index), $q_{B,C} = \frac{|B \cap C|}{|B \cup C|}$, если $B, C \neq \emptyset$ [130]. В вещественном арифметическом пространстве векторов

$A_{2^{|X|-1}}$ будем рассматривать скалярное произведение $(\mathbf{x}, \mathbf{y})_Q := \mathbf{x}^T Q \mathbf{y}$ и соответ-

ствующую норму $\|\mathbf{x}\|_Q := \sqrt{\mathbf{x}^T Q \mathbf{x}}$. В частности, если $q_{B,C}$ – индекс Жаккарда,

то соответствующая метрика $d_J(F_1, F_2) = \frac{1}{\sqrt{2}} \|\mathbf{m}_1 - \mathbf{m}_2\|_Q$, которую часто называют расстоянием Жоусельме (Jousselme distance), широко используется в теории свидетельств [117].

Рассмотрим на симплексе $S_X = \left\{ \mathbf{t} = (t_k)_{k=1}^{2^{|X|-1}} : t_k \in [0,1] \forall k, \sum_k t_k = 1 \right\}$ функционал $E_Q : S_X \rightarrow [0,1]$, $E_Q(F) = E_Q(\mathbf{m}) = 1 - \|\mathbf{m}\|_Q^2 = \sum_B m(B) \left(1 - \sum_C q_{B,C} m(C) \right)$, если $F = (\mathcal{A}, m)$. Этот функционал по некоторым своим свойствам близок к

энтропии: $\mathbf{t}^{(\max)} = \arg \max_{S_X} E_Q(\mathbf{t}) = \arg \min_{S_X} \|\mathbf{t}\|_Q^2$, $\mathbf{t}^{(\min)} = \arg \min_{S_X} E_Q(\mathbf{t})$, если $\exists j$:

$t_j^{(\min)} = 1$ и $t_k^{(\min)} = 0 \quad \forall k \neq j$ (сингулярное распределение, $E_Q(\mathbf{t}^{(\min)}) = 0$).

Тогда из (2.38) следует справедливость утверждения.

Утверждение 2.5. Для меры конъюнктивного конфликта верно равенство

$$K^D(F_1, F_2) = \frac{1}{2} (E_Q(\mathbf{m}_1) + E_Q(\mathbf{m}_2)) + \frac{1}{2} \|\mathbf{m}_1 - \mathbf{m}_2\|_Q^2 - \sum_{B,C} r_{B,C} m_1(B) m_2(C). \quad (2.39)$$

Формула (2.39) показывает, что конъюнктивную меру конфликта можно представить как сумму среднего значения энтропии свидетельств, расстояния между свидетельствами и последнего слагаемого (вычитаемого), которое характеризует взаимодействие слабо пересекающихся фокальных элементов свидетельств.

Следствие 2.1. Если $F_1 = (\mathcal{A}, m_1)$ и $F_2 = (\mathcal{A}, m_2)$, где $A' \cap A'' = \emptyset$ $\forall A', A'' \in \mathcal{A}$ ($A' \neq A''$), то

$$K^D(F_1, F_2) = 1 - \sum_{B \in \mathcal{A}} m_1(B) m_2(B) = \frac{1}{2} (E_I(\mathbf{m}_1) + E_I(\mathbf{m}_2)) + \frac{1}{2} \|\mathbf{m}_1 - \mathbf{m}_2\|_I^2, \quad (2.40)$$

где I – единичная матрица и $\|\mathbf{x}\|_I := \sqrt{\mathbf{x}^T \mathbf{x}}$ – евклидова норма.

Заметим, что функционал $E_I(\mathbf{t}) = \sum_B t(B)(1 - t(B))$, определенный на симплексе S_X , удовлетворяет условиям: $\mathbf{t}^{(\max)} = \arg \max E_I(\mathbf{t})$, если $t_k^{(\max)} = \frac{1}{2^{|X|-1}} \forall k$ (равномерное распределение, $E_I(\mathbf{t}^{(\max)}) = 1 - \frac{1}{2^{|X|-1}}$), $\mathbf{t}^{(\min)} = \arg \min E_I(\mathbf{t})$, если $\exists j: t_j^{(\min)} = 1$ и $t_k^{(\min)} = 0 \forall k \neq j$ (сингулярное распределение, $E_I(\mathbf{t}^{(\min)}) = 0$). Кроме того, учитывая неравенство $1 - \alpha \leq -\log_2 \alpha$, $0 < \alpha \leq 1$, получим, что $E_I(\mathbf{t}) \leq S(\mathbf{t}) := -\sum_k t_k \log_2 t_k$, где $S(\mathbf{t})$ – энтропия Шеннона. Таким образом, конъюнктивная мера конфликта (2.40) в этом случае равна средней энтропии функций масс двух свидетельств и квадрату расстояния между распределениями функций масс двух свидетельств. Энтропия функции масс характеризует величину внутреннего конфликта свидетельства, а расстояние между распределениями функций масс – величину внешнего конфликта.

Отметим, что конъюнктивная мера конфликта на непересекающихся структурах доверия (disjoint belief structures) удовлетворяет неравенству треугольника.

Утверждение 2.6. Если $F_i = (\mathcal{A}, m_i)$, $i = 1, 2, 3$, где $\mathcal{A} = \{A_k\}$, $A_k \cap A_s = \emptyset$ $\forall A_k, A_s \in \mathcal{A}$, $k \neq s$, то $K^D(F_1, F_3) \leq K^D(F_1, F_2) + K^D(F_2, F_3)$.

Справедливость этого утверждения следует из того, что для указанных тел свидетельств $K^D(F_k, F_j) = 1 - \sum_A m_k(A) m_j(A) = 1 - (\mathbf{m}_k, \mathbf{m}_j)$ и следующей геометрической леммы.

Лемма 2.1. Если $\mathbf{a} = (a_i)$, $\mathbf{b} = (b_i)$,

$\mathbf{c} = (c_i) \in D_n = \left\{ (t_i)_{i=1}^n : t_i \geq 0, i = 1, \dots, n, \sum_{i=1}^n t_i \leq 1 \right\}$, то $(\mathbf{a}, \mathbf{b}) + (\mathbf{b}, \mathbf{c}) \leq 1 + (\mathbf{a}, \mathbf{c})$.

Пусть $F_1 = (\mathcal{A}, m_1)$ и $F_2 = (\mathcal{A}, m_2)$, где $A' \cap A'' = \emptyset \quad \forall A', A'' \in \mathcal{A} \quad (A' \neq A'')$.

Тогда в силу Утверждения 2.2 имеем

$$K_1^{DP}(F_1, F_2^{(-)}) = \frac{1}{N-1} K^D(F_1, F_2) = \frac{1}{2(N-1)} (E_I(\mathbf{m}_1) + E_I(\mathbf{m}_2)) + \frac{1}{2(N-1)} \|\mathbf{m}_1 - \mathbf{m}_2\|_I^2.$$

Аналогично,

$$K_2^{DP}(F_1, \bar{F}_2) = K^D(F_1, F_2) = \frac{1}{2} (E_I(\mathbf{m}_1) + E_I(\mathbf{m}_2)) + \frac{1}{2} \|\mathbf{m}_1 - \mathbf{m}_2\|_I^2.$$

2.3.7 Изменение метрической и энтропийной частей меры конфликта

В случае положительной согласованности при комбинировании тел свидетельств мера конфликта не уменьшается. С другой стороны, в мере конъюнктивного конфликта можно выделить метрическую и энтропийную части. Возникает вопрос об изменении именно этих частей при комбинировании тел свидетельств.

Утверждение 2.7. Если $F_1 = (\mathcal{A}, m_1)$ и $F_2 = (\mathcal{A}, m_2)$, где $\emptyset \notin \mathcal{A}$ и $A \cap B = \emptyset \quad \forall A \neq B, A, B \in \mathcal{A}$, то метрическая составляющая меры конъюнктивного конфликта не уменьшится после комбинирования этих тел свидетельств с помощью конъюнктивного правила.

Примечание 2.4. Если при тех же условиях в Утверждении 2.7 $\emptyset \in \mathcal{A}$, то метрическая составляющая меры конъюнктивного конфликта может уменьшиться после комбинирования тел свидетельств с помощью конъюнктивного правила. Например, для $F_1 = \alpha F_{\{x_2\}} + (1 - \alpha) F_{\emptyset}$, $F_2 = \beta F_{\{x_1\}} + (1 - \beta) F_{\emptyset}$,

$\alpha, \beta \in [0, 1]$ имеем $\frac{1}{2} \|\mathbf{m}_1 - \mathbf{m}_2\|_I^2 = \alpha^2 + \beta^2 - \alpha\beta$, $D(F_1, F_2) = F_{\emptyset}$, $\frac{1}{2} \|\mathbf{m}^D - \mathbf{m}_1\|_I^2 = \alpha^2$

. Тогда $\Delta = \frac{1}{2} \|\mathbf{m}^D - \mathbf{m}_2\|_I^2 - \frac{1}{2} \|\mathbf{m}_1 - \mathbf{m}_2\|_I^2 = \beta(\alpha - \beta)$. Последнее выражение может быть любым по знаку.

Энтропийная часть меры конъюнктивного конфликта может, как увеличиться, так и уменьшиться после комбинирования тел свидетельств с помощью конъюнктивного правила.

Утверждение 2.8. Если $F_1 = (\mathcal{A}, m_1)$ и $F_2 = (\mathcal{A}, m_2)$, где $\emptyset \notin \mathcal{A}$ и $A \cap B = \emptyset \ \forall A \neq B, A, B \in \mathcal{A}$, то $E_I(DP(F_1, \bar{F}_2)) \geq E_I(\bar{F}_2)$.

Другими словами, энтропия после комбинирования тел свидетельств F_1 и \bar{F}_2 с помощью дизъюнктивного правила не уменьшится по сравнению с энтропией свидетельства \bar{F}_2 .

Метрическая часть меры дизъюнктивного конфликта может, как увеличиться, так и уменьшиться после комбинирования тел свидетельств с помощью дизъюнктивного правила.

2.3.8 Заключение по подразделу

В этом подразделе отчета приведены результаты по исследованию мер конфликта, индуцированных конъюнктивными и дизъюнктивными правилами комбинирования свидетельств в рамках теории функций доверия. В частности, рассмотрены некоторые условия согласования правил комбинирования с мерами конфликта. Показано, как условия согласования связаны с аксиоматикой мер конфликта.

Кроме того, показано, что в конъюнктивных мерах конфликта можно выделить метрическую и энтропийную составляющие. Показано, что энтропия свидетельств является важной характеристикой наряду со значениями меры конфликта (также как и выбор самой меры) при принятии решений о выборе тел свидетельств для комбинирования.

В некоторых частных случаях (непересекающиеся структуры доверия) показано, что метрическая составляющая меры конъюнктивного конфликта не уменьшится после комбинирования тел свидетельств с помощью конъюнктивного правила. Аналогично показано, что энтропия после комбинирования тел свидетельств с помощью дизъюнктивного правила не уменьшится.

Результаты этого исследования отражены в [132].

2.4 Построение и применение теоретических моделей для расчета степени манипулируемости правил агрегирования предпочтений

В этом подразделе отчета представлены результаты по моделированию и оценке степени манипулируемости правил агрегирования предпочтений. Эта работа ведется в МЛАВР на протяжении последних нескольких лет. В 2018 году предложена модель индивидуального и коалиционного манипулирования правил агрегирования предпочтений в одномерной модели формирования предпочтений. Проведены расчеты манипулируемости для мажоритарных правил выбора для трех альтернатив.

2.4.1 Модель манипулирования

В одномерной модели формирования предпочтений альтернативы и идеальные точки агентов позиционируются на оси. Рассматривается манипулирование со стороны агента. Агент может предъявлять в качестве неискренних предпочтений только те упорядочения альтернатив, которые допустимы текущим позиционированием альтернатив на оси, получающиеся сдвигом идеальной точки агента по оси. Так, например, упорядочение $c > b > a$ допустимо для альтернатив, расположенных на оси, как a, b, c , слева направо, а упорядочение $a > c > b$ недопустимо, т.к. не существует точки-позиции агента на оси, к которой альтернативы a и c одновременно были бы ближе, чем альтернатива b . Для каждого профиля для каждого агента в отдельности рассматриваются все возможные его перемещения (позиционирования) по оси. Для таких позиционирований строятся соответствующие профили (неискренних) предпочтений и рассчитывается коллективный выбор в соответствии с правилом выбора. Результат выбора по неискренним предпочтениям сравнивается с выбором по истинным предпочтениям данного агента с использованием расширенных предпочтений. Если результат коллективного выбора по про-

филю с неискренними предпочтениям лучше для манипулирующего агента, чем результат выбора с его истинными предпочтениями, то профиль считается манипулируемым.

2.4.2 Правила выбора

Манипулируемость в условиях одномерных предпочтений рассчитана для следующих мажоритарных правил выбора.

1. Минимальное доминирующее множество (Minimal dominant set)

Множество S называется доминирующим, если каждая альтернатива из множества S доминирует каждую альтернативу вне множества S , т.е.

$$\forall x \in A : x \in S \Leftrightarrow [\forall y \in A \setminus S, x \mu y].$$

Доминирующее множество S называется минимальным, если никакие из его собственных подмножеств не являются доминирующими.

2. Минимальное недоминируемое множество (Minimal undominated set)

Множество S называется недоминируемым, если никакая альтернатива вне S не доминирует никакую альтернативу внутри S . Недоминируемое множество S называется минимальным, если никакие из его собственных подмножеств не являются недоминируемыми. Если минимальное недоминируемое множество не единственно, то в качестве выбора берется объединение всех таких множеств.

3. Минимальное слабоустойчивое множество (Minimal weakly stable set) (MWSS)

Множество $S \subseteq A$ называется слабоустойчивым, если выполнено следующее условие:

$$\forall x \in A : x \in S \Leftrightarrow (\exists y \in A \setminus S : y \mu x \Rightarrow \exists z \in S : z \mu y),$$

т.е. говорят, что x принадлежит S , если выполнено следующее: если существует альтернатива $y \in A \setminus S$, которая доминирует x по мажоритарному отно-

шению μ , то существует некоторая альтернатива $z \in S$, которая доминирует y , т.е., $z\mu y$.

Слабоустойчивое множество S называется минимальным, если никакие из его собственных подмножеств не являются слабоустойчивыми. Если минимальное слабоустойчивое множество не единственно, то в качестве выбора берется объединение всех таких множеств.

4. Правило Фишберна (Fishburn's rule)

Бинарное отношение γ строится следующим образом:
 $x \gamma y \Leftrightarrow D(x) \subset D(y)$. В выбор включаются альтернативы, недоминируемые по γ (являющимся антирефлексивным и транзитивным), т.е.

$$x \in C(\vec{P}) \Leftrightarrow [\bar{\exists} y \in A \mid y \gamma x].$$

5. Непокрытое множество I (Uncovered set I)

Бинарное отношение δ строится следующим образом:

$$x \delta y \Leftrightarrow L(x) \supset L(y).$$

Отношение δ также является частичным порядком. Выбираются альтернативы, недоминируемые по δ , т.е.

$$x \in C(\vec{P}) \Leftrightarrow [\bar{\exists} y \in A \mid y \delta x].$$

6. Непокрытое множество II (Uncovered set II)

Говорят, что альтернатива x B -доминирует альтернативу y (xBy), если верно $x\mu y$ и $D(x) \subseteq D(y)$ в мажоритарном отношении μ . Выбор состоит из альтернатив, недоминируемых по B , т.е.

$$x \in C(\vec{P}) \Leftrightarrow [\bar{\exists} y \in A \mid yBx].$$

7. Правило Ричелсона (Richelson's rule)

По мажоритарному отношению μ , бинарное отношение σ строится следующим образом:

$$x \sigma y \Leftrightarrow [L(x) \supseteq L(y) \wedge D(x) \subseteq D(y) \wedge ([L(x) \supset L(y)] \vee [D(x) \subset D(y)])].$$

Отношение σ также является частичным порядком. Выбираются альтернативы, недоминируемые по σ : $x \in C(\vec{P}) \Leftrightarrow [\nexists y \in A \mid y \sigma x]$.

8. Правило Копланда I (Copeland's rule I) (C1)

Строится функция $u(x)$, равная разности мощностей нижнего и верхнего контуров мажоритарного отношения μ , т.е., $u(x) = \text{card}(L(x)) - \text{card}(D(x))$. Выбираются альтернативы с максимальным значением u :

$$x \in C(\vec{P}) \Leftrightarrow [\forall y \in A, u(x) \geq u(y)].$$

9. Правило Копланда II (Copeland's rule II)

Функция $u(x)$ определена, как мощность нижнего контура альтернативы x в мажоритарном отношении μ . Выбираются альтернативы с максимальным значением u .

10. Правило Копланда III (Copeland's rule III)

Функция $u(x)$ определена, как мощность верхнего контура альтернативы x в мажоритарном отношении μ . Выбираются альтернативы с минимальным значением u .

11. k-устойчивое множество (k-stable set) (k=1, 2, 3)

Говорят, что альтернатива x принадлежит k-устойчивому множеству, если выполнено одно из следующих условий:

а) x не доминируется никакой альтернативой вне множества kSS по мажоритарному отношению μ , или, в противном случае,

б) Если $\exists y \notin kSS : y \mu x$ то $\exists z_1, \dots, z_k, z_1 \in kSS$ и $z_1 \mu z_2 \mu \dots \mu z_k \mu y$. Множество kSS называется минимальным, если никакие из его собственных подмножеств не являются k-устойчивыми. Если минимальное k-устойчивое множество не единственно, то берется объединение всех таких множеств. Для $k=1$ в таком определении k-устойчивое множество совпадает с минимальным слабоустойчивым множеством (Minimal weakly stable set).

k-устойчивое множество Π (k-stable set Π) ($k=1, 2, 3$) является одним из вариантов правила k-устойчивое множество (k-stable set).

2.4.3 Индекс манипулируемости и методика расчета

Для оценки степени манипулируемости рассчитывается индекс манипулируемости правил коллективного выбора Ницана-Келли, определенный следующим образом:

$$NK = \frac{d_0}{d_{total}},$$

где d_0 – число манипулируемых профилей, d_{total} – общее число профилей.

При расчете индекса манипулируемости используются ранее введенные нами расширенные предпочтения, в которых упорядочиваются множества альтернатив исходя из предпочтений агентов на единичных альтернативах. Расширенные предпочтения задаются аксиоматически. Степень манипулируемости правил коллективного выбора рассчитана для 4 расширений предпочтений для случая множественного выбора для 3 альтернатив, для 10 расширений для 4 альтернатив, и для 12 расширений для 5 альтернатив.

2.4.4 Результаты расчетов

Для нечетного числа агентов в одномерной модели манипулируемость всех рассматриваемых правил равна нулю. Для четного числа агентов, для обоих способов манипулирования, индивидуального и коалиционного, манипулируемость с ростом числа агентов уменьшается.

Для коалиционного случая одномерной модели, как и для индивидуального, выявлены три группы мажоритарных правил, имеющих совпадающие между собой значения степени манипулируемости. Первая группа содержит только правило Minimal dominant set, вторая группа – правила Uncovered set I, Richelson's, Copeland I и II, третья группа – правила Minimal undominated set,

Minimal weakly stable set, Fishburn, Uncovered set II, Copeland III, k-stable set и k-stable set II.

Обозначим группы правил следующим образом: G1: Minimal dominant set, G2: Uncovered set I, Richelson's, Copeland I и I,I, G3: Minimal undominated set, Minimal weakly stable set, Fishburn, Uncovered set II, Copeland III, k-stable set и k-stable set II.

Проведем сравнение степени манипулируемости для индивидуального и коалиционного случаев.

Для 3 альтернатив рассчитывается индекс манипулируемости мажоритарных правил выбора для расширений Leximin, Leximax, Risk-averse и Risk-lover. Результаты вычислений представлены на рисунке 2.1 для расширений Leximin и Leximax, и на рисунке 2.2 для расширений Risk-averse и Risk-lover.

Для расширения предпочтений Leximin и Risk-lover манипулируемость правил G3 равна 0. Для расширения Leximin при коалиционном манипулировании правила G2 становятся более манипулируемыми, чем правило G1, начиная с 8 агентов, тогда как при индивидуальном манипулировании G2 менее манипулируема вплоть до 20 агентов.

Для расширенных предпочтений Leximax правило G1 менее манипулируемо, чем G2 для индивидуального и коалиционного манипулирования. Однако, правила G3 в коалиционном случае более манипулируемы, чем две другие группы, до 8 агентов, и более манипулируемы, чем правило G1 для рассматриваемого числа агентов, тогда как для индивидуального случая правила G3 менее манипулируемы, чем две остальные группы правил.

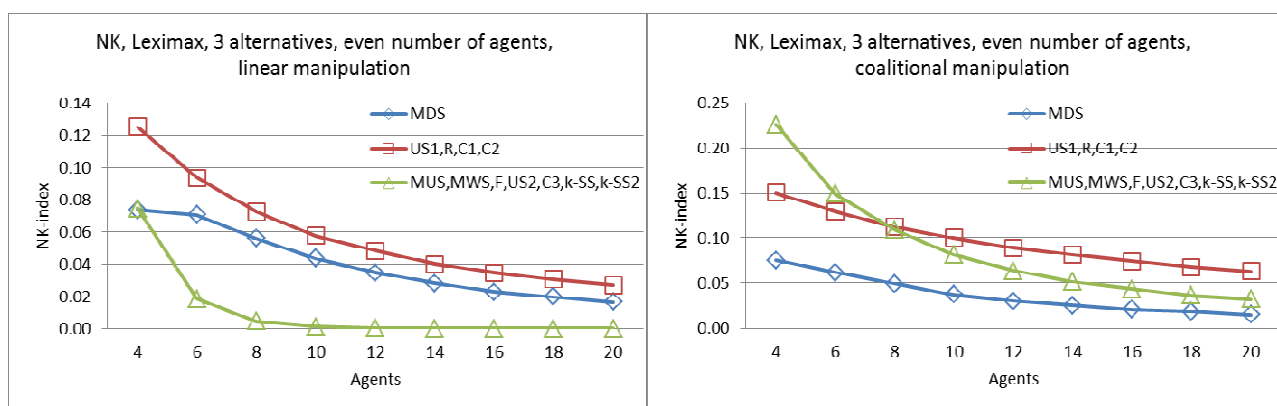


Рисунок 2.1 – индекс манипулируемости мажоритарных правил выбора для расширений Leximin, Leximax,

Для расширения Risk-averse уровень манипулируемости для коалиционного случая G1 и G3 совпадает и значения манипулируемости превышают, чем маипулируемость G2 для числа агентов не более 8. Для индивидуального случая манипулируемость G3 меньше манипулируемости G2, которая, в свою очередь, меньше G1.

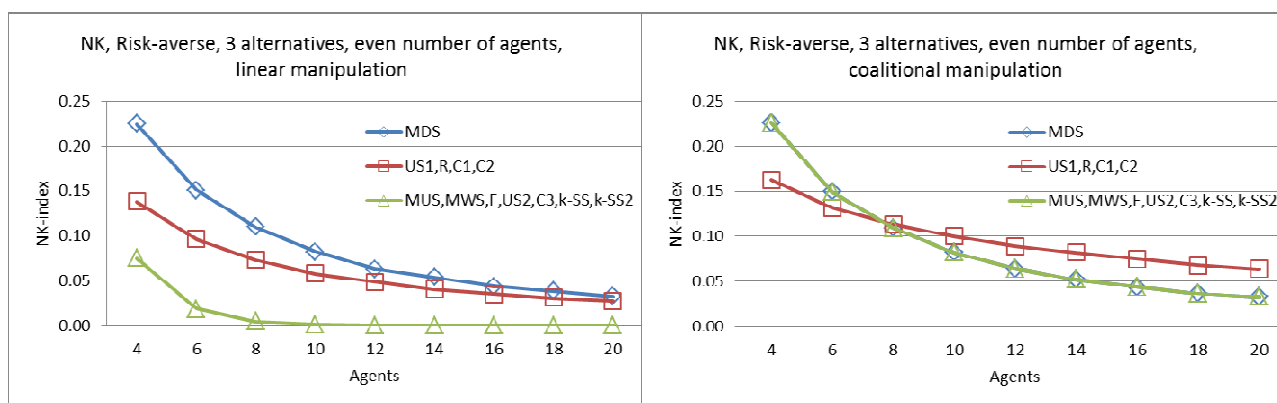


Рисунок 2.2 – Индекс манипулируемости мажоритарных правил выбора для расширений Risk-averse и Risk-lover

Для расширения Risk-lover значения манипулируемости G3 нулевые. При индивидуальном манипулировании манипулируемость G2 меньше, чем манипулируемость G1. При коалиционном манипулировании манипулируемость G2 меньше, чем манипулируемость G1 только для небольшого числа агентов, до 12.

2.4.5 Некоторые выводы

Предложена модель индивидуального и коалиционного манипулирования правил агрегирования предпочтений в условиях одномерной модели. Для трех альтернатив выявлены минимально манипулируемые группы мажоритарных правил.

Результаты исследований отражены в работах [133, 134].

3 Анализ моделей в финансовой и банковской сферах

В этом разделе отчета приведены результаты анализа некоторых моделей в финансовой и банковской сферах, а именно:

- определение адекватного метода валидации рейтинговой модели с учетом бимодальности распределения доли дефолтов, непредусмотренное стандартами Базельского комитета;
- исследование критического уровня мошенничества в банковской системе: агентно-ориентированный подход.

3.1 Анализ методов валидации моделей оценки вероятности дефолта

3.1.1 Введение и постановка задачи

Точность моделей оценки вероятности дефолта (PD) проверяется через сравнение количества фактических реализаций дефолтов и их прогнозного числа. Такое сравнение называется биномиальным тестом. Считается, что если количество фактических реализаций входит в доверительный интервал прогноза среднего числа единичных реализаций, то модель является точной, в противном случае — нет, и ее необходимо переработать, отказаться от нее или для сферы управления кредитным риском ввести поправку на оценку компонент риска.

Биномиальный тест рассматривают в двух видах: без корреляции и с корреляцией. Первая разновидность имеет узкий доверительный интервал (заниженный уровень дефолтов). Поскольку кредитоспособность заемщиков в реальности зависит от макроэкономических факторов, то заемщики взаимосвязаны, что в итоге приводит к большему числу дефолтов, чем давал бы процесс генерации некоррелированных случайных величин. В итоге часто на практике модели PD не проходят валидацию по биномиальному тесту без корреляции. По этой причине используют вторую разновидность биномиального теста. В ее рамках ожидается наличие большего числа дефолтов по сравнению с моделью без корреляции. На практике модели PD валидацию,

как правило, проходят. Однако возникают вопросы, что понимать под корреляцией, как ее посчитать, какое ее значение использовать при валидации моделей PD в российских условиях, как валидировать модель с учетом корреляции. Этим вопросам посвящено данное исследование.

В стандартах Базель II при расчете требований к капиталу используется корреляция активов, но она имеет ряд недостатков. Во-первых, Базель II считает, что корреляция активов имеет отрицательную зависимость с PD, в то время как выводы [135] обратные [136]. Во время экономического кризиса мелкие фирмы (с высоким PD) не имеют доступа к финансовым рынкам по сравнению с крупными (с низким PD), поэтому более подвержены изменению состояния экономики. Следовательно, отрицательная связь между PD и корреляции активов в Базельской формуле ставится под вопрос. Во-вторых, [136] и [137, 138, 139] отмечали, что в целом корреляция активов в [140] завышена. Вместо 12%-24% для корпоративных заемщиков истинная корреляция активов колеблется в размере от 0% до 5%. Однако, [141] на данных США за 18 лет (с 1981 г. по 1999 г.) получили, что значения корреляции могут выходить за диапазон от 0% до 5%, принимая отрицательные значения и достигая максимума в 39%. Такие оценки корреляции дефолтов ближе к оценкам [140].

Помимо неоднозначности значений корреляции в академических статьях термин «корреляция активов» подменяется термином «корреляция дефолтов». Так, в [142] для биномиального теста на проверку точности рейтинговой модели предполагается использование корреляции активов. В приведенной в данной работе формуле указано $\rho = R = AC$ и говорится, что $\delta = DC$ [142]. При этом в [137, р.259] ρ ошибочно используется как корреляция дефолтов (DC).

В [137] также предложены две вариации биномиального теста: без корреляции и с учетом корреляции. Если в тест с корреляцией подставить корреляцию, равную 0, и сравнить значения с результатами теста без учета корреляции, то получаются различные границы доверительного интервала (см. ри-

сунок 3.3). Это говорит о несогласованности подходов, которую сами авторы не замечают и не обсуждают.

Работа с коррелированными случайными (бинарными) величинами ведется с 1961 г. [143, 144]. Такие величины часто встречаются в области медицины при анализе эффекта от применения лекарственного препарата, когда в группах пациентов (кластерах, грейдах) исследуется случайная величина бинарного отклика (выздоровел пациент или нет). Поскольку совместное распределение таких случайных величин сводится к нормальному (гауссовскому), или q -гауссовскому [145], то можно ожидать, что идея модели кредитного риска была заимствована О. Васичеком [146] из области медицины, хотя явных ссылок в его работе не сделано.

3.1.2 Модель дефолтов по Васичеку и на основе бинарного распределения

Целью расчетов является оценка влияния корреляции дефолтов на уровень их частоты. Есть два подхода: первый, предложенный [144], соответствует природе процесса генерации данных, второй, представленный [146], не соответствует, но он лежит в основе регулирования банков. Начнем с рассмотрения процесса генерации данных.

Если [144] моделируют компоненты модели Z и Y_i , используя биномиальное распределение, то [146] применяет к данным компонентам (x и ε_i) стандартное нормальное распределение. Влияние компонент на результирующую переменную в модели определяется корреляцией ρ , в подходе О. Васичека она именуется корреляцией активов (АС). В обеих моделях предполагается, что значение системного фактора (Z или x) едино для кластера (системы). Если в [144] объясняемая переменная X_i является бинарной, то для соответствующей переменной A_i в модели [146] требуется дополнительное преобразование. A_i (активы) сравниваются с долгом (квантиль ρ стандартного нормального распределения) аналогично логике нахождения точки дефолта (Default Point), используемой, например, в [147]. В обоих ме-

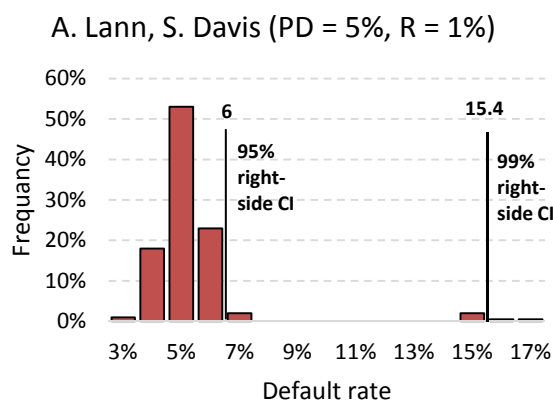
тодах последовательность СВ при бесконечно большом количестве наблюдений стремится к стандартному нормальному распределению. Сравнение рассматриваемых подходов представлено в Таблице 3.1.

Таблица 3.1 – Сравнение методов

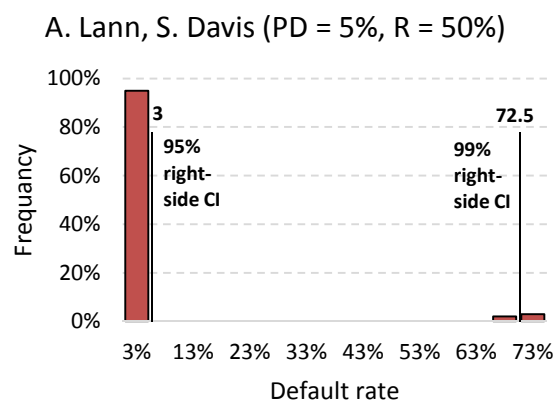
Параметр	[146]	[144]
Системный фактор	$x \sim N(0,1)$, где N – нормальное распределение	$Z \sim Bi(1, p)$, где Bi – бимодальное распределение
Идиосинкратический фактор	$\varepsilon_i \sim N(0,1)$	$Y_i \sim Bi(1, p)$
Корреляция	$R = \rho$	$R = \rho = r^2$
Модель	$A_i = x\sqrt{\rho} + \varepsilon_i\sqrt{1-\rho}$	$X_i = ZU_i + (1 - U_i)Y_i$, where $U_i \sim Bi(1, r)$
Вероятность дефолта	$PD = p = Prob(A_i < D_i)$, где D_i – долг	$PD = p$
Зависимая переменная	Стоимость активов $A_i \sim N(0,1)$	Дефолт $X_i = \begin{cases} 0 \\ 1 \end{cases}$
Распределение суммы случайных величин	N	N

В [144] Z , так же как и Y , имеет биномиальное распределение. В наших расчетах в рамках одной итерации Z едино для всех заемщиков, но различается между итерациями. Под кластером можно понимать портфель заемщиков или только группу заемщиков в составе одного разряда рейтинговой шкалы банка, если корреляция внутри рейтинговой шкалы варьируется.

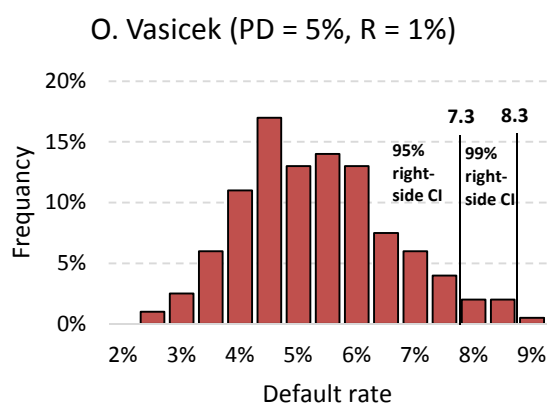
Возврат и невозврат кредита — это два события. Доля дефолтов в выборке может меняться из-за случайности процесса. Когда события независимы (некоррелированы), распределение доли дефолтов в выборке имеет колоколообразную форму (примерно как на рисунке 3.1B). Когда эти события взаимосвязаны (коррелированы), то недефолты и дефолты происходят чаще, поэтому распределение доли дефолтов раздваивается (становится бимодальным) (как на рисунках 3.1A и 3.1C). Однако в методе [146] раздвоение не учтено (рисунок 3.1D). При этом частота экстремальных (высоких) значений DR возрастает. Это видно по границе 99% правостороннего доверительного интервала: с вероятностью 1% DR превысит значение 15,4% на рисунке 3.1A и значение 8,3 на рисунке 3.1B. При увеличении корреляции распределение по [144] становится бимодальным. В итоге доверительные интервалы для метода [144] шире, чем для метода [146].



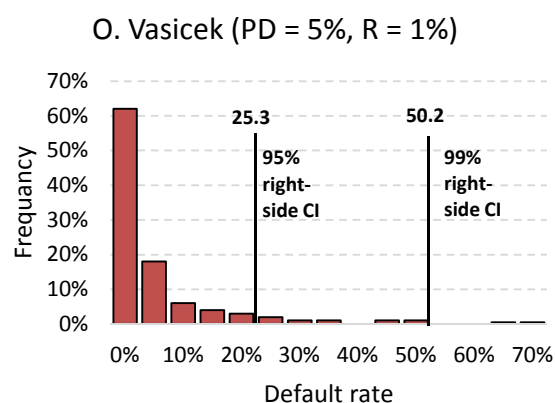
(A)



(C)



(B)



(D)

Рисунок 3.1 – Функция распределения дефолтов в портфеле из 1000 заемщиков, результат на основе 1000 итераций: сравнение методов [146] и [144]

3.1.3 Эмпирическое распределение уровня частоты дефолтов

3.1.3.1 Проверка бимодальности по регионам

Бимодальность распределения DR наблюдается для рынков корпоративных облигаций США, Европы, России и развивающихся рынков, хотя бимодальность не выражена (рисунок 3.2). Корреляция дефолтов в Европе выше, чем в США, но сопоставима с корреляцией на развивающихся рынках и в России. В среднем корреляция дефолтов для всех четырех регионов составляет около 2% (Таблица 2.2). Ожидается, что при условии большей однород-

ности данных (например, с учетом отдельных секторов, таких как судоходство) бимодальность будет более отчетливой.

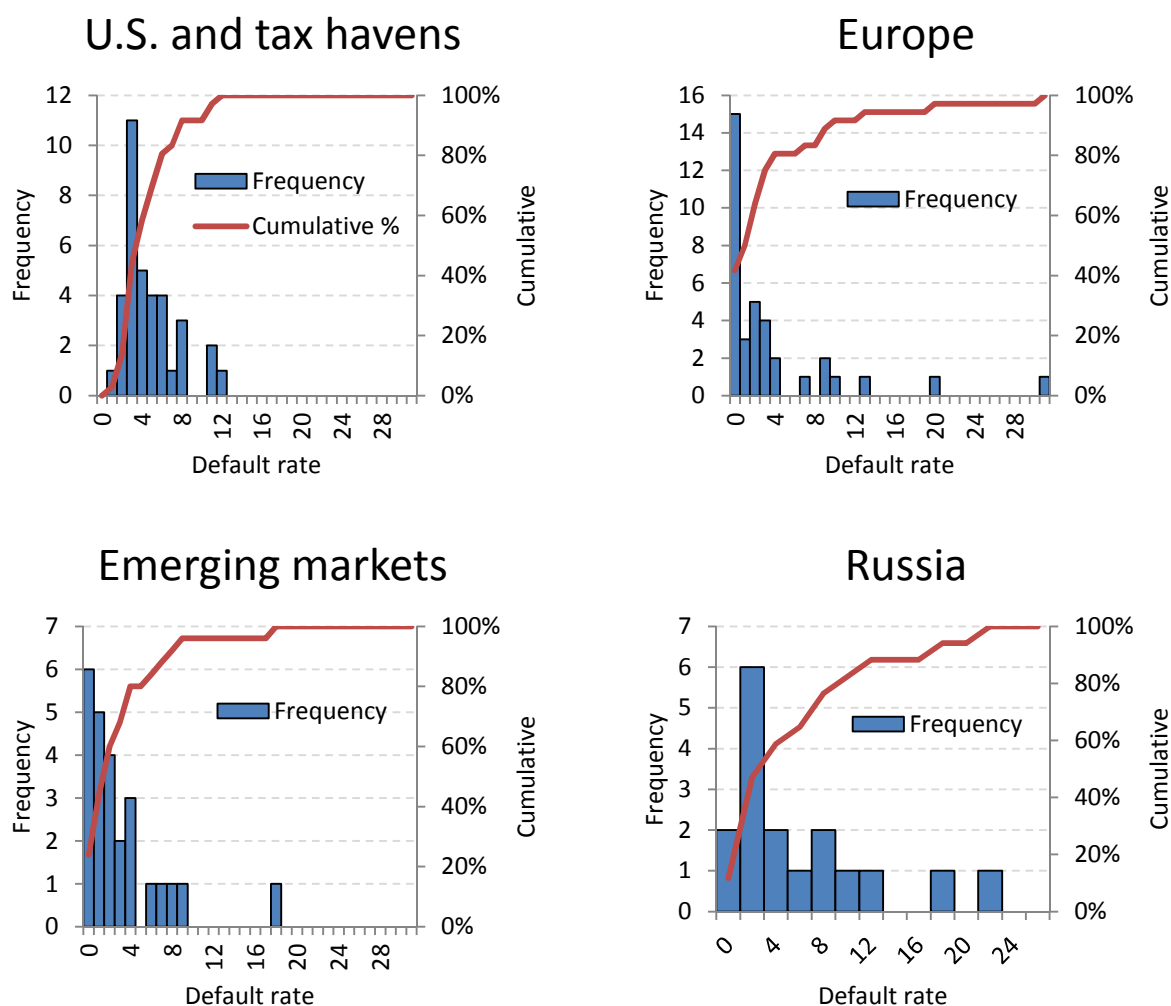


Рисунок 3.2 – Распределение уровня частоты дефолтов по регионам

Корреляция уже отражена в данных заемщика. Это результат системного фактора. Например, из-за перебоев между поставщиками во время кризиса меняется дебиторская или кредиторская задолженность; изменение обменного курса увеличивает финансовый рычаг для заемщиков с кредитами в иностранной валюте. По этой причине эмпирическое распределение уровня частоты дефолта по экономическому циклу также согласуется с бимодальным распределением, поскольку корреляция существует.

Таблиц 3.2 – Корреляция дефолтов по регионам на данных корпоративных облигаций

Регион	Период	Корпоративные облигации
--------	--------	-------------------------

		Средний DR, %	Макс DR, %	Корреляция дефолтов, %
США	1981-2016	4.2	11.7	0.7
Европа	1981-2016	4.4	66.7	40.0
	1992-2016	3.7	20.0	3.0
Развивающиеся рынки	1993-2016	2.8	17.6	2.5
Россия	2002-2018	5.3	21.2	2.0
Среднее (исключая выброс – Европа, 1981-2016)		4.0	17.6	2.0

3.1.3.2 Проверка бимодальности по модели PD по рынку российских корпоративных облигаций

Ключевой вывод состоит в том, что при проведении оценки точности модели следует проверять бимодальное распределение прогнозных DR относительно бимодального распределения фактических DR, т.е. не использовать биномиальный тест с корреляцией (который все еще предполагает одномодальность распределения DR и недооценивает хвост распределения потерь). Проблема сравнения фактического бимодального распределения и теоретического одномодального состоит в том, что, наблюдая одну моду и укладываясь в правую границу доверительного интервала, мы можем получить неточную модель лишь потому, что вторая мода еще не реализовалась.

Приведем пример использования биномиального теста на практике. Для этого возьмем публичные данные о дефолтах российских корпоративных эмитентов облигаций и модель PD, построенную в [148], уровень DR в 21,2% достигается при корреляции в 2%.

В примере корпоративных заемщиков РФ построенная в 2015 г. модель не точна на уровне 10% (переоценивает низкий уровень DR), однако способна улавливать правую моду с DR в период рецессии. Данный вывод сделан на основе сравнения распределения фактических и прогнозных DR (см. рисунок 3.3) за экономический цикл с использованием статистики Колмогорова–Смирнова. Максимальная разница кумулятивной кривой равна 49% в точке DR = 4%. 49% сравниваем с критическим значением статистики Колмогорова–Смирнова D_α :

$$D_{\alpha} = c(\alpha) \sqrt{\frac{n_1 + n_2}{n_1 n_2}},$$

где n_1 – количество наблюдений в первом распределении (в примере $n_1 = 15$); n_2 – количество наблюдений во втором распределении (в примере $n_2 = 15$); α – уровень значимости (в примере $\alpha = 10\%$); $c(\alpha)$ – статистика Колмогорова–Смирнова (в примере $c(\alpha) = 1,22$). При 10% уровне статистической значимости нулевая гипотеза о равенстве распределений отвергается, распределения прогнозного и модельного PD отличны друг от друга (рисунок 3.3).

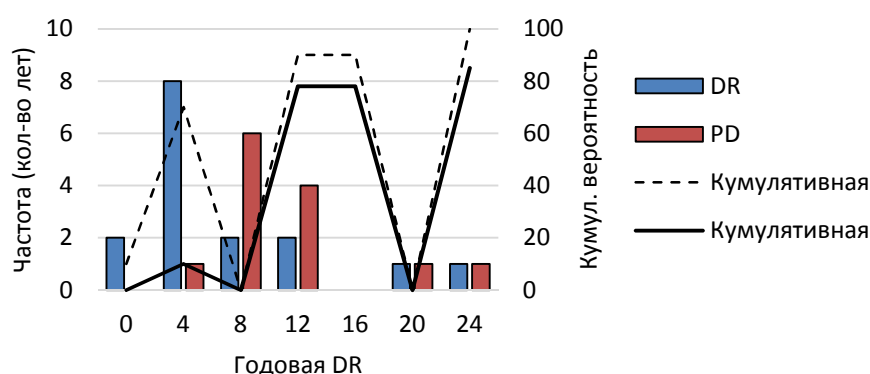


Рисунок 3.3 – PD модель для корпоративных облигаций [148]

3.1.3.3 Проверка бимодальности по модели PD судовых компаний

Далее мы анализируем модель PD для судоходных компаний, разработанную в [149].

Средний уровень дефолтов для судоходных компаний за период с 1989 по 2015 год составляет 1,4%. Максимальный уровень дефолтов в 1999 году составил 16,7%, при этом корреляция составила 2,5% (оценки основаны на методе [144]. За период с 2002 по 2015 год (выборка для разработки [149]) корреляция дефолтов – около 0,1%, поскольку максимальный уровень DR за 2002-2015 достигал только 3,7% (Таблица 3.3). Сравним распределения фактических и прогнозных DR (см. рисунок 3.4) за экономический цикл с использованием статистики Колмогорова–Смирнова.

Таблица 3.3 – Корреляция дефолтов для судоходной отрасли

Отрасль	Период	192 судоходные компании (на фонд. бирже)		
		Средний DR, %	Макс DR, %	Корреляция дефолтов, %
Судоходство	2002-2015	1.6	3.7	0.1
	1989-2015	1.4	16.7	2.5

На рисунок 3.4А показано, что оценки PD «logit 1» смещены вправо относительно значений фактических значений DR. Поэтому модель завышает уровень дефолтов. Максимальное расстояние между кумулятивными кривыми PD и DR достигает 64%. При 5% уровне статистической значимости отвергается нулевая гипотеза равенства распределений прогнозируемых и модельных DR. Модель не является точной. В [149] сделан такой же вывод, поэтому «logit 2» выбран в качестве ориентира. На рисунок 3.4В показано, что сдвиг оценок PD, рассчитанных по модели «logit 2», относительно DR меньше, чем по методу «logit 1». В то же время модель по-прежнему завышает DR за четыре года (PD = 6%, 7%, 9%). Тем не менее, разрыв в кумулятивных кривых меньше и достигает 35%. Распределения схожи при 5% уровне статистической значимости. Следовательно, модель является точной.

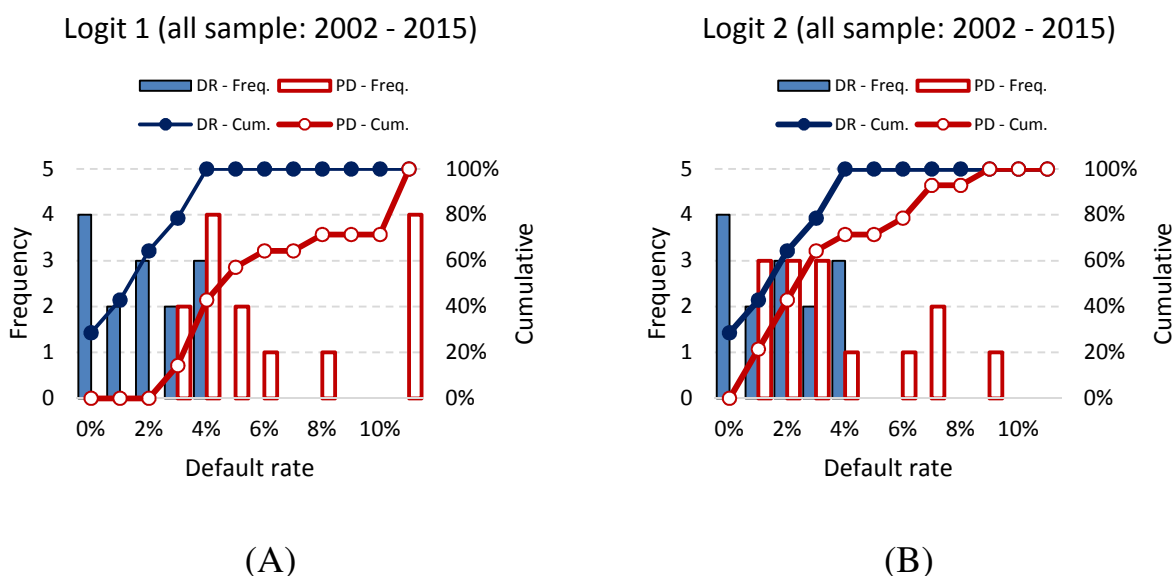


Рисунок 3.4 – PD модель для судоходных компаний: logit 1 и logit 2

На рисунок 3.5А показано, что порядковый логит недооценивает полу-дефолты в рецессию (DR выше 3%), но переоценивает полу-дефолты в пе-

риоды подъема (DR ниже 3%). Здесь и далее под полудефолтами понимается поглощение, изменение имени, быстрое выздоровление после дефолта. Несмотря на то, что распределения прогнозируемого и фактического DR схожи между собой на 5% уровне значимости. Таким образом, недооценка высокого уровня DR незначительна, но требует особого внимания.

Модель лучше улавливает дефолты (рисунок 3.5B), чем полудефолты. Порядковый логит для дефолтов более чувствителен к кризисным годам по сравнению с моделью полудефолтов. Распределения прогнозных и фактических DR схожи при 5% значимости. Следовательно, модель является точной.

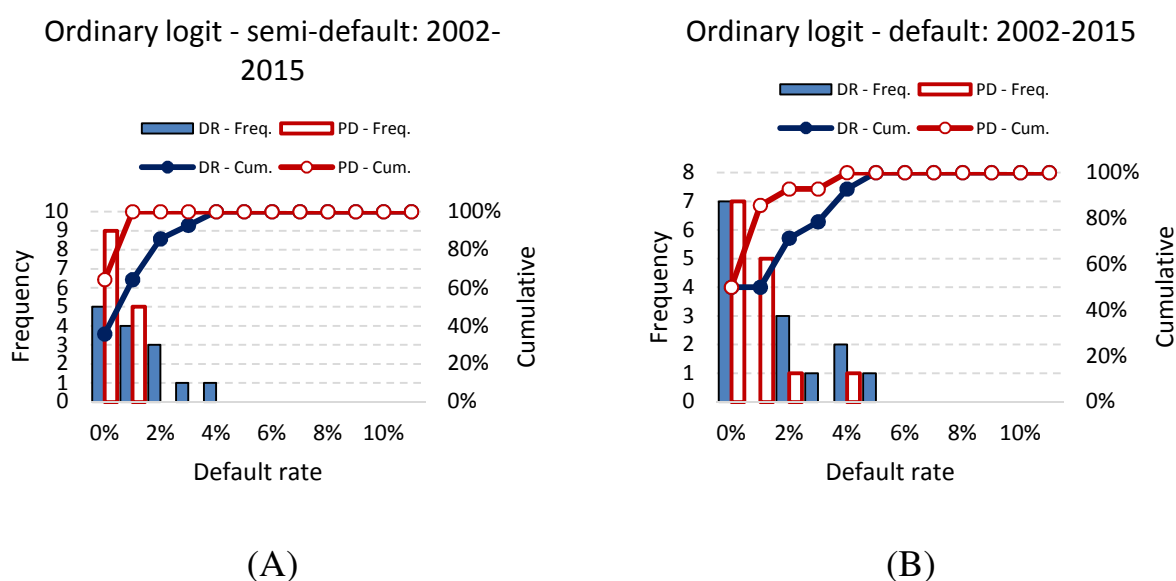


Рисунок 3.5 – PD модель для судоходных компаний: порядковый логит для дефолтов и полу-дефолтов

3.1.4 Заключение и выводы

Метод моделирования коррелированных случайных (бинарных) величин был упомянут еще в [143, 144]. Коррелированные величины и проверка на их основе эффекта применения лекарственного препарата широко используются в медицинских исследованиях. При решении задач управления кредитным риском необходимость в проведении биномиального теста коррелированных случайных бинарных величин появляется при проверке точности прогноза количества дефолтов PD с помощью моделей банка. Благодаря использова-

нию результатов медицинских исследований в данной работе мы смогли найти решение по применению биномиального теста для задачи управления кредитным риском. Кратко перечислим основные выводы.

Во-первых, мы уточнили, что в тесте, предложенном в [142], требуется подставлять корреляцию активов, а не корреляцию дефолтов с целью обеспечения согласованности с подходом для расчета оценки кредитного риска.

Во-вторых, мы объяснили применение биномиального теста и указали на его неверную интерпретацию в [137]. Такой тест на основе метода [144] соответствует природе процесса генерации данных, подход [146] – нет. В данной работе предлагается определять корреляцию через наблюдаемую DR как единственный фактически наблюдаемый компонент на выборке для разработки и использовать параметр корреляции СВ для валидации моделей. Однако предложенный метод применения биномиального теста имеет ограничение в виде предпосылки о неизменности корреляции дефолтов во времени.

В-третьих, мы показали, что корреляция дефолтов для рынка российских корпоративных облигаций определена на уровне 2%. Это близко к интервалам от 0% до 3%, упомянутым в [142], и в [137], хотя значительно меньше среднего значения корреляции дефолтов в США [150] (Таблица 3.4).

В заключение отметим два момента. Во-первых, корреляция может отличаться в разрядах рейтинговой шкалы. Из-за этого требуется повторять данный тест не только на уровне портфеля, но и на уровне отдельного разряда со свойственной ему корреляцией.

Во-вторых, существует необходимость в регулярной валидации модели корреляции активов со стороны регулятора, в идеале Базельского комитета по банковскому надзору. Параметр корреляции, заложенный в [140], калибровался до 2003 г. и на сегодняшний день не пересматривался уже 15 лет. В данном случае появляется методологическое противоречие, когда модели оценки входящих в общую формулу взвешенных по риску активов (RWA) параметров валидируются ежегодно, а сама формула не валидируется, не пе-

рекалибровывается и не перерабатывается, хотя необходимость в этом присутствует. Со стороны банков требуется активная работа по оценке корреляции дефолтов на уровне их кредитных портфелей и, более того, на уровне каждого разряда рейтинговой шкалы.

Таблица 3.4 – Оценки корреляции дефолтов предыдущих и текущего исследования

Работы	Параметр	Страна	R, %	
			Min	Max
[141]	ICC	США	-6.0	39.0
[142]	DC	Не указано	0.0	3.0
[137]	DC	Не указано	0.0	3.0
[140]	AC	Не указано	12.0	24.0
[138, 150]	AC	Германия	0.0	5.0
Настоящий отчет	DC	Россия	2.0	2.0
	DC	Судоходные компании	0.1	2.5

Результаты этого пункта отчета отражены в [151].

3.2 Исследование критического уровня мошенничества в банковской системе: агентно-ориентированный подход

3.2.1 Введение постановка задачи

Мошеннические операции могут инициировать кризис системы и значительные потери общества. Примерами в истории были такие банки как Citi, Goldman Sachs, BCCI, Barrings Bank, Johnson Matthey Bank [152], банк «ФК Открытие» (российская практика), Societe Generale и др. В российской банковской системе за последние 10 лет мошенничеством занимались около 300 российских банков, чьи лицензии были аннулированы с 2013 года.

Исследование [152] отмечает, что при дерегулировании количество мошеннических операций увеличивается (на примере дерегулирования в 1980-е годы), из чего можно сделать вывод, что, в основном, только политика регулятора позволит сократить размеры мошеннической деятельности с целью обеспечить финансовую стабильность. Однако регуляторы сталкиваются с проблемой выбора (в том числе из-за общественного давления): затраты от предотвращения мошеннических операций должны покрывать выгоды (сни-

жение регуляторных издержек (средства налогоплательщиков) на восстановление банковской системы). Последнее требует определить, каково влияние на систему каждого вида мошеннических операций, и каков критический уровень мошеннической деятельности, способный сделать финансовую систему уязвимой.

Несмотря на значительный масштаб мошеннических операций и их влияние на финансовую стабильность, мошенничество изучается в отдельности от макроэкономических моделей и моделей банковской системы. Неучет мошеннических операций в моделях может вести к недооценки риска системы.

Предлагаемая работа определит критические уровни мошеннической деятельности, которые делают финансовую систему уязвимой, на основе агентно-ориентированной подхода моделирования. Подобная задача в академической отечественной и зарубежной литературе полностью еще не рассматривались. Однако спрос на знание критического уровня мошенничества часто поднимается в выступлениях представителей Центральные банков, в частности Банка России (например, при возросшей проблеме фиктивного капитала). В данной работе будет проведена оценка влияния мошеннических операций на уязвимость всей системы, в отличие от предыдущих работ, где изучались только изменения на микро-уровне, например, изменение поведения / политик конкретного банка. На текущий момент в исследовании рассмотрены два вида мошенничества: финансовая пирамида и скрытый риск. Следующим этапом планируется рассмотреть неучет пассивов и фиктивный капитал.

Рассмотрение мошеннических операций банков в рамках агентно-ориентированной модели позволит ответить на следующие вопросы:

- каково влияние на систему каждого вида мошеннической операцией?
- что должно быть учтено в регулировании для уменьшения мошеннических операций и потерь от них?
- каких ресурсов не хватает в борьбе с мошенническими операциями?

3.2.2 Обзор литературы

В последнее десятилетие особое внимание уделяется работам, изучающим причины возникновения кризисов и факторы, усиливающие потери при рецессии. Тем не менее, большое количество примеров на практике, когда мошенническая деятельность банков вела к очень большим потерям не находит частого отражения в макроэкономических моделях или моделях банковской системы. Большая часть работ направлена на создание методов идентификации мошеннических операций. Меньшее внимание уделено вопросам причин усиления мошеннической деятельности и ее влияние на финансовую устойчивость системы. Это будет доказано ниже.

Для оценки современного состояния исследования по данной проблеме анализировались работы по банковской мошеннической деятельности из базы данных зарубежной периодики Science Direct, содержащей около 1 700 журналов по экономике. Было найдено около 90 работ за период с 1986 по 2018 гг., которые прямо или косвенно относятся к теме мошеннической деятельности банков. Около 70% работ представляли собой исследования, предлагающие новые методы идентификации мошеннических операций (в основном внешних, то есть мошенничества против банков, а не мошенничества самих банков) с использованием современных методов анализа данных (например, машинное обучение). Системный обзор таких исследований за период 1997-2008 был сделан в [153]. В [153] разделили работы по типам объекта исследований: мошенничество с банковскими продуктами (рассматриваются только внешние мошенники, не банки) – 8 статей; мошенничество с продуктами страховых компаний – 24 статьи; мошенничество с ценными бумагами и товарами – 1 статья; другие виды мошенничества – 17 статей. Эмпирических статей, рассматривающих мошенничество с ценными бумагами и товарами, мало из-за отсутствия данных. Среди российских работ по идентификации мошеннических операций следует выделить исследования [154, 155], где оценивается вероятность и размеры «дыры» капитала в обанкротившихся

и еще здоровых российских банках за период с 2013-2016 с использованием модели Хекмана.

Мы обратили в большей степени внимание на теоретические работы, которые рассматривают мошеннические операции в контексте влияния таких операций на финансовую стабильность. Наиболее значимыми являются работы [156, 157, 158, 159, 152, 160].

Статья [156] показывает на примере Израильского банковского кризиса 1983, как условия в банковской системе провоцируют банки на мошеннические операции и как такое поведение приводит к значительным потерям (около 40% ВВП). Возникновение Израильского банковского кризиса 1983 года было обусловлено следующими условиями, созданными в системе. Закон о ценных бумагах (1981) и освобождение от налога на оборот позволяли банкам покупать и продавать собственные акции, повышая таким образом цену акций. Банки участвовали в данном виде мошенничества с целью обеспечить капитал выше минимального нормативного требования, что было сделать значительно труднее при высокой инфляции. Последствия данного кризиса оказались серьезными (падение рыночной капитализации на 60%, что соответствует 40% ВВП Израиля), так как банковский сектор характеризовался высокой концентрацией (все 6 крупных банков, составляющих 80% банковской системы, понесли высокие потери при падении цен на акции). Из-за закрытого рынка капитала Израиля все последствия кризиса приняла внутренняя экономика Израиля.

В [157] рассматривается, как сокрытие проблемных активов влияет на принятие рисков банками и на банковские кризисы, в данной работе разрабатывается динамическая модель финансового посредничества. [157] строит динамическую модель финансового посредничества, в рамках которых существует четыре типа активов, различающихся по сроку жизни, уровню риска (дефолтные и недефолтные) и графиком платежей. Распределение активов в кредитном портфеле банков задается случайным образом в каждом временном периоде. Модель предусматривает аудит банков со стороны регулятора,

где результаты аудита приводят к ликвидации банка в случае обнаружения у него скрытых проблемных активов. Результаты исследования показывают, как банки изменяют кредитную политику, уровень морального риска в зависимости от частоты мониторинга регулятора.

В [158] продолжают рассматривать модель общего равновесия с включением возможности дефолта, которую впервые ввели Shubik и Wilson в 1977 (цит. по [158]). В [158] рассматривают, при каких условиях заемщику не выгодно инициировать процесс бесконечного перекредитования с целью покрыть прошлые кредиты (схема Понци). Авторы отмечают, если штрафы за дефолт очень высокие, вероятность схемы Понци значительно увеличивается.

В [159] изучает, какие факторы определяют доверие людей к схемам Понци. В частности, в [159] строится эмпирическая модель, которая затем проверяется на выборке 402 инвесторов Ямайки.

В [152] рассматривается финансовый сектор Великобритании, который являлся одним из секторов с высокой долей экономических преступлений с 1970-х годов.

В [160] отмечают, что падение финансовых пирамид наносит не только прямые потери обществу, но и косвенные, увеличивая уровень преступности в целом.

Таким образом, можно заключить, что большая часть работ направлена на создание методов идентификации мошеннических операций. Меньшее внимание уделено вопросам причин усиления мошеннической деятельности и ее влияние на финансовую устойчивость системы. Предлагаемое исследование обратит внимание на вторую часть незатронутых вопросов, а именно будут определены критические уровни мошеннической деятельности, которые делают финансовую систему уязвимой.

3.2.3 Методология исследования

Операцию будем считать мошеннической, если одновременно выполняются следующие условия:

- операция произведена банком преднамеренно;
- операция противоречит требованиям регулятора;
- совершение операции причинило крупный ущерб гражданам, организациям или государству.

На текущий момент в работе выделено три основных вида мошеннических операций: (1) финансовая пирамида, (2) неучет пассивов, (3) скрытый риск. Далее будут разобраны данные типы мошеннических операций банков и представлены способы их внедрения в модель.

Банк, который реализует финансовую пирамиду, принимает только вклады и не выдает кредиты. Ставка по вкладам может отличаться от рыночной. В рамках данного типа мошеннических операций могут быть рассмотрены случаи перекредитования (часть агентов берет кредит в сумме к возврату ранее взятого кредита и начисленных по нему процентов). При данном виде мошенничества банки больше испытывают потребность в МБК кредитах и более подвержены другим типам мошеннических операций (например, созданию фиктивного капитала).

Банк, который скрывает часть своего риска, из-за отсутствия реального капитала покрывает свой риск за счет средств кредиторов банка. Недостаток ликвидности покрывается межбанковскими займами. Таким образом, возможен эффект заражения в системе.

Неучет пассивов отражается в том, что банки несут меньше расходов по страхованию депозитов (возможно увеличение прибыли за счет этого). В данном сценарии мошенничества оценивается устойчивость системы при текущих ставках Агентства по страхованию вкладов.

Долю банков с каждым видом мошеннических операций задается пользователем. Банк может участвовать сразу в нескольких видах мошенничеств.

Для внедрения всех типов мошеннических операций банков требуется создание двойных профилей банков. Первый профиль – отчетность, которую банк раскрывает другим участникам рынка (может быть фальсифицированной у банка-мошенника). Второй профиль – отчетность по факту. Если банк не проводит мошеннические операции, то первый профиль идентичен второму; иначе – у банка-мошенника. Второй профиль может быть обнаружен только в случае проведения аудита. Во втором профиле возможен отрицательный капитал и невыполнение нормативов, которые не приводят к банкротству банка до проведения аудита.

3.2.4 Результаты

На данный момент в агентно-ориентированную модель, представленную на конференциях МАСЭП 2016 г. и 2017 г. исследовательской группой в составе Леонидов А.В., Нечитайло В.А., Пеникас Г.И., Пильник Н.П. добавлены следующие модули:

- финансовая пирамида как одна из возможных стратегий банка;
- неучет пассивов как одна из возможных стратегий банка;
- реализация дефолтов в системе;
- расчет взвешенных по риску активов и их использование в принятии решения банками при одобрении кредита;
- регулирование банков в виде норматива достаточности капитала и требования создавать резервы под возможные потери.

Программный модуль создается на программном языке Delphi.

4 Разработка и применение методов анализа данных и прогнозирования

В этом разделе отчета отражены результаты по разработке и применению методов анализа данных и прогнозирования. Работа велась в следующих направлениях:

- исследование по сетевому анализу данных;
- анализ сообщества экономистов и построение рейтингов экономических журналов;
- развитие методов анализа паттернов;
- разработка нового частотного механизма оценивания структурной сложности графов;
- исследования по кластерному анализу.

4.1 Исследование по сетевому анализу данных

4.1.1 Введение

Исследование посвящено выявлению влиятельных элементов в сетевых структурах. Работа является продолжением прошлых исследований по сетевому анализу и включает в себя новые приложения ранее разработанных индексов, модернизацию и снижение сложности предложенных алгоритмов, а также новые методы учёта дальних взаимодействий между элементами сети.

Из литературы известно большое количество индексов центральности, оценивающих степень влияния вершин сети. Некоторые из них основаны на степенях вершин [161], степенях их соседей [162, 163, 164], на поиске кратчайших путей между узлами сети или степени посредничества различных вершин [165, 166, 167]. В основе каждого из этих индексов лежит своя идея, какие именно элементы сети считать центральными, а общее число этих индексов исчисляется сотнями.

К сожалению, существующие методы не в полной мере оценивают интенсивность дальних взаимодействий и зачастую учитывают незначимые связи между элементами сети в силу того, что исходный набор вершин, а также структура связей между ними могут быть неоднородными. Влияние на одну вершину может цениться больше, чем влияние на другую вершину, что также зависит как от структуры сети, так и от характеристик самих вершин. Наконец, влияние на вершины сети может осуществляться в совокупности с другими вершинами (групповое влияние), а может осуществляться и косвенно, то есть через другие вершины. Все эти особенности, которые регулярно возникают в реальных задачах, делают невозможным применение известных в литературе методов определения ключевых вершин.

В [168, 170, 169] были предложены новые модели оценки влияния в сетях (SRIC и LRIC), отличительной особенностью которых является то, что они учитывают индивидуальные характеристики вершины, возможность группового влияния, а также влияние через промежуточных посредников. Несмотря на то, что индексы SRIC и LRIC учитывают данные особенности сети, основным их недостатком является высокая вычислительная сложность. В рамках этой работы была рассмотрена проблема вычислительной сложности индексов SRIC и LRIC, а также был предложен новый алгоритм расчета индексов дальних взаимодействий LRIC, который позволяет определять ключевые элементы сети за меньшее количество времени. Эффективность алгоритма проверена на основе вычислительных экспериментов, в рамках которых рассматривались графы с разным числом вершин и ребер.

Помимо снижения сложности ранее разработанных моделей был применен новый подход по учёту дальних взаимодействий между элементами. Так, подход, основанный на случайных блужданиях в графе, показал хорошо коррелированные результаты с индексами дальних взаимодействий LRIC, при этом вычислительная сложность этого подхода ниже, чем подход, основанный на переборе всех простых путей между вершинами. В частности, в работе был использован подход, основанный на вычислении индивидуального

PageRank [171] для каждой из вершин, после чего была составлена матрица попарных взаимодействий между вершинами, которая может быть агрегирована в вектор влияния или вектор зависимости.

В работе также проводится анализ влияния стран на различных уровнях взаимодействия, таких как, например, международные миграционные отношения и экспортно-импортные отношения в сфере торговли продовольствием.

Международная миграция является важным аспектом межстрановых взаимодействий, а также социальным феноменом, влияющим на многие стороны жизни общества. Международная миграция между странами была исследована во многих работах, в том числе с помощью сетевого подхода (см. обзор литературы в [172]). Данная работа продолжает исследование сетевого анализа международной миграции на основе новой, расширенной базы данных [1]. В рамках данного анализа произведена оценка влияния стран в сети международной миграции по периодам с 1970 по 2014 гг.

Межстрановые отношения в контексте торговли продовольствием также являются неотъемлемым процессом экономических, социальных и политических взаимодействий между государствами и территориями. В рамках данного направления были проанализированы известные статистические показатели продовольственного баланса стран, отражающих уровень запасов, производства, потерь и прочих показателей по конкретным видам продовольствия, а также рассчитаны индексы влияния и зависимости от импорта на сетях общей торговли и основных продуктовых категорий. Корреляционный анализ индексов показал, что влияние в сети по общей торговле не в полной мере отражает реальное влияние стран, так как в каждой продуктовой категории выявляются свои лидеры, отличные от лидеров в других категориях.

Другим направлением по исследованию сетей является анализ террористических организаций и их взаимосвязей. С момента терактов 11 сентября 2001 года терроризм стал важным фактором мировой политики, а террористические группы зачастую претендуют на то, чтобы быть самостоятельным

актором на международной арене. Крупные и известные террористические организации как Исламское государство, Талибан и Аль-Каида являются яркими тому примерами. Терроризм включил в повестку дня вопросы безопасности во многих странах мира, в связи с чем можно говорить о безусловном влиянии террористических групп на политические процессы и принятие государственных решений.

При этом многие террористические организации в целях собственной безопасности не имеют какого-либо территориального центра и явной структуры, в связи с чем оценка их силы и влияния представляется затруднительной.

В научной литературе тема терроризма освещена широко; отдельным предметом интереса исследователей являются террористические сети. При этом стоит отметить, что подавляющее большинство научных исследований в данной области посвящены сетевым структурам отдельных террористических организаций, где в фокусе внимания находятся связи между террористами-членами организации. С другой стороны, связи между террористическими группами как акторами сети рассматриваются исследователями крайне редко, несмотря на то, что сеть террористических групп представляет собой важный коммуникационный и инфраструктурный канал для организаторов террористических атак.

4.1.2 Описание индексов взаимодействия

Рассмотрим граф $G = (V, E, W)$, $V = \{1, \dots, n\}$ – множество вершин, $|V| = N$, $E \subseteq V \times V$ – множество ребер, а $W = \{|w_{ij}|\}$ – множество весов – действительных чисел, соответствующих каждому ребру $(i, j) \in E$. Дополнительно каждая вершина характеризуется пороговым значением q_i , при достижении которого вершина i подвержена влиянию других вершин.

Обозначим также через N_i набор вершин, имеющих прямую связь с вершиной i . Тогда группа вершин $\Omega(i) \subseteq N_i$ является критической для вер-

шины i , если $\sum_{j \in \Omega(i)} w_{ij} \geq q_i$, а вершина $l \in \Omega(i)$ – ключевой, если ее исключение из данной группы делает группу некритической. Набор ключевых вершин в группе $\Omega(i)$ обозначим через $\Omega_p(i)$.

Наконец, простым путем между вершинами i и j в графе G будем считать последовательность ребер, которая соединяет данные вершины, причем весь набор вершин на пути должен быть уникальным. Другими словами,

$$(i, k_1), (k_1, k_2), (k_2, k_3), \dots, (k_{s-1}, j), \quad i \neq k_1 \neq \dots \neq k_{s-1} \neq j$$

называется простым путем, который можно обозначить через $P_{i-j}^t(s)$, где t – номер простого пути.

Вычисление индексов ближних и дальних взаимодействий (SRIC и LRIC соответственно) может быть поделено на несколько этапов.

4.1.2.1 Индекс ближних взаимодействий (SRIC)

На первом этапе происходит определение того, как отдельная вершина сети подвержена влиянию других участников. Для индекса ближних взаимодействий происходит оценка прямых и непрямых интенсивностей взаимодействия между вершинами, рассчитанная по следующим формулам

$$p_{ij} = \frac{w_{ij}}{\sum_k w_{kj}}$$

– интенсивность прямого влияния вершины i на вершину j ,

$$p_{ihj} = \begin{cases} \frac{w_{ih}}{\sum_k w_{kj}}, & \text{если } w_{ij} > 0, \quad w_{hj} > 0, \quad w_{hj} \geq w_{ih}, \quad i \neq h \neq j, \\ \frac{w_{hj}}{\sum_k w_{kj}}, & \text{если } w_{ij} > 0, \quad w_{hj} > 0, \quad w_{hj} < w_{ih}, \quad i \neq h \neq j, \\ 0, & \text{в ином случае.} \end{cases}$$

После данного этапа для индекса ближних взаимодействий происходит рассмотрение всех возможных критических групп, где вершина i влияет на вершину j , а также оценка вклада вершины i в данную группу, рассчитанная по формуле

$$f(i, \Omega_k(j)) = \frac{p_{ij} + \sum_{h \in \Omega_k(j)} p_{ihj}}{|\Omega_k(j)|}, \quad i \in \Omega_k^p(j).$$

После рассмотрения всех возможных критических групп для каждой вершины и вычисления общего влияния вершины i на вершину j через все возможные критические группы ($\chi_i(j) = \sum_k f(i, \Omega_k(j))$), индекс ближних взаимодействий нормирует влияние на каждую вершину по формуле

$$\hat{\chi}_i(j) = \frac{\chi_i(j)}{\sum_k \chi_k(j)},$$

после чего преобразует данную информацию в индекс ближних взаимодействий (SRIC) с учетом важности влияния на отдельные вершины.

4.1.2.2 Индекс дальних взаимодействий (LRIC)

Аналогично индексу ближних взаимодействий, основной целью первого этапа вычисления индекса LRIC является расчет влияния участников сети на каждую индивидуальную вершину. Тем не менее, основной отличительной особенностью данного индекса заключается в том, что определение влияния вершины i на вершину j заключается в поиске такой критической группы $\Omega_k(j)$, в которой вершина i имеет максимальное влияние (см. формулу (4.1)).

$$c_{ij} = \begin{cases} \max_{\Omega_k(j): i \in \Omega_k^p(j)} \frac{w_{ij}}{\sum_{h \in \Omega_k(j)} w_{hj}}, & \text{если } \exists \Omega_k^p(j): i \in \Omega_k^p(j), \\ 0, & \text{в ином случае.} \end{cases} \quad (4.1)$$

Значение c_{ij} показывает, как вершина i влияет на вершину j напрямую. Найденные значения можно представить в виде матрицы прямых влияний $C = [c_{ij}]_{n \times n}$.

Тем не менее, так как вершины могут влиять друг на друга через какие-то промежуточные вершины, на втором этапе расчета индекса производится оценка непрямого влияния между вершинами.

Рассмотрим все возможные простые пути между вершинами. Тогда оценка непрямого влияния вершины i на вершину j через путь $P_{i-j}^t(s)$ определяется с учетом всех прямых влияний следующим образом:

$$f_{mult}(P_{i-j}^t(s)) = c_{ik_1(t)} \times c_{k_1(t)k_2(t)} \times \dots \times c_{k_{s-1}(t)j}, \quad (4.2)$$

Таким образом, согласно формуле (4.2) совокупное влияние элемента i на элемент j рассчитывается с учетом всех прямых влияний между элементами на этом пути.

В случае если между двумя вершинами присутствует большое число путей, агрегирование их для построения итоговой оценки влияния вершины i на вершину j может осуществляться разными способами. Например, общее влияние может быть рассчитано как сумма всех возможных путей, через которые одна вершина влияет на другую

$$c_{ij}^*(s) = \min \left\{ \sum_{k: |P_k^{ij}| \leq s} f(P_k^{ij}), 1 \right\}. \quad (4.3)$$

Комбинируя формулы (4.2) и (4.3) можно получить индекс общего влияния, обозначаемый как LRIC (Sum). После расчета попарного общего влияния, данная информация агрегируется в индекс дальних взаимодействий с учетом важности влияния на ту или иную вершину.

4.1.3 Вычислительная сложность индексов SRIC и LRIC

Основным недостатком индексов ближних и дальних взаимодействий является их высокая вычислительная сложность. Индекс ближних взаимодействий при расчете прямого влияния требует рассмотрения всех возможных критических групп. В общем случае, если вершина j имеет порядка n соседей, общее количество рассматриваемых групп достигает $2^n - 1$. Соответственно, в случае, если исходный граф состоит из нескольких десятков или сотен вершин, а количество связей между ними близко к количеству ребер в полном графе, вычисление индекса SRIC не представляется возможным.

Что касается индекса дальних взаимодействий, то в отличие от индекса SRIC, индекс LRIC требует нахождения такой группы, где вершина k имеет максимальное влияние на другую вершину. Данная особенность позволяет существенно снизить вычислительную сложность индекса дальних взаимодействий.

4.1.3.1 Определение прямого влияния индекса LRIC

Рассмотрим прямое влияние вершины i на вершину j . Так как в общем случае необходимо рассмотреть все возможные группы вершин, то обозначим через $x_l = 1$ присутствие вершины l в рассматриваемой группе и через $x_l = 0$ ее отсутствие в группе.

В таком случае задачу поиска наибольшего влияния вершины i на вершину j можно представить в следующем виде

$$f(x_1, \dots, x_{|N_j|}) = \sum_l w_{lj} \cdot x_l \rightarrow \min$$

при ограничениях

$$\sum_l w_{lj} \cdot x_l \geq q_j - w_{ij}$$

$$\sum_l w_{lj} \cdot x_l < Q_{kj}$$

$$\forall l \neq i \ x_l = \{0,1\},$$

где $Q_{kj} = q_j$ на первом этапе.

Данная задача является задачей целочисленного линейного программирования. Следовательно, решив данную задачу, можно определить наибольшее прямое влияние вершины k на вершину j .

Наконец, обратим внимание на некоторые свойства ключевых участников в критических группах. Расположим элементы k из множества N_j в порядке возрастания веса w_{kj} и запишем их в множество $N'_j = \{1,2,3, \dots, |N_j|\}$.

Другими словами, $\forall i. i + 1 \in N_j' w_{ij} \leq w_{i+1 j}$. Нетрудно заметить из предыдущих двух пунктов, что если $i \in N_j'$ не является ключевой, то любая вершина $l \in N_j' l < i$ также не является ключевой. Наоборот, если вершина $i \in N_j'$ является ключевой, то любая вершина $l \in N_j' i < l$ также является ключевой в группе. Более того, если $i \in N_j'$ не является ключевой ни в одной из групп, то никакая минимальная критическая группа не содержит вершину i . Данная особенность позволяет резко сократить число рассматриваемых вершин, которые могут влиять на вершину j .

Таким образом, если представить задачу нахождения наибольшего прямого влияния между вершинами в виде задачи целочисленного линейного программирования, а также осуществлять вычисление прямого влияния на основе бинарного поиска, можно существенно снизить вычислительную сложность индекса LRIC.

4.1.3.2 Определение непрямого влияния индекса LRIC

Основной проблемой вычисления индексы LRIC (Sum) является необходимость всех возможных простых путей в графе. В худшем случае, существует $\sum_{l=1}^{n-2} \left(\frac{(n-2)!}{(n-2-l)!} \right)$ простых путей между двумя вершинами в графе и $\frac{n \cdot (n-1)}{2} \cdot \sum_{l=1}^{n-2} \left(\frac{(n-2)!}{(n-2-l)!} \right)$ простых путей в общем случае. Данная особенность достаточно существенна и не позволяет производить расчет индекса ближних взаимодействий на графах большого размера, обладающих свойством высокой плотности.

Одним из способов решения данной проблемы является адаптация алгоритма поиска в глубину (DFS) [173], который рассматривает каждое ребро в графе только 1 раз, с возможностью хранения информации о силе того или иного участка пути. Тем не менее, данный способ решения задачи накладывает дополнительные требования на объем используемой памяти для решения задачи.

Альтернативным способом решения задачи является операция матричных умножений. Напомним, что если представить граф в виде матрицы связности $A = [a_{ij}]_{n \times n}$, где $a_{ij} = 1$ при $(i, j) \in E$ и $a_{ij} = 0$ в ином случае, то значение $[A^k]_{ij}$ будет показывать общее число путей длины k между вершинами i и j . В силу того, что индекс LRIC (Sum) требует рассмотрения всех возможных путей между вершинами длины k , нахождение значения $[C^k]_{ij}$ позволит определить влияние вершины i на вершину j через все возможные пути длины k .

К сожалению, результат умножения матрицы прямых влияний в степень k включает помимо простых путей пути, содержащие циклы. Например, $[C^2]_{ij}$ будет включать следующий путь с циклом (см. рисунок 4.1).



Рисунок 4.1 – Путь с циклом длины 2

Таким образом, для расчета попарного влияния вершин через все возможные простые пути из общего значения $[C^k]_{ij}$ необходимо удалить информацию о путях, которые содержат циклы. Для случая $k=2$ общее влияние $c_{ij}(k)$ вершины i на вершину j через все возможные пути длины k можно вычислить через

$$c_{ij}(2) = \begin{cases} [C^2]_{ij}, & \text{если } i \neq j, \\ 0, & \text{в ином случае.} \end{cases}$$

При $k=3$ существует 2 возможных типа путей с циклами (см. рисунок 4.2).

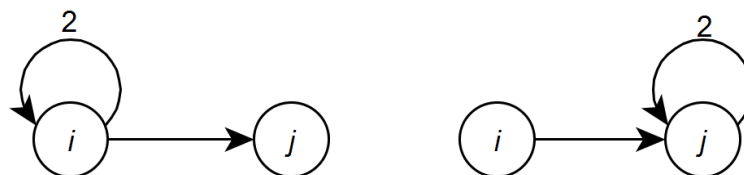


Рисунок 4.2 – Путь с циклом длины 3

Общее влияние каждого пути с циклом из рисунка 4.2 можно определить как

Тип 1: $[C^2]_{ii} \cdot c_{ij}$;

Тип 2: $c_{ij} \cdot [C^2]_{jj}$.

Обратим внимание, что данные типы путей с циклами пересекаются между собой. На рисунке 4.3 представлен путь с циклом, который входит в каждый из типов на рисунке 4.2

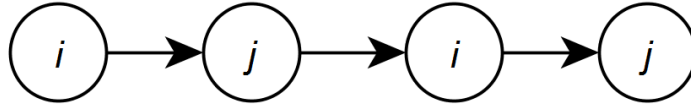


Рисунок 4.3 – Путь с циклами, который входит в тип 1 и тип 2

Таким образом, интенсивность всех возможных простых путей между вершинами i и j длины 3 можно рассчитать по формуле

$$c_{ij}(3) = \begin{cases} [C^3]_{ij} - ([C^2]_{ii} + [C^2]_{jj}) \cdot c_{ij} + (c_{ij})^2 \cdot (c_{ji})^2, & \text{если } i \neq j, \\ 0, & \text{в ином случае.} \end{cases}$$

Аналогично, можно представить пути длины 4, 5 и более. В рамках нашей работы был разработан алгоритм, который позволяет определить вид формулы для вычисления интенсивности влияния $c_{ij}(k)$ длины k . Следовательно, так как индекс дальних взаимодействий рассматривает все возможные простые пути длины не более s , итоговое попарное влияние можно считать через

$$c_{ij}^*(s) = \sum_{k=1}^s c_{ij}(k).$$

Таким образом, вычисление не прямых влияний можно привести к операциям с матрицами, что сильно позволяет снизить вычислительную сложность данной операции. Стоит отметить, что оба предложенных метода вычисления прямых и не прямых влияний не могут быть адаптированы к вычислению индекса ближних взаимодействий (SRIC).

4.1.4 Вычислительные эксперименты

Для оценки вычислительной сложности алгоритма и сравнения его с предыдущей версией, проведем вычислительный эксперимент. Сгенерируем два типа графов:

- полный граф;
- граф с экспоненциальным распределением количества ребер.

Для сравнения эффективности предложенных подходов было рассмотрено три версии реализации алгоритма дальних взаимодействий:

- $LRIC_{naive}$: существующий алгоритм индекса LRIC (рассмотрение всех критических групп и возможных путей между вершинами);
- $LRIC_{direct}$: алгоритм индекса LRIC с оптимизированным вычислением прямого влияния;
- $LRIC_{opt}$: алгоритм индекса LRIC с оптимизированным вычислением прямого и непрямого влияния.

Сравнение алгоритма на полном графе позволит понять, насколько эффективен алгоритм при большом числе соседних вершин. Сравнение на графе с экспоненциальным распределением позволит определить, как изменяется его вычислительная сложность при увеличении числа вершин при относительно небольшом числе соседей.

Результаты вычисления индекса LRIC на полном графе размера от 10 до 60 вершин приведены на рисунке 4.4.

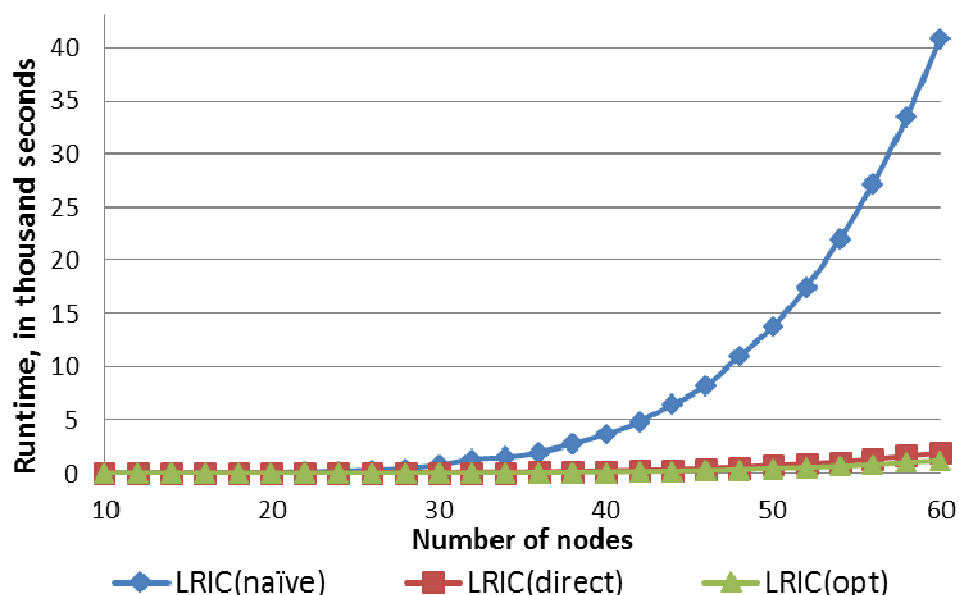


Рисунок 4.4 – Вычислительный эксперимент на полном графе

Как показано выше, время вычисления алгоритмом $LRIC_{naïve}$ на полных графах растет экспоненциально. Если на графах размера до 20 вершин время вычисления составляло менее 1 минуты, то на графах размера до 30 вершин время выполнения составляет до 10 минут. При увеличении числа вершин до 40 вершин время выполнения достигает 1 часа на обычном компьютере, при размере до 60 вершин расчеты занимают до 10 часов. Тем не менее, обратим внимание, что алгоритм $LRIC_{direct}$ позволяет в среднем в 14 раз сократить вычисление индекса дальних взаимодействий. Что касается алгоритма $LRIC_{opt}$, то данный алгоритм дает улучшение примерно на 100% по сравнению с алгоритмом $LRIC_{direct}$. Таким образом, можно заметить, что благодаря предложенным моделям время расчета индекса дальних взаимодействий на полных графах размера 60 вершин можно сократить с 10 часов, до 20 минут.

Обратим внимание, что алгоритм $LRIC_{opt}$ может вычисляться параллельно. На рисунке 4.5 показано, насколько снижается время алгоритма $LRIC_{opt}$ при параллельных вычислениях на полных графах размера от 50 до 100 вершин (в среднем – в 3 раза).

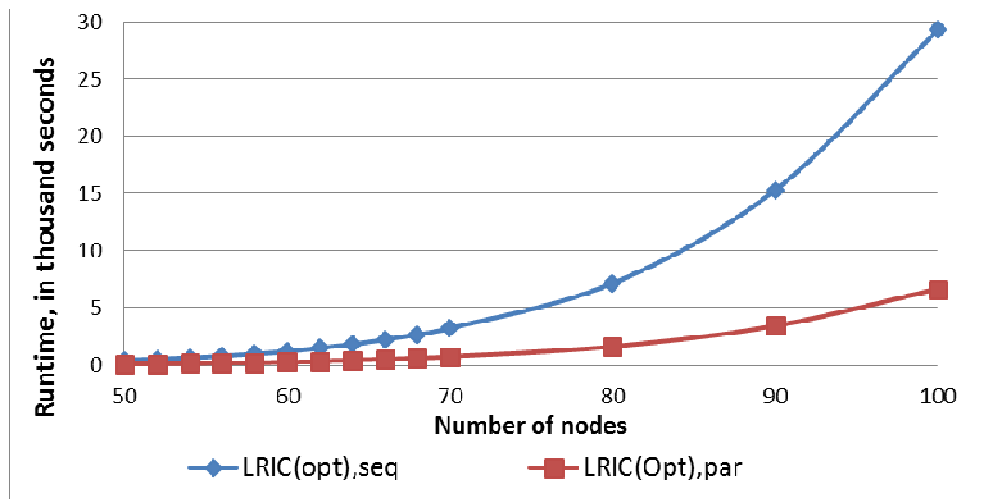


Рисунок 4.5 – Время алгоритма $LRIC_{opt}$ при параллельных вычислениях

Результаты вычисления индекса LRIC на графе с экспоненциальным распределением степеней вершин размера от 50 до 1200 вершин приведены на рисунке 4.6.

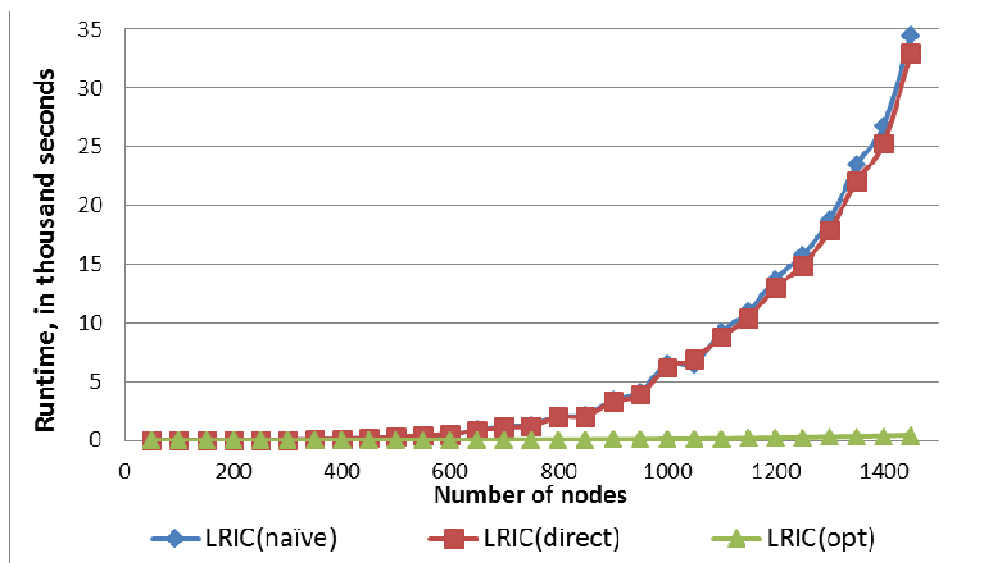


Рисунок 4.6 – Вычислительный эксперимент на графе с экспоненциальным распределением

Как показано выше, время вычисления обоим алгоритмом растет экспоненциально. Так как в экспоненциальных графах количество вершин, обладающих большим числом соседей относительно небольшое, преимущества алгоритма вычисления прямых влияний по сравнению с предыдущим неочевидны. Если графы размера до 400 вершин вычисляются менее чем за минуту, то вычисление индекса LRIC на графах размера до 1000 вершин требует

около часа. При увеличении числа вершин до 1200 время вычисления достигает уже 2 часов. В среднем, время выполнения алгоритма $\text{LRIC}_{\text{direct}}$ меньше времени старого на 1,5%. Общая высокая вычислительная сложность алгоритмов $\text{LRIC}_{\text{naive}}$ и $\text{LRIC}_{\text{direct}}$ в первую очередь может быть объяснена тем, что основной проблемой в данной задаче является рассмотрение различных не-прямых путей между вершинами, что не решается данными алгоритмами.

В отличие от других алгоритмов, алгоритм LRIC_{opt} решает задачу вычисления непрямого влияния. Графы размера 900 узлов вычисляются менее чем за 1 минуту, в то время как графы размера 1000 и 1500 вершин требуют 2 и 7 минут соответственно. Таким образом, наблюдается существенное снижение сложности при вычислении индекса дальних взаимодействий с помощью алгоритма LRIC_{opt} .

Стоит отметить, что предложенные алгоритмы реализованы в качестве библиотеки на языке R.

4.1.5 Оценка непрямого влияния в сетевых структурах

В предыдущем разделе была рассмотрена модель, учитывающая дальние взаимодействия между вершинами, основанная на рассмотрении всевозможных простых путей между вершинами. В частности был рассмотрен подход, основанный на перемножении прямых влияний на каждом из простых путей между вершинами (см. формулу (4.2)). Такой подход схож с вычислением совместной вероятности некоторых событий, а именно с одновременным влиянием всех вершин в цепочке друг на друга. Данное предположение и последующее вычисление влияния даёт верхнюю оценку влияния начальной вершины на конечную вершину в рассматриваемой цепочке в терминах вероятностей.

Другой подход, основанный на рассмотрении простых путей между вершинами, также описан в работе [169]. Этот подход связан с вычислением минимального влияния в рассматриваемой цепочке, а именно

$$f_{\min}(P_{i-j}^t(s)) = \min(c_{ik_1(t)}, c_{k_1(t)k_2(t)}, \dots, c_{k_{s-1}(t)j}).$$

Данный подход схож с вычислением пропускной способности через рассматриваемый путь (эффект бутылочного горлышка).

Как было показано ранее, вычислительная сложность данных моделей оценки непрямого влияния достаточно высока. Для вычисления попарного влияния необходимо перебирать всевозможные простые пути между всеми парами вершин.

Существует ряд других подходов оценки непрямого влияния в сетевых структурах. В задачах, где важно только итоговое ранжирование вершин и не требуется искать попарное влияние элементов друг на друга, применимо вычисление классических мер центральностей на графе прямого влияния. При этом для учёта дальних взаимодействий необходимо использовать соответствующие меры центральности (Katz, eigenvector, PageRank и другие, учитывающие длинные цепочки в графах). Но если нас также интересует общее попарное влияние между вершинами, подход, основанный на вычислении классических мер центральностей, будет неприменим.

В случае, когда важно получить информацию о попарном влиянии элементов друг на друга, может быть использован подход оценки непрямого влияния, основанный на случайных блужданиях в графе. В частности, может быть использован подход на основе оценки индивидуального PageRank для каждой из вершин. Напомним, что классический индекс центральности PageRank выявляет элементы, которые с более высокой вероятностью встречаются на путях при случайном блуждании по графу. Точнее говоря,

$$PR(i) = d \left(\sum_j \frac{PR(j)}{\sum_k w_{jk}} \right) + (1 - d)E,$$

где d – вероятность продолжать блуждания по текущему пути, $1 - d$ – вероятность начать блуждания с новой вершины, $E = [E_i]$ – вектор вероятностей, $E_i = \frac{1}{n} \forall i$, n – общее число вершин в графе. Данная формула комбинирует

подход случайных блужданий (первое слагаемое) и случайные переходы к новым вершинам, с которых начинается новый процесс блужданий. При этом в классической версии индекса переходы к новым вершинам совершаются с одинаковой вероятностью. Отличие индивидуального PageRank от классического заключается в том, что индивидуальный PageRank допускает разные вероятности переходов к каждой из вершин. Другими словами, вектор E может содержать разные вероятности для каждого из элементов. Естественно, центральность элементов возрастает, если вероятность случайного перехода в них выше, чем в другие вершины. При этом максимальное значение центральности, которое может получить вершина, достигается при вероятности перехода в неё равной единице. Напомним, что общая сумма вероятностей переходов равна единице, поэтому при вероятности перехода в какую-то из вершин равной единице, все остальные вершины получают нулевую вероятность перехода в них. Поэтому, если мы хотим оценить влияние одной вершины на все остальные вершины, мы можем вычислить индивидуальный PageRank для каждой вершины с максимальной вероятностью перехода в неё. Таким образом, для каждой вершины может быть вычислен вектор влияния на остальные вершины, что в конечном итоге может быть перестроено в матрицу общего попарного влияния вершин друг на друга.

Отметим, что сложность случайных блужданий в графах ниже, чем перебор всех простых путей в графе.

4.1.6 Анализ международной миграции

Для анализа международной миграции были использованы базы данных по потокам миграции между странами мира. Комбинирование различных источников данных сделано, прежде всего, с целью наиболее полно отобразить потоки международной миграции, а именно: увеличить количество стран, вовлеченных в процесс международной миграции за период с 1970 по 2014. В первых, в работе были использованы данные ООН [174, 175] по странам ми-

ра. Во-вторых, данные по потокам международной миграции ОЭСР и Евросоюза [176, 177].

Сетевой анализ миграционных потоков был проведен на данных по потокам международной миграции за период с 1970 по 2014. В течение указанного периода в миграции между странами происходили различные изменения. В работе был осуществлен анализ суммарного миграционного оттока из стран (рисунок 4.7), миграционного притока (рисунок 4.8) и оттока по регионам (рисунок 4.9). Данный анализ позволил выделить наиболее существенные изменения, произошедшие в международной миграции с 1970 по 2014. В работе были выделены следующие исторические события, которые оказали влияние на потоки миграции между странами.

В Европе к данным событиям относятся окончание программы трудовой миграции как следствие энергетического кризиса в 1973 г и последующее изменение характера иммиграции. Территориальные изменения в составе стран ЕС также оказывали влияние на потоки миграции в Европу. Однако особенно сильно на изменение миграционной ситуации повлияло объединение Германии (1990) и распад СССР (1991) [178]. Последнее также повлияло на иммиграцию в США и Канаду. Всплеск в иммиграции в США после 1987 г. объясняется миграционной амнистией, т.е. переучетом нелегальных мигрантов, уже находящихся в стране, преимущественно из стран Латинской Америки [179]. После 2000г. наблюдался дальнейший рост миграции из Мексики в США (рисунок 4.8). Финансовый кризис 2008 г. вызвал возвратную миграцию в Европе и США, ввиду снижения экономической привлекательности принимающих стран.

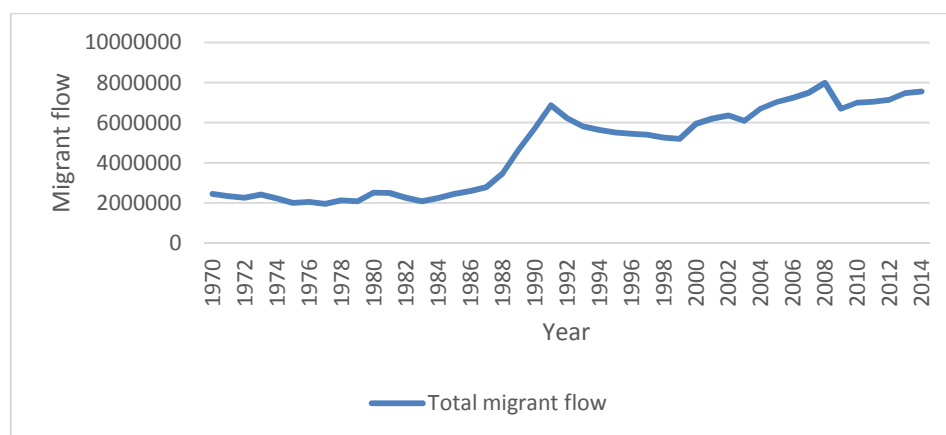


Рисунок 4.7 – Суммарный поток международной миграции с 1970 по 2014

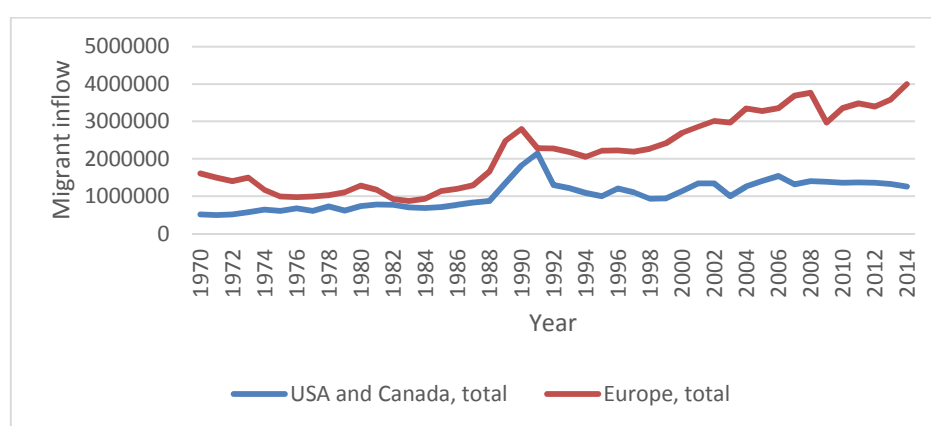


Рисунок 4.8 – Суммарный приток международной миграции по регионам с 1970 по 2014

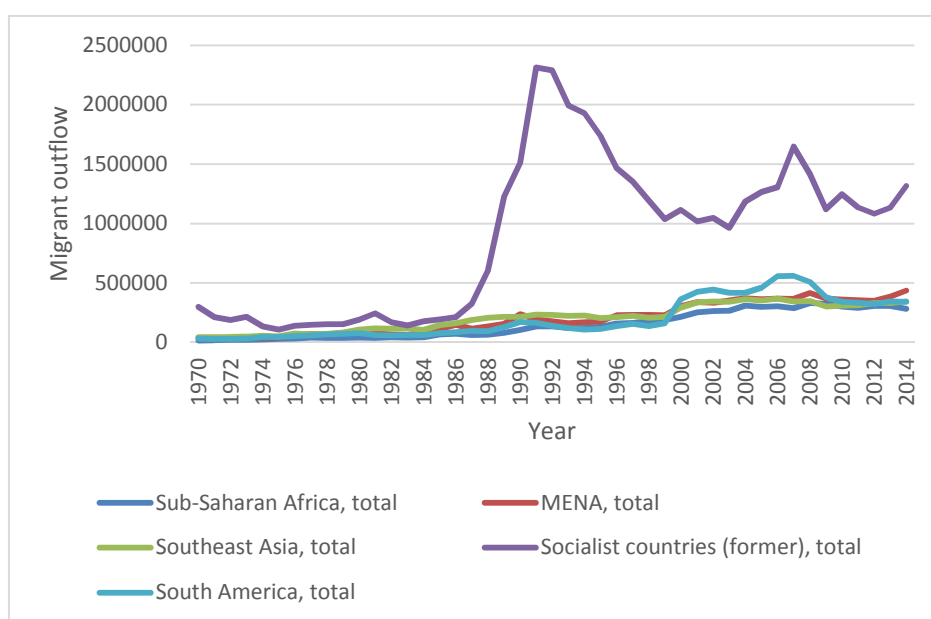


Рисунок 4.9 – Суммарный отток международной миграции по регионам с 1970 по 2014

Таким образом, для более корректного анализа временной промежуток был разбит на периоды: 1970-1987, 1988-1991, 1992-1999, 2000-2008, 2009-2014. Международная миграция в каждом периоде была проанализирована с помощью сетевого анализа как классическими мерами центральности, так и новыми, учитывающими не прямые взаимодействия между странами, групповое влияние и население в качестве индивидуальной характеристики стран [169].

В Таблице 4.1 приведены результаты анализа за 2014г. Данный период интересен тем, что он предшествовал миграционному кризису в Европе. Результаты классических индексов показывают, что Германия была на первом месте по притоку и оттоку мигрантов (InWdeg и OutWdeg соответственно). Наибольший приток иммигрантов также наблюдался в США, Великобритании и Испании. По оттоку мигрантов (OutWdeg) лидировали Китай, Румыния, Испания и Польша. Центральности EV и PR выделили такие страны, как Румыния, Польша, Япония и Китай. Данные страны были связаны в сети международной миграции со странами с наибольшим оттоком или притоком мигрантов.

Результаты, полученные с помощью индексов ближних (SRIC) и дальних (LRIC) взаимодействий позволяют обнаружить новые страны, обладающие значительным влиянием в процессе международной миграции. К примеру, среди стран, обнаруженных только с помощью индексов LRIC и SRIC можно выделить Сирию, Индию, Вьетнам и Филиппины.

Таблица 4.1 – Рейтинг влияния стран на основе индексов центральности

Страна	InW _{deg}	OutW _{deg}	EV	PR	LRIC (Max)	SRIC
Германия	1	1	1	1	8	5
США	2	13	14	11	14	15
Великобритания	3	8	11	8	16	11
Республика Корея	4	6	6	4	44	20
Испания	5	4	5	2	5	10
Япония	6	9	12	3	48	54
Италия	7	12	8	12	7	12
Китай	8	2	4	5	6	1
Канада	9	37	43	43	52	48
Австралия	10	47	52	47	53	28
Франция	11	14	18	16	15	9
Румыния	12	3	2	6	1	2
Польша	13	5	3	7	2	3
Нидерланды	17	20	20	10	21	40
Бельгия	18	25	34	9	49	44
Болгария	21	15	7	15	3	13
Венгрия	22	18	10	22	9	18
Индия	27	7	16	19	10	4
Вьетнам	32	16	21	14	34	6
Россия	35	23	19	24	17	33
Филиппины	37	10	29	23	18	8
Сирия	189	17	9	18	4	7

4.1.7 Анализ межстрановой торговли продовольствием

Данное направление посвящено анализу отношений между странами в терминах экспорта/импорта продовольствия. Для более глубокого понимания продовольственной ситуации в каждой стране и определения лидеров на международной арене, во-первых, был проанализирован продовольственный баланс стран. Продовольственный баланс включает в себя такие показатели, как общее производство, нормы потребления, потери, остатки и прочие показатели. При этом оценка данных показателей приводится для различных категорий продуктов. В Таблице 4.2 приведены лидеры по каждому из основных показателей за 2010 год для трёх выбранных категорий продуктов (зерновые, мясо и овощи) [180].

Так, объёмы внутренних запасов включают в себя такие показатели как производство, общий экспорт, импорт и изменение уровня запасов. Общее продовольствие показывает уровень запасов того или иного продукта, доступного для населения в рассматриваемый период. Показатель пищевых ресурсов измеряется как общий запас в килограммах в расчёте на человека в год.

Таблица 4.2 – Страны с высокими показателями продовольственного баланса

Внутренние запасы			Продовольствие			Потери		
Зерн.	Мясо	Овощи	Зерн.	Мясо	Овощи	Зерн.	Мясо*	Овощи
Китай	Китай	Китай	Китай	Китай	Китай	Китай	Перу	Китай
США	США	Индия	Индия	США	Индия	Индия	Канада	Индия
Индия	Бразилия	США	Индонезия	Бразилия	США	Бразилия	Аргентина	Турция
Бразилия	Россия	Турция	США	Россия	Египет	Индонезия	Япония	США
Индонезия	Германия	Египет	Бангладеш	Германия	Иран	Мексика	Мьянма	Иран

Производство			Пищевые ресурсы		
Зерн.	Мясо	Овощи	Зерн.	Мясо	Овощи
Китай	Китай	Китай	Марокко	НК	Китай
США	США	Индия	Египет	США	Армения
Индия	Бразилия	США	Лесото	Австралия	Тунис
Бразилия	Германия	Турция	Азербайджан	Багамы	Черногория
Франция	Россия	Иран	Мали	Макао	Албания

*Ограниченные данные

Данные показатели учитывают статистические характеристики каждой из страны, но не учитывают данные по торговым взаимоотношениям между ними. Поэтому также необходимо изучить, как страны взаимодействуют друг с другом на основе информации об экспортно-импортных взаимосвязях.

Для построения сети торговли продовольствием была использована база данных World Integrated Trade Solution Database (WITS Comtrade) [181]. Данная база предоставляет данные по двусторонней торговой статистике между странами по отдельным категориям товаров (в частности, по продовольствию).

Большинство потоков между странами описывается двумя показателями (зеркальные данные): значение экспорта по мнению экспортёра и значение импорта по мнению импортёра (в тыс. долларов). В большинстве случаев эти числа расходятся в несколько десятков раз (см. рисунок 4.10).

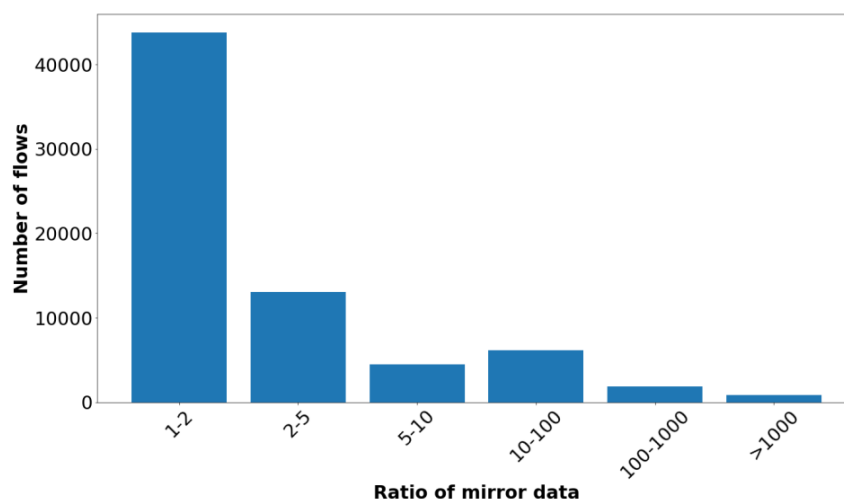


Рисунок 4.10 – Распределение расхождения зеркальных данных

Для построения графа, отражающего торговые отношения между странами, по каждому потоку было выбрано единое число. Выбор данного числа основывается на опросе ООН [182], позволяющего оценить качество сбора информации и процедуру отчётности о торговых отношениях со странами-партнёрами.

Для построения графа экспорта/импорта продовольствия были выбраны 10 основных категорий продуктов, соответствующих классификации SITC Revision 3. Эти категории включают в себя живых животных (Ж.Ж.), мясо, молочные продукты и яйца, рыбу, зерновые, фрукты и овощи (Фр./Ов.), сахар и мёд, кофе, чай, какао и специи, корм для животных и прочие съедобные продукты. Мы также построили граф общей торговли по всем категориям продуктов для сравнения результатов с отдельными видами продуктовых товаров.

Для применения индексов дальних взаимодействий LRIC [169] стоит необходимость выбора порога для каждой вершины, начиная с которого вершина считается подверженной влиянию. Иными словами, необходимо выбрать такой объем товара, потеря которого будет критична для данной страны. Для расчёта этого порога мы ориентировались на общие объёмы экспорта и импорта рассматриваемого вида продовольствия. Если страна *A* является крупным экспортёром и при этом мало импортирует, то никакая другая стра-

на не сможет повлиять на A через сокращение её импорта. В противном случае страна A может сократить свой экспорт рассматриваемого товара и восполнить потери импорта. С другой стороны, если страна B является крупным импортёром и мало экспортирует, то потеря части импорта будет критичной для B , т.к. эта страна не сможет покрыть потери, просто сократив экспорт в другие страны. Поэтому выбор порога должен зависеть и от экспорта, и от импорта. В частности, в нашей работе мы выбрали $q_i = p\%$ от $\max(\text{общий экспорт страны } i; \text{общий импорт страны } i)$, где p установили на уровне 15%. При этом другие проценты потерь были также последовательно проанализированы, и корреляционный анализ показал незначительные отклонения в результатах в зависимости от разного выбора порога. В Таблице 4.2 приведен индекс корреляции Кендалла итогового ранжирования по LRIC в зависимости от выбранного порога для общей торговли продуктами.

Таблица 4.1 – Корреляция Кендалла финального ранжирования индексов LRIC в зависимости от выбранного порога влияния

%	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	1	.89	.85	.84	.83	.82	.82	.81	.81	.81	.80	.79	.79	.79	.78	.78	.78	.77	.77	.76	.77	.77	.76	.76	.75
2	.89	1	.95	.93	.92	.91	.90	.90	.89	.89	.87	.86	.86	.86	.86	.85	.84	.84	.83	.83	.83	.83	.82	.82	.81
3	.85	.95	1	.97	.96	.94	.94	.93	.92	.91	.89	.89	.88	.88	.88	.87	.86	.86	.85	.85	.85	.84	.84	.83	.82
4	.84	.93	.97	1	.98	.96	.95	.94	.93	.93	.91	.90	.89	.89	.89	.88	.87	.87	.86	.86	.86	.85	.85	.84	.83
5	.83	.92	.96	.98	1	.98	.97	.96	.95	.94	.92	.91	.91	.90	.90	.89	.89	.88	.87	.87	.87	.86	.86	.85	.84
6	.82	.91	.94	.96	.98	1	.98	.97	.96	.95	.93	.93	.92	.92	.91	.90	.90	.89	.89	.88	.88	.88	.87	.86	.85
7	.82	.90	.94	.95	.97	.98	1	.99	.97	.97	.94	.94	.93	.93	.92	.92	.91	.90	.90	.89	.89	.89	.88	.87	.86
8	.81	.90	.93	.94	.96	.97	.99	1	.98	.98	.95	.95	.94	.94	.93	.93	.92	.91	.91	.90	.90	.89	.89	.88	.87
9	.81	.89	.92	.93	.95	.96	.97	.98	1	.98	.97	.96	.95	.95	.95	.94	.93	.92	.91	.91	.91	.90	.89	.89	.88
10	.81	.89	.91	.93	.94	.95	.97	.98	.98	1	.97	.96	.96	.95	.95	.94	.93	.93	.92	.91	.92	.91	.90	.90	.89
11	.80	.87	.89	.91	.92	.93	.94	.95	.97	.97	1	.99	.98	.98	.97	.97	.95	.94	.94	.93	.94	.93	.92	.91	.90
12	.79	.86	.89	.90	.91	.93	.94	.95	.96	.96	.99	1	.99	.99	.98	.97	.96	.95	.95	.94	.94	.93	.92	.92	.91
13	.79	.86	.88	.89	.91	.92	.93	.94	.95	.96	.98	.99	1	.99	.99	.98	.96	.95	.95	.94	.94	.93	.92	.92	.91
14	.79	.86	.88	.89	.90	.92	.93	.94	.95	.95	.98	.99	.99	1	.99	.98	.97	.96	.95	.95	.95	.94	.93	.93	.92
15	.78	.86	.88	.89	.90	.91	.92	.93	.95	.95	.97	.98	.99	.99	1	.99	.97	.96	.96	.95	.95	.94	.93	.93	.92
16	.78	.85	.87	.88	.89	.90	.92	.93	.94	.94	.97	.97	.98	.98	.99	1	.98	.97	.96	.96	.96	.95	.94	.94	.93
17	.78	.84	.86	.87	.89	.90	.91	.92	.93	.93	.95	.96	.96	.97	.97	.98	1	.99	.98	.97	.97	.96	.95	.95	.94
18	.77	.84	.86	.87	.88	.89	.90	.91	.92	.93	.94	.95	.95	.96	.96	.97	.99	1	.99	.98	.98	.97	.96	.96	.94
19	.77	.83	.85	.86	.87	.89	.90	.91	.91	.92	.94	.95	.95	.95	.96	.96	.98	.99	1	.99	.99	.98	.97	.96	.95
20	.76	.83	.85	.86	.87	.88	.89	.90	.91	.91	.93	.94	.94	.95	.95	.96	.97	.98	.99	1	.99	.98	.97	.97	.96
21	.77	.83	.85	.86	.87	.88	.89	.90	.91	.92	.94	.94	.94	.95	.95	.96	.97	.98	.99	.99	1	.99	.97	.97	.95
22	.77	.83	.84	.85	.86	.88	.89	.89	.90	.91	.93	.93	.93	.94	.94	.95	.96	.97	.98	.98	.99	1	.98	.97	.96
23	.76	.82	.84	.85	.86	.87	.88	.89	.89	.90	.92	.92	.92	.93	.93	.94	.95	.96	.97	.97	.97	.98	1	.99	.98
24	.76	.82	.83	.84	.85	.86	.87	.88	.89	.90	.91	.92	.92	.93	.93	.94	.95	.96	.96	.97	.97	.97	.99	1	.98
25	.75	.81	.82	.83	.84	.85	.86	.87	.88	.89	.90	.91	.91	.92	.92	.93	.94	.94	.95	.96	.95	.96	.98	.98	1

В результате были рассчитаны индексы дальних взаимодействий для каждой из 10 категорий продуктов. В Таблице 4.3 приведены результаты расчёта для одного из показателей LRIC за 2010 год. Остальные индексы дальних взаимодействий, включая индивидуальный PageRank, были также проанализированы и показали высокую корреляцию между собой (на уровне 0.9 – 0.99).

Таблица 4.2 – Ранжирование стран (TOP-10) согласно LRIC за 2010 год

Ранжирование	Категории продуктов									
	Ж.Ж.	Мясо	Молоч.	Рыба	Зерн.	Фр./Ов.	Сахар	Кофе	Корм	Другое
	FRA	BRA	DEU	NOR	CAN	ESP	BRA	BRA	ARG	IRL
	DNK	DEU	NLD	CHN	USA	MEX	FRA	CIV	BRA	NLD
	DEU	NLD	FRA	THA	FRA	CHN	DEU	NLD	USA	DEU
	NLD	BEL	BEL	RUS	DEU	NLD	BEL	BEL	NLD	CAN
	BEL	DNK	NZL	CAN	ITA	BEL	NLD	DEU	DEU	USA
	CAN	CAN	ITA	VNM	BEL	USA	MEX	GHA	PER	BEL
	MEX	FRA	DNK	USA	AUS	CAN	CAN	CMR	BEL	FRA
	USA	USA	AUT	IDN	ARG	CHL	THA	FRA	FRA	ITA
	AUS	ESP	USA	SWE	HUN	ITA	CUB	IDN	PRY	SGP
	HUN	AUS	POL	CHL	NLD	BRA	ESP	CHE	IND	CHE

Как видно из Таблицы 4.3, наборы наиболее влиятельных стран по каждой из категорий продуктов сильно различаются между собой. Действительно, корреляционный анализ не выявил никакой зависимости между ранжированием относительно различных продуктовых категорий, а также относительно общей торговли.

4.1.8 Анализ террористических сообщений

4.1.8.1 Описание данных

Целью нашей работы стало изучение влияния в международной террористической сети, охватывающей период активности разных групп с 2001 по 2016 годы. В работе были использованы данные из базы Global Terrorism Database, представленной исследователями из Университета Мэриленда [183]. База содержит информацию о 170 350 террористических атак с 1970 по 2016, включая дату, место, использованное оружие, цель атаки, число жертв

и раненных, а также информацию об ответственных за атаку группах. В период наблюдения (2001-2016) вошли 83 635 случаев террористических атак в 158 странах мира.

Для того чтобы получить информацию о связях между террористическими организациями, нами были выбраны случаи терактов, за организацию которых ответственна более, чем одна группа. В период с 2001 по 2016 годы имели место 1216 подобных случаев.

Таким образом, для анализа были выбраны 426 групп, совершившие теракты в 90 странах; все группы взаимодействовали хотя бы с одной другой группой при совершении атаки, на основании чего был построен граф. Были выделены компоненты связности графа (подграфы), каждый из которых включает различные по идеологической направленности и географическому положению террористические группы. Всего в сети было выделено 60 компонент связности. Мы сосредоточили наше внимание на пяти самых крупных подграфах. Первый включает в себя 253 группы религиозной (исламистской), сепаратистской, этнической и левой направленности и распространяется почти на все регионы мира, включая, в первую очередь, Ближний Восток, Северную и Центральную Африки, Центральную, Южную и Юго-Восточную Азии. Число атак, совершенных группами первой компоненты достигает 24 284, а число жертв данных атак – 104 392. Второй подграф относится к Центральной Африке и включает 20 групп религиозной и этнической направленности. Данные группы несут ответственность за организацию 2078 терактов с 6151 жертвами. Третий подграф состоит из 13 анархистских и экологических террористических групп из Европы. Группы организовали 245 атак, и от их действий пострадал только 1 человек. Четвёртый подграф включает 7 сепаратистских террористических организаций Непала, совершившие 41 теракт, в результате которых погибло 46 человек. Пятый подграф из 5 левых террористических групп полностью сконцентрирован в Колумбии. Данные организации совершили 1374 террористические атаки и несут ответственность за гибель 1449 человек.

4.1.8.2 Основные результаты

Для групп из пяти самых крупных подграфов были посчитаны индексы ближних (SRIC) [186] и дальних взаимодействий (LRIC) [170] с целью выделить наиболее влиятельные группы в каждом подграфе, учитывая при этом как прямое влияние групп друг на друга, так и не прямое (через посредников). При подсчете индексов были использованы индивидуальные характеристики групп – число совместных атак и количество погибших в результате атак людей.

По результатам подсчёта индекса ближних взаимодействий для трёх разных периодов между 2001 и 2016 (2001-2007, 2008-2011, 2012-2016) были выявлены наиболее влиятельные группы в каждом рассматриваемом подграфе. Так, на основании количества совместных атак в период с 2001 по 2007 в первом подграфе наиболее влиятельными являлись Al-Aqsa Martyrs Brigade, Hamas, Palestinian Islamic Jihad и Al-Qaida; в 2008-2011 – Lashkar-e-Taiba, Al-Aqsa Martyrs Brigade, Indian Mujahideen и Students Islamic Movement of India; в 2012-2016 – Al-Nusrah Front и Islamic State of Iraq and the Levant. С другой стороны, на основании количества жертв совместных атак группы с наивысшими показателями влияния: в 2001-2007 – Al-Qaida, Hamas, Jemaah Islamiya; в 2008-2011 – Lashkar-e-Taiba, Deccan Mujahideen, Tehrik-i-Taliban Pakistan, Lashkar-e-Jhangvi; в 2012-2016 – Al-Nusrah Front и Tehrik-i-Taliban Pakistan. Во втором подграфе в 2001-2007 годы на основании и числа совместных атак, и числа жертв наиболее влиятельными группами являлись Party for the Liberation of the Hutu People и National Council for Defense of Democracy; в 2008-2011 – Al-Shabaab, Hizbul al Islam и Ogaden National Liberation Front; в 2012-2016 – Democratic Front for the Liberation of Rwanda, Allied Democratic Forces и Mayi Mayi.

С помощью индекса дальних взаимодействий мы смогли выделить группы в первом и втором подграфе, которые могут оказывать скрытое влияние на сеть через другие группы. Так, индекс выделил группы, оказывающие

влияние на сеть через трёх посредников. В первой компоненте в период с 2001 по 2007 годы на основании числа совместных атак лидировали Great Eastern Islamic Raiders Front, Kurdistan Freedom Hawks, Al-Qaida; в 2008-2011 – Harkatul Jihad-e-Islami, 313 Brigade, United Liberation Front of Assam; в 2012-2016 – Jaysh al-Islam, Ansar al-Sharia Operations Room, Islamic State of Iraq and the Levant. На основании числа жертв совместных атак в данной компоненте выделились другие группы: Lashkar-e-Jhangvi, Harakat ul-Mujahidin, Harkatul Jihad-e-Islami (2001-2007); 313 Brigade, Harkatul Jihad-e-Islami, Students Islamic Movement of India (2008-2011); Asa'ib Ahl al-Haqq, Syrian Resistance, Military Council of the Tribal Revolutionaries (2012-2016).

4.1.9 Индексы, учитывающие предпочтения индивидов по созданию коалиций

4.1.9.1 Актуальность задачи и методы решения

В этом подразделе отчета приведены результаты исследования аксиоматики индексов влияния в сетях и разработки такой аксиоматики для индексов, учитывающих предпочтения агентов по созданию коалиций.

Индексы центральности, учитывающие предпочтения агентов по образованию коалиций, описанные в [184], использовались в нескольких эмпирических исследованиях [185, 186]. В них такие индексы рассчитаны вместе с классическими определителями влияния, приведены сходства и различия полученных разными способами распределений переговорной силы в реальных управляющих органах.

Такие индексы уже используются в исследованиях распределения влияния в принимающих решения структурах, что обосновывает необходимость создания теоретических работ по исследованию их свойств. Целью данного направления работы является определение аксиоматики для индексов, учитывающих силу связи между участниками, и обобщение классических индексов влияния.

4.1.9.2 Методология и основные результаты исследования

Основные определения.

Модель выборного органа представляет собой сеть, в которой набор вершин (V) представляет агентов или партии (N), набор направленных взвешенных ребер (E) представляет отношения между партиями (P_{ij}). Вес ребра P_{ij} отражает желание партии i вступить в коалицию с партией j или вероятность образования связи для этих партий.

Решение в таком выборном органе принимается голосованием и считается принятым, если за него проголосовали q (квота) человек или больше. Игроки могут объединяться в коалиции. Коалиция считается выигрывающей, если имеет q или больше голосов. Агент считается ключевым в выигрывающей коалиции, если без его голосов она перестает быть выигрывающей. Обозначим через W_i множество коалиций, в которых игрок i является ключевым.

Свойства классических индексов.

Все классические индексы обладают набором свойств, полностью определяющих эти индексы. Для основных индексов существует несколько таких определяющих наборов, описывающих разные свойства и подходящих для различных задач. Целью данного направления исследований является определение набора свойств, определяющих индекс влияния, учитывающий предпочтения агентов по образованию связей. Исследование различных методов определения распределения влияния в сети включает в себя описание и обобщение, а так же сравнение между собой их свойств.

К классическим мерам влияния относятся индексы Банцафа, Пенроуза, Шепли-Шубика и основанные на них. Их свойства описаны в [187, 188, 189, 190].

Индекс Банцафа рассчитывается в зависимости от числа коалиций, в которых игрок является ключевым: $Bz_i = \frac{|W_i|}{\sum_{j \in N} |W_j|}$. Индекс Шепли-Шубика рас-

считывается как число таких коалиций, взвешенное на размер коалиции и количество игроков: $SS_i = \sum_{S \in W_i(v)} \frac{(|N|-|S|)!(|S|-1)!}{|N|!}$. Оба индекса удовлетворяют аксиомам о болване (индекс равен нулю для игрока, не являющегося ключевым ни в одной коалиции), анонимности (все игроки рассматриваются одинаково), трансфера (об изменениях при объединении коалиций). Эффективность определена для этих индексов по-разному, но близка по смыслу, с учетом того, что индекс Шепли-Шубика с большим весом учитывает большие коалиции, а индекс Банцафа считает все выигрывающие коалиции одинаково.

На основе классического индекса Шепли-Шубика в [191] разработана мера Шепли для ориентированных графов. Это мера для распределения выигрыша в кооперативной игре, в которой с помощью ориентированного графа задается иерархия между игроками. Полная аксиоматика для этой меры, а также ее связь с вектором Шепли, лежащим в основе индекса Шепли-Шубика, приведена в [191].

Мера Майерсона определена в [192], полное описание свойств показано в [193]. Эта мера основана на векторе Шепли и отражает распределение выигрыша в играх с ограничением на кооперацию. В сети, отражающей связи между игроками, наличие ребра показывает возможность создания коалиции, отсутствие ребра отражает невозможность. Следовательно, эта мера близка к индексам влияния, учитывающим предпочтения по кооперации, но рассматривает только детерминистический случай, в котором агенты либо могут вступать в коалицию, либо нет. В реальности же чаще ограничения вызваны не правилами принятия решений, а взаимоотношениями между партиями. И тогда лучшей моделью будет создание вероятностной меры, отражающей большую или меньшую вероятность для некоторой партии i образовать связь с партией j , чем с некоторой третьей. Основные свойства меры Майерсона по [193] – эффективность (вся создаваемая коалицией стоимость распределяется между входящими в нее игроками) и справедливость (два игрока получают одинаковый излишек от образованной между ними связи). Эти свойства на-

прямую следуют из определения меры Майерсона как решения класса игр на сетях.

Обратимся к мерам центральности для сетей с целью выделения наиболее влиятельных узлов. Наиболее известные меры центральности – степенная (degree, число ребер вершины), близости (closeness, среднее расстояние до других узлов), по посредничеству (betweenness, доля кратчайших путей между другими вершинами, проходящих через данную). Меры, учитывающие характеристики других вершин (связи с более важными вершинами повышают важность вершины): по собственному вектору (eigenvector), центральность Боначича (Bonacich), престиж Катца (Katz prestige). Меры центральности позволяют определить более влиятельные вершины, следовательно, эти методы могут использоваться для определения распределения переговорной силы в парламенте.

Еще одна мера влияния вершин в сети это функции согласия (consensus functions). Они появились как решение проблемы общественного выбора голосованием, но могут быть использованы как мера влияния в сети или ее части. Функция может быть задана в зависимости от задачи, разные функции определяют наиболее влиятельными разные вершины. Аргумент такой функции — это профиль из вершин графа, причем они могут повторяться, что отражает наличие нескольких голосов. Значением функции будет набор вершин, причем не обязательно входящих в профиль. Центральная функция (center function) возвращает вершины, минимизирующие максимальное расстояние до вершин профиля. Свойства такой функции исследованы в [194, 195]. Медианная функция (median function) возвращает вершины, минимизирующие среднее расстояние до вершин профиля. Свойства медианной функции и некоторые доказательства показаны в [196]. Существуют и другие функции согласия.

Разработанные в [187, 197] и описанные в предыдущих отчетах индексы ближних и дальних взаимодействий позволяют определить влиятельность каждой вершины в сети. Основное преимущество этих индексов в том, что

как и меры центральности, основанные на собственных векторах или престиже, они учитывают влияние соседних узлов. При этом они значительно проще вычисляются (в смысле количества операций), и поэтому могут быть рассчитаны для больших сетей. Эти индексы использовались ранее в нескольких исследованиях лаборатории [1].

Индексы влияния, учитывающие предпочтения агентов по созданию коалиций, приведенные в [184], описываются функцией, схожей с индексом Банцафа:

$$\alpha_i = \frac{\chi_i}{\sum_j \chi_j},$$

где $\chi_i = \sum_{w_i} f(i, w)$ – сумма функций интенсивности связи для каждого из участников по тем коалициям, в которых он является ключевым игроком. Функция интенсивности может определяться различным образом в зависимости от задачи. Она зависит от взвешенных ребер внутри коалиции и отражает силу связи между агентами. Например, если это сумма весов всех ребер, входящих в вершину i , исходящих из вершин остальных членов коалиции, то эта функция отражает, насколько все агенты коалиции хотели бы объединиться с агентом i . Другие возможные определения для функции интенсивности связи приведены в [184].

В [190] показаны свойства описанных индексов влияния в зависимости от определения функций интенсивности. Эти свойства сформулированы на основе описанных в [188] для классических индексов [187, 197].

Заметим, что некоторые индексы влияния могут быть обобщены до исследуемых индексов. Например, если граф неориентированный и невзвешенный, и функция связи описывается количеством коалиций, в которых игрок является ключевым, то, в соответствии с приведенной формулой, индекс α_i будет в точности равен индексу Банцафа. Другие подобные соответствия приведены в [190]. Следовательно, при определенных условиях, индексы

влияния, учитывающие предпочтения по созданию коалиций, обладают свойствами, разработанными для классических индексов влияния.

4.1.9.3 Дальнейшее развитие работы

На текущий момент исследованы существующие подходы к определению влияния в сетевых структурах и их свойства. Наборы свойств, выполняющихся для некоторого индекса и только для него (определяющие его), являются аксиоматикой, описывающей меру влияния. Рассмотрены подходы к определению степени влияния со стороны теории графов, теории сетей и теории игр.

Дальнейшим развитием работы в этом направлении является разработка аксиоматики для индексов влияния, учитывающих предпочтения по образованию коалиций, описание их свойств и взаимосвязей с другими мерами влияния в сетях. На основе существующих подходов, описанных в этой части отчета, разрабатывается аксиоматика для новых индексов.

4.1.10 Некоторые выводы

На основе разработанных ранее моделей влияния в сетях был предложен эффективный алгоритм вычисления индексов ближних и дальних взаимодействий. Эффективность алгоритма достигается за счет более низкого времени выполнения вычисления прямого и непрямого влияния между вершинами, не требующего перебора всех возможных групп вершин и рассмотрения всех возможных путей между вершинами.

Вычислительная сложность алгоритма оценена на основе вычислительных экспериментов, в рамках которых было сгенерировано множество полных графов и графов с экспоненциальным распределением по числу ребер. Было показано, что в обоих случаях время вычисления предложенного алгоритма значительно ниже, что означает, что данный алгоритм может быть

применен на графах с сотнями и тысячами вершин. Предложенный алгоритм реализован в качестве библиотеки на языке R.

С помощью сетевых моделей были также проанализированы отношения между странами на различных уровнях взаимодействия. В работе проведен анализ международной миграции с 1970 по 2014гг. на основе данных по потокам миграции между странами. Сделан анализ динамики миграционных потоков по регионам, на основе которого выделены периоды. С помощью сетевого анализа определены рейтинги влиятельных стран в процессе международной миграции с помощью индексов центральностей, как классических, так и учитывающих не прямые взаимодействия. Данные результаты могут быть полезны при планировании миграционной политики принимающими странами, прогнозировании будущих миграционных потоков между странами.

В этом подразделе также представлены результаты сетевого анализа торговли продовольствием. Были проанализированы известные меры продуктового баланса, а также построена сеть экспорта/импорта продовольствия по различным категориям продуктов и общей торговле. Был проведен анализ выбора порога, когда вершина считается подверженной влиянию. Также были проанализированы итоговые ранжирования по различным категориям продуктов, и было показано, что зависимости между влиятельными странами по различным продуктам, а также общей торговле не наблюдается.

Индексы дальних взаимодействий были также применены к анализу террористических сообществ. С помощью новых индексов центральности можно выявить группы, влияние которых на первый взгляд не очевидно и имена которых неизвестны широкой публике. Также, используя сетевой подход к анализу террористической активности, мы можем наблюдать распространённость террористической сети в разных регионах мира и наблюдать динамику влияния террористических групп в различные периоды.

Наконец в последнем подразделе отчета приведены некоторые предварительные результаты по исследованию индексов влияния в сетях с учетом предпочтений индивидов по созданию коалиций.

4.2 Анализ сообщества экономистов и построение рейтингов экономических журналов

4.2.1 Введение и постановка задачи

Задача построения рейтинга является актуальной задачей последних лет в связи с возрастающей популярностью построения всевозможных рейтингов в научном сообществе – рейтингов университетов [198], журналов [190, 200, 201, 202], научных работников [203, 204] – для оценки эффективности их деятельности. Однако единой общепризнанной методологии их построения до сих пор нет. В данной работе мы сосредоточимся на построении рейтингов научных журналов по экономике и смежным дисциплинам, однако подобная методология применима и к другим задачам рейтингования с соответствующими поправками на объект рейтингования.

Ранжирование научных журналов является многокритериальной задачей, как правило, исследователи учитывают библиометрические показатели – цитируемость, импакт-фактор, средний объем публикаций и др., а также экспертные оценки научного уровня и престижности журналов. Чаще всего для агрегирования информации используется взвешенная сумма критериев [201, 204, 205], хотя известно, что этот метод обладает рядом серьёзных неустраняемых недостатков [206] и должен использоваться с большой осторожностью.

Есть рейтинги, основанные на лексикографическом упорядочении нескольких критериев. Например, в 2007-2008 в ВИНТИ РАН проводилось экспертное оценивание российских и зарубежных научных журналов [199] для чего было предложено лексикографическое правило на трех критериях «Максимальная оценка», «Обобщенная оценка» и «Нормированный ранг».

Эта статья вызвала в свое время бурную дискуссию, и ей был посвящен отдельный сборник (Управление большими системами, Выпуск 27) с детальным разбором предложенной методики. Критики отметили, что она имеет целый ряд недостатков:

- 1) неустойчивость к ошибкам экспертов [207] и подверженность манипулированию со стороны экспертов [208];
- 2) использование суммарных, а не усредненных величин неполных экспертных оценок приводит к смещению оценок – журналы, получившие много низких оценок и одну высокую, могут занять в рейтинге более высокое место по сравнению с журналами, имеющими только высокие оценки, но в малом количестве [207, 209].

Более убедительным представляется подход, основанный на теории коллективного выбора. Задачей теории коллективного выбора является агрегирование индивидуальных мнений участников голосования в единое коллективное мнение. Известно, что не существует «идеального» правила голосования (теорема Эрроу о невозможности), однако предложены ряд правил принятия коллективных решений, которые обладают «хорошими» свойствами – более подробно о правилах коллективного выбора и их соответствии нормативным свойствам см., например, в [210]. Подобный подход уже использовался для ранжирования экономических журналов [202], журналов по менеджменту [200] и вклада научных работников [203].

Однако в указанных работах для ранжирования журналов использовались только библиометрические показатели, хотя известно, что этими показателями легко можно манипулировать и «накручивать» их для повышения рейтинга [201, 203, 205]. Более адекватным представляется использование экспертной информации для оценки и рейтингования журналов, однако и здесь встает вопрос о подборе экспертов – является ли выборка экспертов представительной и насколько выбранных экспертов можно считать экспертами в данной области. Существующие работы по рейтингованию научных журналов на основе экспертного мнения либо полагаются на метод рекомен-

даций и метод «снежного кома» при формировании пула экспертов (рейтинг Управления академической экспертизы НИУ ВШЭ, 2015), что не гарантирует полноту и репрезентативность выборки, т.к. метод сильно зависит от первоначальной выборки экспертов, либо обходят стороной вопрос о выборе экспертов, например, в работе [199] опрашивались 237 редакторов реферативных журналов ВИНТИ РАН, что обусловлено спецификой поставленной задачи, но не гарантирует компетентности экспертов.

Еще одна проблема с определением экспертного сообщества связана с тем, что сообщество российских экономистов является неоднородным [198, 211] и журналы, значимые для одних экономистов, могут быть малозначимыми для других. Кроме того, в работе [212] в ходе анализа анкет экономистов указано на «расслоение» российского академического сообщества экономистов (РАСЭ): «Наше исследование подтверждает тезис о "немейнстримности" отечественного сообщества экономистов: 39,6% респондентов однозначно не идентифицируют себя с "основным течением" экономической мысли, а доля сторонников мейнстрима оказалась в 1,5 раза ниже находящихся за пределами "магистрального течения" исследователей и составила 25,7%».

В нашей работе мы постараемся описать неоднородность российского сообщества экономистов, выделить подмножество экономистов, которых можно считать «продвинутыми» исследователями, обладающих современными знаниями в области экономики и смежных направлений, которых можно считать экспертами, и учесть эту неоднородность сообщества экономистов при построении рейтинга экономических журналов.

Данная работа является частью совместного проекта Института экономики Российской академии наук, Национального исследовательского университета Высшая школа экономики и Новой экономической ассоциации «Стратификация научного сообщества экономистов и ранжирование экономических журналов», разрабатываемого временным коллективом сотрудников в составе: В.С. Автономов (НИУ ВШЭ), Ф.Т. Алескеров (НИУ ВШЭ),

Н.А. Бураков (ИЭ РАН), Л.Г. Егорова (НИУ ВШЭ), А.Л. Мячин (НИУ ВШЭ), О.А. Славинская (ИЭ РАН), А.Я. Рубинштейн (ИЭ РАН, руководитель исследования). Результаты этой работы обсуждались на XIX Апрельской международной научной конференции по проблемам развития экономики и общества.

4.2.2 Данные и методология

4.2.2.1 Описание анкеты

В основе нашего исследования лежат результаты анкетирования российского сообщества экономистов, проводившегося в декабре 2016-апреле 2017гг. На анкету отвечали участники Третьего Российского Экономического Конгресса (РЭК-2016), XVIII Апрельской международной научной конференции по проблемам развития экономики и общества, а также участники Московского экономического форума 2017 (МЭФ-2017). Всего получено 1028 ответов, из них 645 – от участников РЭК-2016, 130 – от участников Апрельской конференции и 254 – от участников МЭФ-2017.

Анкета реализована на основе сервиса «Google Forms». Персональный доступ респондентов обеспечен направлением на личную почту каждого соответствующей Интернет-ссылки. Данная анкета состоит из 30 вопросов, которые можно условно объединить в три блока. Первый блок служит для выявления структуры сообщества экономистов и включает в себя такие вопросы как наиболее важный вид деятельности респондента, тип организации, в которой он работает, распределение его рабочего времени, его оценку структуры российского экономического сообщества как соотношения между преподавателями, экспертами и академическими исследователями, и др. Второй блок вопросов служит для оценки респондентом предлагаемой выборки журналов – их научного уровня, общественного престижа, интереса к публикациям этих журналов, а также для выявления мнения респондента относительно различных библиометрических показателей. Третий блок включает

в себя вопросы для определения социально-демографических характеристик респондентов – пола, возраста, наличия ученой степени и др.

Отдельно необходимо упомянуть используемый нами список журналов, который оценивали респонденты. Данный список основан на списке RSCI (раздел «Экономика. Экономические науки», 29 журналов), из которого были исключены 16 изданий и добавлены 13 журналов, относящихся к разделам «Междисциплинарные журналы в области общественных и гуманитарных наук», «Организация и управление». Изменение списка связано с желанием авторов проекта учесть известные и значимые журналы смежных с экономикой наук, представляющих интерес для экономистов и в которых многие также публикуются или читают. Понятно, что такой экспертный подход к составлению исходной выборки журналов для построения рейтинга может показаться неполным и быть подвергнут сомнению в правильности отбора журналов, но мы в этом вопросе солидарны с коллегами и можем привести выдержку из [213]: «В рамках обсуждения на круглом столе «Экономическое сообщество и ранжирование российских экономических журналов» Апрельской конференции (13 апреля 2017 года) было высказано критическое замечание в отношении формирования фиксированного списка анализируемых журналов. Так, по мнению А. Муравьева, более корректным является подход НИУ ВШЭ (Проект НИУ ВШЭ, 2015), согласно которому респондентам предоставляется возможность расширять список анализируемых изданий, добавляя в него журналы, не вошедшие в анкету. Полезными также являются рекомендации Н. Тихоновой и В. Полтеровича о целесообразности разделении совокупности изданий на научные журналы экспертного уровня и журналы, обладающие своей особой аудиторией читателей. Принимая во внимание эти предложения надо иметь в виду все же, что расширение списка журналов в анкете порождает известные проблемы, связанные с качеством ответов респондентов. И, как свидетельствуют социологические исследования, перегруженность вопросов, содержащих слишком много позиций, которые должен оценить респондент, приводит часто к искаженным ответам и/или к отказам

давать необходимые оценки. Очевидно, что исходный список экономических журналов в данном проекте, как и любой другой их перечень, не может быть полным».

Также относительно небольшой размер выборки журналов в нашей анкете продиктован тем, что при увеличении количества вопросов в анкете и объектов для анализа респондентами падает вероятность заполнения анкеты и точность оценок. Впрочем, респонденты могли самостоятельно указать, публикации в каких журналах, помимо предложенных, они считают интересными, и в каких журналах, помимо предложенных, они публикуются – на каждый из этих вопросов мы получили более сотни ответов. По объективным причинам, включить подобный список в анкету было бы невозможно.

4.2.2.2 Анализ латентных классов

Метод латентных классов (Latent Class Analysis, LCA) – это статистический метод для нахождения классов (кластеров) в многомерных категориальных данных [214]. Предполагается, что наблюдаемые объекты имеют скрытые (латентные) характеристики, которые позволяют распределить их в однородные группы. Метод латентных классов приписывает каждому объекту вероятности нахождения в каждом из классов. Естественно, далее можно отнести объект к тому классу, вероятность нахождения в котором является наибольшей, и получить однозначную классификацию всех исследуемых объектов.

Приведем общее описание метода. Пусть L – число наблюдаемых переменных (число вопросов в анкете, которые используются для кластеризации), $Y_l, l = 1, \dots, L$ – вопрос анкеты под номером l , y_l – ответы респондента на вопрос Y_l . Обозначим через C число латентных классов, а через X – сами латентные классы $x \in \{1, 2, \dots, C\}$. Также для удобства будем использовать векторные обозначения Y и y для описания совокупности ответов респондента на вопросы анкеты.

Вероятность $P(Y = y)$ получения конкретного ответа y на анкету по формуле полной вероятности есть

$$P(Y = y) = \sum_{x=1}^C P(X = x)P(Y = y | X = x),$$

где $P(X = x)$ – это доля людей, относящихся к латентному классу x , $P(Y = y | X = x)$ – вероятность получения совокупности ответа y от людей из этого класса.

Модель латентных классов предполагает независимость наблюдаемых переменных, поэтому вероятность получения паттерна ответа y от людей из класса x равна

$$P(Y = y | X = x) = \prod_{l=1}^L P(Y_l = y_l | X = x).$$

Таким образом, вероятность получения ответов будет равна

$$P(Y = y) = \sum_{x=1}^C P(X = x) \prod_{l=1}^L P(Y_l = y_l | X = x).$$

Далее методом максимального правдоподобия оцениваются параметры модели. После оценки параметров модели по формуле Байеса можно рассчитать вероятности нахождения респондента в определенном классе в зависимости от полученных ответов на вопросы анкеты, что позволяет описать полученные классы объектов.

Для расчетов параметров модели мы использовали пакет `poLCA`, созданный для языка программирования R [215].

4.2.2.3 Метод анализа паттернов

Для исследования обобщённого представления экономистов обо всей структуре экономического сообщества в работе предложено использование метода анализа паттернов.

Приведем краткое описание метода. В качестве характеристик множества респондентов использовались ответы на вопрос об удельном весе каждого из трех групп экономистов: преподаватели в университетах (ПУ), эксперты-аналитики (ЭА) и академические исследователи (АИ).

Таким образом, каждому респонденту i можно поставить во взаимно однозначное соответствие вектор $x_i = (x_{i1}, x_{i2}, x_{i3})$, где x_{ij} – присвоенный i -ым респондентом удельный вес j -ой группе. Под паттерном в данном случае будем понимать определённые качественно похожие объекты (вектора x_i).

Для визуализации паттернов используется систему параллельных координат [216], состоящая в данном случае из трех равномерно распределённых линий, каждая из которых дает численную характеристику одной из следующих групп: ПУ, АИ и ЭА. На j -ой оси отмечается фактическое значение показателя x_{ij} , после чего данные значения соединяются отрезками, образуя некоторые ломанные. Другими словами, по значениям каждого вектора строится функция $f^x : R \rightarrow R; f^x(j) = x_{ij} \forall j = 1, 2, 3$. Как правило, в качестве f^x выбирается кусочно-линейная функция: $f^x(z) = \{k_j^x z + c_j^x; j \leq z \leq j+1\}, k_j^x j + c_j^x = x_{ij}, k_j^x(j+1) + c_j^x = x_{ij+1}, j = 1, 2$. Для интервалов $[j, j+1]$ ставится в соответствие пара $G_j^x = (k_j^x, c_j^x)$, а каждой функции f^x – вектор $G_j^x = (G_1^x, G_2^x, G_3^x)$.

В качестве меры близости будем использовать расстояние Хемминга и применим порядково-инвариантную паттерн-кластеризацию [217, 218], результат которой не зависит от выбора исходной последовательности показателей. С этой целью вводится кодировка каждого респондента, формируемая следующим образом: во-первых, попарно сравниваем ответы каждого респондента на процентное распределение по группам. Формируем дополнительные показатели r_{ij} так, что $r_{ij} = 1$ если $x_{ij} < x_{ij+1}, x_{ij} < x_{ij+1}, r_{ij} = 0$, если $x_{ij} = x_{ij+1}$, и $r_{ij} = 2$, если $x_{ij} > x_{ij+1}$, и сформируем кодировки c_i объектов

$c_i = \sum_{j=1}^2 \sum_{i=1}^3 10^{j-(i+1)} r_{ij}$. При нулевом расстоянии Хемминга между кодировками

двух респондентов они объединяются в единую группу, во всех других случаях – разделяются.

Основным результатом применения описанного выше метода к данным респондентов российского экономического сообщества является выделение 7 основных паттернов и их сопоставление с полученными ранее результатами с использованием метода латентных классов и исследованием распределения их рабочего времени.

При анализе всех 7 полученных паттернов возможен следующий вывод: большая часть респондентов представляют структуру российского экономического сообщества согласно собственному распределению рабочего времени между преподаванием в университетах, академическими исследованиями и проведением экспертно-аналитических работ.

4.2.2.4 Методы ранжирования журналов

Для анализа мнения различных групп респондентов о российских журналах по экономике были использованы ответы респондентов на вопросы второго блока:

- какие отечественные издания публикуют наиболее интересные для Вас статьи?
- оцените, пожалуйста, научный уровень журналов;
- какие российские журналы, на Ваш взгляд, считаются в экономическом сообществе наиболее престижными?

Респонденты выставляли каждому из 26 предложенных в анкете журналов ранги, где 3 обозначал наивысший ранг, 1 – самый низкий ранг и 0 – без оценки.

Для анализа мнения разных групп РАСЭ применялась двухступенчатая схема: на первом этапе строились три частных рейтинга (Интерес, Научный уровень и Престиж), представляющих мнения всех участников опроса по трем указанным выше вопросам, а затем на втором этапе с помощью различ-

ных правил принятия коллективных решений по этим трем частным рейтингам строилось итоговое агрегированное ранжирование журналов. Каждый метод был использован на всей выборке ответов респондентов, на подвыборках ПУ, ЭА и АИ, а также по подвыборкам «продвинутые» и «традиционные».

Построение частных ранжирований по трем вопросам на первом этапе проводилось двумя методами – методом Борда и методом порогового агрегирования.

На втором этапе из трех частных рейтингов (Интерес, Научный уровень и Престиж) необходимо составить один агрегированный рейтинг, для чего мы использовали несколько часто используемых правил коллективного выбора – три правила Коупленда, правило Кумбса и правило Нансона.

Правила Коупленда [219] основаны на построении числовой функции для каждого журнала на основе мажоритарного отношения. Правило Кумбса состоит в итеративном удалении журналов, которые считают наихудшим по максимальному количеству рейтингов. Процедура Нансона – это итеративная процедура, использующая схему правила Борда для промежуточных расчетов. Из списка альтернатив последовательно исключаются альтернативы, у которых ранг Борда больше, чем среднее значение этих рангов по всему множеству журналов.

4.2.3 Результаты

Ниже приведены основные результаты нашего исследования. Подробные описания и расчеты приведены в [220].

4.2.3.1 Структура РАСЭ по видам деятельности

Целесообразно выделить три латентных класса респондентов анкеты по профилю их основной деятельности: ПУ, АИ и ЭА. В качестве наблюдаемых признаков были выделены следующие 6 вопросов анкеты:

- 1) какой из указанных видов деятельности Вы считаете для себе наиболее важным?
- 2) что Вы читаете чаще – отечественные или зарубежные журналы?
- 3) оцените, пожалуйста, важность для российских экономистов знания экономико-математического аппарата (моделей равновесия, теории игр, эконометрики и т. п.).
- 4) какую квалификацию (степень) Вы имеете?
- 5) укажите, пожалуйста, тип организации Вашего основного места работы.
- 6) оцените, пожалуйста, в баллах, как распределяется Ваше рабочее время?

Проведем анализ латентных классов по всей выборке респондентов, а также отдельно по двум подвыборкам (Таблица 4.4): участники РЭК-2016 (далее обозначаем как РЭК) и участники XVIII Апрельской конференции (далее обозначаем АпрК). По выборке респондентов МЭФ-2017 не удастся выделить значимое разделение на три класса, поскольку все участники выбрали либо академические исследования, либо преподавание как наиболее важный для себя вид деятельности, что не позволяло получить значимое разбиение на классы.

Таблица 4.4 – Доли респондентов по классам

	$X = AI$			$X = ЭА$			$X = ПУ$		
	РЭК	АпрК	Вся выборка	РЭК	АпрК	Вся выборка	РЭК	АпрК	Вся выборка
$P(X = x)$	0.28	0.3	0.24	0.23	0.36	0.23	0.49	0.34	0.53

Видно, что по числу академических исследователей по всем трем выборкам есть согласованный результат на уровне 24-30%, однако результаты по участникам Апрельской конференции существенно завышают число экспертов-аналитиков (0,36 против 0,23 по всей выборке) в ущерб числу преподавателей. Интересно отметить, что наши результаты не совсем совпадают с тем, как сами респонденты себя оценивают: в вопросе о наиболее важном виде деятельности 41% респондентов указали «Академические исследования», 38% – «Преподавание» и 20% – «Экспертную деятельность». Анализ

латентных классов показывает, что часть респондентов, указавших академические исследования как наиболее важный вид деятельности, на самом деле являются представителями класса ПУ. Вероятно, это связано с набирающей обороты ориентацией ВУЗов на исследовательский профиль деятельности и ужесточением требований к преподавателям по публикационной активности.

4.3.3.2 Выделение группы «продвинутых» экономистов

Для выделения двух групп экономистов – «традиционных» экономистов, следующих традициям российской экономической школы, и «продвинутых» экономистов, обладающих современными знаниями в области экономики – мы оценивали, как респонденты оценивают распределение своего рабочего времени. Все множество респондентов можно разбить на 4 группы преподавателей в университетах (ПУ-1 - ПУ-4), 2 группы академических исследователей (АИ-1 и АИ-2), 2 группы экспертов-аналитиков (ЭА-1 и ЭА-2), а также отдельная группа, равномерно распределяющая рабочее время между тремя видами деятельности. Результаты представлены в Таблице 4.5. Предварительно отнесем к группе «продвинутых» тех респондентов, у которых определенный вид деятельности является преобладающим.

Таблица 4.5 – Разбиение экономического сообщества согласно распределению рабочего времени

Группа	Описание	Процент	Условное разбиение
ПУ-1	$ПУ > (АИ = ЭА)$	0,15	Продвинутые
ПУ-2	$ПУ > АИ > ЭА$	0,16	Продвинутые
ПУ-3	$(ПУ = ЭА) > АИ$	0,08	Традиционные
ПУ-4	$(ПУ = АИ) > ЭА$	0,08	Традиционные
АИ-1	$АИ > (ПУ = ЭА)$	0,24	Продвинутые
АИ-2	$(АИ = ЭА) > ПУ$	0,07	Традиционные
ЭА-1	$ЭА > АИ > ПУ$	0,07	Продвинутые
ЭА-2	$ЭА > (ПУ = АИ)$	0,06	Традиционные
Отдельная группа	$ПУ = АИ = ЭА$	0,03	

Из полученных групп «продвинутые» отберем тех респондентов, которые соответствуют 4 критериям:

- владеют хотя бы одним иностранным языком;
- имеют степень кандидата наук, доктора наук либо PhD;
- читают в основном зарубежные журналы (либо отечественные и зарубежные в равной степени);
- считают важным знание экономико-математического аппарата для экономистов.

Таким образом, мы можем отнести к группе «продвинутые» 21% всех респондентов. При этом получается, что в эту группу попадают 34,8% представителей класса АИ, 20,4% представителей класса ПУ и всего 8,3% из класса ЭА.

4.2.3.3 Результаты анализа паттернов

Применив описанный выше метод к исходному множеству респондентов X , мы выделили семь основных паттернов (рисунок 4.11).

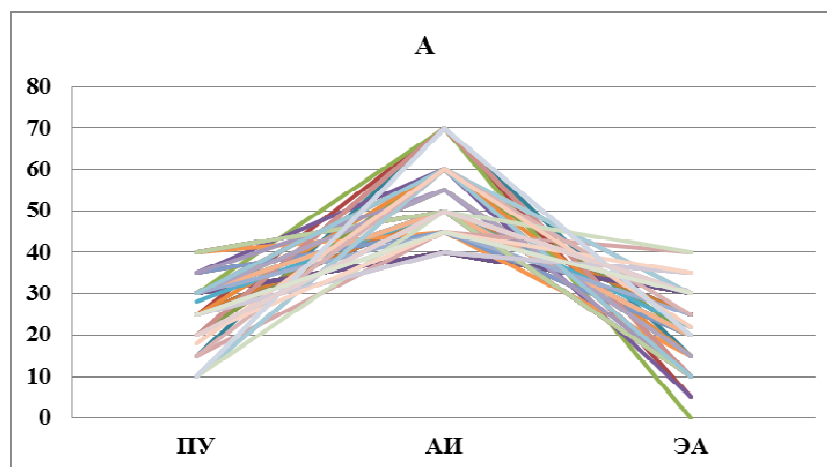


Рисунок 4.11 – Кусочно-линейные функции паттерна «А»

Паттерн «А» описывает для 11% принявших участие в опросе экономистов. Характерным для данного паттерна является преобладание академических исследователей в представлении респондентами структуры экономического сообщества. Если данный паттерн с результатами, полученными при

анализе латентных классов, можно отметить, то 38% респондентов отнесены к ПУ, 34% – АИ, 28% – ЭА. С другой стороны, если сопоставить с распределением рабочего времени респондентов, около 36% можно отнести к ПУ, 53% – к АИ, и 10,5% – ЭА. Таким образом, можно отметить, что в паттерне «А» преобладают специалисты, которые можно отнести к АИ, и именно на академические исследования респонденты, вошедшие в данный паттерн, тратят больше всего рабочего времени. Наименьшее число респондентов (10,5%) можно отнести к ЭА. И именно на экспертно-аналитическую работу респонденты, принадлежащих паттерну «А», тратят меньше всего рабочего времени.

Мужчин описывают 48% кусочно-линейных функций данного паттерна, женщин – 52%. Если анализировать возраст респондентов, вошедших в данный паттерн, можно отметить, что 6% принадлежат к возрастной группе до 25 лет; 26% – от 25 до 45; 32% – от 45 до 60; 33% – более 60 лет; 3% респондентов не указали свой возраст.

Паттерн «В» характерен для 46,5% респондентов, считающих, что подавляющее большинство экономического сообщества должно состоять из преподавателей в университетах. В отличие от паттерна «А», большая часть респондентов считают наиболее важным для себя преподавание в университетах (41%). Второе место занимают академические исследования (36%), третье – экспертно-аналитическая работа (16%). Около 6% не указали наиболее важный вид деятельности либо указали иное

Паттерн «С» содержит 18% кусочно-линейных функций, для которых характерно преобладание показателей «Преподаватели в университетах» и «Эксперты-аналитики» над показателем «Академические исследователи». 58% респондентов мужчины, 42% – женщины.

Паттерн «D» содержит менее 5% от общего числа полученных кусочно-линейных функций. Отличительной особенностью являются близкие значения показателей «Преподаватели в университетах» и «Академические исследователи».

дователи», а также их существенное преобладание над показателем «Эксперты-аналитики».

Паттерн «Е» описывает 4% респондентов, участвующих в опросе. Отличительной особенностью данного паттерна является преобладание удельного веса экспертов-аналитиков в общем представлении респондентами структуры сообщества российских экономистов, а также равенство удельных весов преподавателей в университетах и академических исследователей.

Паттерн «F» состоит из 3% респондентов, ответы которых учитывались при построении кусочно-линейных функций. Особенностью шестого паттерна является относительно небольшая разница в значениях всех показателей (значения показателя «ПУ» находятся в пределах [20;40]; «АИ» – [30;40]; «ЭА» – [30;40]).

Паттерн «G» описывает 2% респондентов и отличается преобладанием показателя «ЭА». Значения показателя «АИ» больше либо равны значениям показателя «ПУ».

4.2.3.4 Анализ мнения разных групп РАСЭ о российских экономических журналах

Результаты построения агрегированных рейтингов по всей выборке респондентов, по группе АИ и по группе «Advanced» приведены в [220]. Расчеты по разным классам респондентов дают в целом согласованные результаты, есть значимые отличия только для класса «Академические исследователи». Для этой группы респондентов журналы "Квантиль (РЭШ)", "Общественные науки и современность", "Прикладная эконометрия", "Пространственная экономика", "Экономика и математические методы" и "Экономическая наука современной России" получили более высокие позиции в рейтингах в сравнении с позициями этих журналов в рейтингах остальных групп, а такие журналы, как "Вестник Института экономики РАН", "Проблемы теории и практики управления", "Финансы", "Финансы и бизнес" и "Эконо-

мист", наоборот, получили более низкие позиции в рейтинге АИ, нежели в рейтингах остальных групп.

По группе «Advanced» в целом более высокие позиции (по сравнению с рейтингом по всей выборке) в рейтинге получили журналы "Квантиль (РЭШ)", "Прикладная эконометрика", "Пространственная экономика", "Российский журнал менеджмента", "Экономика и математические методы" и "Экономическая наука современной России", а более низкие позиции в рейтинге – "Вестник Института экономики РАН", "Вопросы государственного и муниципального управления", "Общество и экономика", "Проблемы теории и практики управления", "Финансы" и "Финансы и бизнес".

В целях сравнения разных рейтингов и составления одного итогового ранжирования мы сравнивали средние и максимальные ранги журналов. Можно выделить подгруппу ТОП-журналов по нашему списку из 26 журналов: "Вопросы экономики", "Журнал Новой экономической ассоциации", "Экономический журнал ВШЭ", "Мировая экономика и международные отношения" и "Экономика и математические методы" – эти журналы во всех рейтингах по обеим выборкам имеют ранг не ниже 5. Также мы условно разделили нашу выборку из 26 журналов на 4 группы (А+, А, В+ и В), ориентируясь на указанные показатели (см. Таблица 12 в [220]).

Не стоит забывать, что мы сознательно в целях отработки данной методики рейтингования и удобства наших респондентов взяли небольшую и разнородную выборку журналов, ориентированных на разную читательскую аудиторию. Поэтому более целостным является подход с комплексной оценкой рейтингов журналов по разным параметрам и разным группам респондентов, хотя, разумеется, в таком случае воспринимать весь объем информации и сравнивать различные журналы становится затруднительно.

4.2.4 Некоторые выводы

Предложены несколько методов составления рейтингов научных журналов по экономике с учетом неоднородности сообщества российских эконо-

мистов на основе результатов анкетирования, проведенного в конце 2016 – начале 2017 гг. Первый рассмотренный подход использует вероятностную оценку соотнесения каждого респондента к одной из трёх групп: преподаватели в университетах, академические исследователи и эксперты аналитики. Второй предполагает исследование распределения рабочего времени респондента для предварительного разбиения и выделения группы «продвинутых» экономистов на основе четырех критериев (знание иностранного языка, наличие ученой степени, признание важности экономико-математического аппарата и чтение зарубежных научных журналов). Третий подход предполагает распределение респондентов по паттернам на основе их представления о структуре экономического сообщества. Данные подходы позволяют получить представление о структуре экономического сообщества России и строить рейтинги журналов на основе полученных данных.

Результаты исследований отражены в работах [221, 222, 223].

4.3 Развитие методов анализа паттернов

4.3.1 Введение

На сегодняшний день, в связи с накоплением большого рода различных данных, очень востребованными становятся различные методы их анализа, позволяющие не только выявлять различные закономерности, но и затрачивать на данный процесс минимально возможное количество времени. При всё большем накоплении различного рода известных параметров, необходимым условием успешного их применения при максимально точном решении поставленных задач является развитие как можно более мощных алгоритмов, позволяющих в сжатые сроки провести их обработку.

Поскольку настоящие исследования относятся к данному направлению, для частичного решения задачи поиска закономерности в данных при ограниченных временных ресурсах, в работе развиваются существующие разновидности методов анализа паттернов, и частично решается проблема устой-

чивости конечного разбиения при различных исходных последовательностях показателей.

Ранее методы анализа паттернов были успешно применены при решении множества прикладных задач в самых различных областях: макроэкономике [224, 225], анализе банковской сферы [226], политике [227], менеджменте и управлении персоналом [228], анализе данных науки, образования и инновационной деятельности [229].

Предпосылкой к данному исследованию послужила следующая идея: возможно ли создать определенный метод выявления паттернов среди разнородных данных, результат которого не будет зависеть от выбора последовательности входных показателей, описывающих исследуемые объекты? И какими свойствами будет обладать данный метод?

4.3.2 Анализ паттернов: общее описание

Для начала, приведем определение из [230]: «Анализ паттернов – это новая область анализа данных, связанная с поиском взаимосвязей исследуемых объектов, построением их классификации и исследованием развития объектов во времени». В общем случае, в качестве исходных данных представлено некоторое множество объектов $X: |X| = k$, каждый из которых характеризуется значениями n переменных. Для простоты поставим во взаимно однозначное соответствие каждому объекту $x_i \in X$ вектор $x_i = (x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{in})$, где x_{ij} – значение j -ого показателя i -ого объекта. Для визуализации используется система параллельных координат [231], состоящая, в данном случае, из n равномерно распределенных линий, характеризующих n показателей, на которых отмечаются их фактические (либо нормированные) значения.

Для каждого $x_i \in X$ строится функция $f^x: R \rightarrow R; f(j) = x_{ij} \forall j = 1, 2, \dots, n$. При этом

$$f^x(y) = \{k_j^x y + g_j^x; j \leq x \leq j + 1\}$$

$$k_j^x j + g_j^x = x_{ij}$$

$$k_j^x (j + 1) + g_j^x = x_{ij+1}, j = 1, \dots, n - 1,$$

а для интервалов $[j; j+1]$ ставится в соответствие пара $P_j^x = (k_j^x, g_j^x)$. Задачей анализа является выявление определенных паттернов. Другими словами, разбиение множества X на ν непересекающиеся подмножества, объекты которых близки по некоторой мере близости d . Основные меры близости описаны в [232, 233].

Как правило, выделяются две разновидности анализа паттернов: динамический и статический. Для динамического анализа характерным является построение траекторий развития объектов во времени, тогда как первая разновидность метода предполагает исследование объектов в статическом виде.

Формально, метод анализа паттернов можно разделить на следующие этапы:

1. Исходный этап:
 - 1.1. Моделирование исходной системы показателей;
 - 1.2. Сбор данных;
 - 1.3. Качественная проверка найденных данных;
 - 1.4. Проведение корреляционного анализа используемых данных;
 - 1.5. Агрегирование данных;
 - 1.6. Составление базовой системы показателей.
2. Аналитический этап:
 - 2.1. Построение объектов в виде определённых функций;
 - 2.2. Проведение кластеризации;
 - 2.3. Интерпретация результатов кластеризации.
3. Динамический этап анализа:
 - 3.1. Построение траекторий, выражающих развитие объектов во времени;
 - 3.2. Выявление динамических групп.

Для проведения динамического анализа необходимым условием является построение некоторых кривых, каждая из которых характеризует один из изучаемых объектов во времени. С этой целью каждому объекту приписывается значение паттерна y_i на каждом рассматриваемом периоде (обозначим как y_i^t). Таким образом, каждый объект x_i описывается некоторой траекторией развития $\{y_i^1, y_i^2, \dots, y_i^t, \dots, y_i^w\}$, где w – общее количество наблюдений. Визуальное представление данной траектории позволяет определять устойчивое/неустойчивое развитие объекта x_i во времени.

4.3.3 Анализ паттернов, основанный на парном сравнении при фиксированной последовательности показателей

Порядково-фиксированная паттерн-кластеризация является разновидностью метода анализа паттернов, в связи с чем перед использованием предложенной методологии необходимым условием является выполнение исходного этапа анализа паттернов, а также построение исследуемых объектов в виде определённых функций. Данная методология построена на парном сравнении исходных показателей объектов. При этом возможны три различных варианта: $x_{ij} < x_{ij+1}$, $x_{ij} = x_{ij+1}$, $x_{ij} > x_{ij+1}$ (x_{ij} – j -ый признак i -ого объекта). На основании данных сравнений формируется определённая кодировка исследуемого объекта следующим образом:

$$z_{ij} = \text{при } x_{ij} < x_{ij+1};$$

$$z_{ij} = \text{при } x_{ij} < x_{ij+1};$$

$$z_{ij} = \text{при } x_{ij} < x_{ij+1}.$$

Используя данные значения, объекты соотносятся в разные группы только при их различии на любом этапе сравнения. В противном случае, объекты относятся к единой группе (единому паттерну). Проиллюстрируем работу данного метода на простом примере. Исследуются 3 банка (обозначим как x_1 , x_2 и x_3), причем $x_1 = (42; 21; 41)$, $x_2 = (46; 12; 46)$, $x_3 = (22; 61; 21)$. Изобразим исследуемые банки в системе параллельных координат (рисунок 4.12).

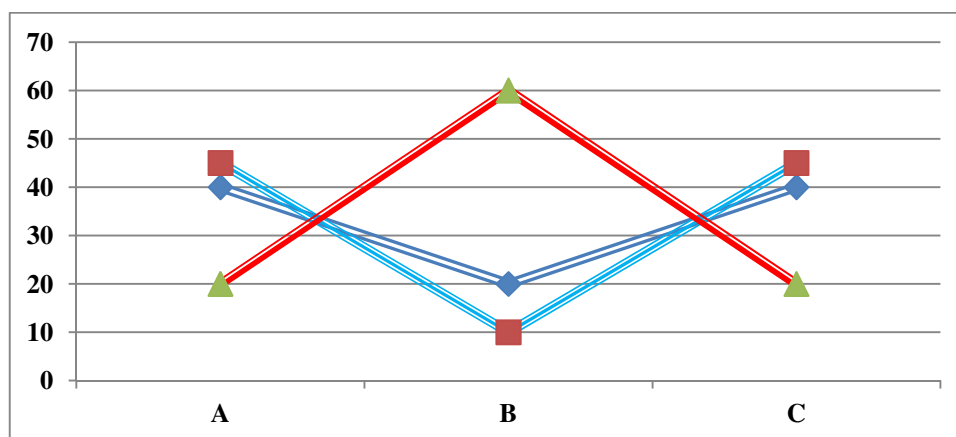


Рисунок 4.12 – Кусочно-линейные функции гипотетических банков

Использование порядково-фиксированной паттерн-кластеризации приведет к следующему результату: банки 1 и 2 отнесены будут объединены в один паттерн, банк 3 – в другой.

Подобный метод, хотя и не решает проблему получения устойчивых результатов при различных последовательностях показателей, но существенно помогает в предварительной обработке данных при наличии больших временных ограничений.

4.3.4 Анализ паттернов с независимыми от выбора исходной последовательности показателей результатами

Порядково-инвариантная паттерн-кластеризация является разновидностью метода анализа паттернов, предназначенная для выявления устойчивых паттернов. В связи с этим, также исследуется множество X состоящее из k элементов (обозначаются через x_i), каждый из которых описывается n параметрами. При необходимости рассмотрения всех возможных перестановок показателей потребуется изучить $n!$ возможных вариантов, что, очевидно, потребует большого количества времени. Исследование усложняется ещё тем фактом, что различные перестановки показателей приводят к различным кусочно-линейным функциям, и, как следствие, к различным результатам. Для получения устойчивого паттерна в [234] предложена порядково-

инвариантная паттерн-кластеризация, однако, в данной работе не приведено подробное описание ее свойств.

Для практической реализации порядково-инвариантной паттерн-кластеризации предложено использовать кодировки объектов (на подобии порядково-фиксированной паттерн-кластеризации). Группы объектов, полученных с использованием порядково-инвариантной паттерн-кластеризации, будем называть порядково-инвариантными паттерн-кластерами.

Порядково-инвариантная паттерн-кластеризация обладает рядом свойств, описанных в [235]:

Свойство 1. При использовании порядково-фиксированной паттерн-кластеризации один объект не может принадлежать двум кластерам.

Свойство 2. При использовании порядково-инвариантной паттерн-кластеризации один объект не может принадлежать двум кластерам.

Свойство 3. Если для любых объектов существует порядок показателей, формирующий строго монотонно убывающую/возрастающую последовательность, то такие объекты образуют порядково-инвариантный паттерн-кластер.

Свойство 4. Показатели объектов, объединенных в единый порядково-инвариантный паттерн-кластер, всегда могут образовать монотонно невозрастающую/неубывающую последовательность.

Свойство 5. Если два объекта x_a и x_b принадлежат одному порядково-инвариантному паттерн-кластеру, то и объект, показатели которого сформированы на основе суммы соответствующих показателей объектов x_a и x_b , также принадлежит данному кластеру.

Свойство 6. Если объект x_a принадлежит одному порядково-инвариантному паттерн-кластеру, то и объект $x_\lambda = \lambda x_a$ ($\lambda > 0$) также принадлежит данному кластеру.

Доказательство данных свойств приведено в [236]. Важным их следствием (и одним из основных направлений исследований в работе) является возможность использования алгоритмов сортировки для нахождения поряд-

ково-инвариантных паттерн-кластеров, и, как следствие, существенное снижение вычислительной сложности порядково-инвариантной паттерн-кластеризации для работы с большими объемами данных.

4.3.5 Некоторые выводы

Продолжена работа над изучением основных свойств методов анализа паттернов, основанных на парном сравнении показателей, конечный результат которых не зависит от выбора их исходной последовательности. Сформулированы и доказаны шесть свойств, что позволяет использовать алгоритмы сортировки данных для практической реализации порядково-инвариантной и диффузионно-инвариантной паттерн-кластеризаций. Использование подобных алгоритмов, в свою очередь, существенно снижает вычислительную сложность новых методов анализа паттернов и позволяет работать с крупными массивами данных.

4.4 Разработка нового частотного механизма оценивания структурной сложности графов

4.4.1 Введение и постановка задачи

Представленный подраздел отчёта посвящён развитию нового подхода к понятию сложности графа и приложениям этого понятия, предложенного в публикациях [237, 238, 239, 240, 241] и цитированных в них работах. Основные отличия между данной и предшествующими работами могут быть кратко выражены следующим образом.

1. В предыдущих вариантах основной этап – построение пути, увеличивающего максимальную частоту в рёбрах графа – осуществлялось много раз, однако дихотомия строилась только один раз, после выполнения заданного количества основных этапов. Для построения семейства дихотомий весь процесс повторялся многократно. В новой версии алгоритма каждая дихотомия строится сразу после очередного выполнения основного этапа. Это и приво-

дит к значительному – в 100 и более раз ускорению работы алгоритма. Это не просто «техническое» преимущество. Быстрый алгоритм позволяет обработать данные такого объёма, который был ранее недоступен. Не менее важна и возможность более тщательного подбора параметров и анализа деталей алгоритма.

2. В предыдущих публикациях анализ курсов акций, находящихся на рынке в день x , осуществлялся с учётом времени дальнейшего пребывания этих акций на рынке. В настоящем отчёте для анализа функционирования фондового рынка в каждый день x рассматривались только данные, непосредственно доступные в этот день. Это обстоятельство усложняет работу, делая в то же время её более реалистичной.

В настоящем отчете все необходимые понятия и определения (так же, как некоторые новые) введены в рассмотрение. Материал не требует от читателя знакомства с предыдущими публикациями.

Этот подраздел отчета структурирован следующим образом. Пункт 4.4.2 посвящён центральному понятию работы – новому частотному алгоритму дихотомии. В пункте 4.4.3 вводится формальное понятие сцепленности графа. В пункте 4.4.4 это понятие применяется для анализа фондового рынка. Другие примеры применения предложенного подхода в данном отчете не рассмотрены. Напомним, что некоторые из них рассмотрены в публикациях [237, 238, 239, 240]. В пункте 4.4.5 сделаны некоторые выводы по этому подразделу.

4.4.2 Частотный алгоритм дихотомии

В настоящем отчете предлагается новый алгоритм построения семейства дихотомий произвольного неориентированного простого графа (поскольку графы других типов в работе не рассматриваются, будем далее пользоваться термином «граф» без дополнительных уточнений).

Дадим описание общей структуры частотного алгоритма дихотомии, далее для краткости обозначаемого ЧАД.

1. На предварительном этапе каждому ребру исходного графа сопоставляется равномерно распределённое случайное целое число из диапазона $[1 - d_{\max}]$ (типичное значение $d_{\max} = 10$). Далее числа, сопоставляемые рёбрам графа, будут называться частотой соответствующего ребра. Находим также максимальную (по всем рёбрам графа) частоту ребра F_{\max} .

2. Как и в предыдущих публикациях [239, 240, 241], сам ЧАД состоит в многократном последовательном выполнении основного этапа. В результате каждого выполнения этого этапа частоты некоторых рёбер увеличиваются на 1, а во всех остальных рёбрах частоты не изменяются. Опишем основной этап подробнее.

1. С помощью стандартного датчика равномерно распределённых случайных чисел выбираются две различные вершины графа.

2. Для двух выбранных вершин алгоритмом Дейкстры находится соединяющий их кратчайший путь. Длиной ребра является его текущая частота. Длина пути равна длине его самого длинного ребра, а не сумме длин всех его рёбер. Хорошо известно, что алгоритм Дейкстры применим и в подобных случаях, с единственным изменением: при определении продолженного пути вместо суммы длин начального участка и добавляемого ребра записывается максимальное из этих же двух чисел.

3. Определяется максимальная частота рёбер F_p в найденном на шаге 2 пути.

4. Если $F_p = F_{\max}$, то переходим к шагу 5. В противном случае осуществляется модификация частот: число 1 прибавляется к частотам всех рёбер пути, найденного на шаге 3. Текущее выполнение основного этапа завершено; переходим к очередному выполнению основного этапа.

5. При выполнении условия $F_p = F_{\max}$ рёбра с максимальной частотой образуют разрез графа. Если бы это было не так, то в силу минимаксного критерия алгоритм Дейкстры нашёл бы путь, не проходящий по рёбрам с максимальной частотой.

5.1. Находим компоненты связности исходного графа, которые получились бы при удалении всех рёбер с максимальной частотой.

5.2. Компоненту с максимальным числом вершин объявляем 1-ой частью текущей дихотомии графа, а другую компоненту (если она одна) или объединение всех остальных компонент (если их больше одной) объявляем 2-ой частью текущей дихотомии.

5.3. Запоминаем найденную дихотомию.

5.4. Как и в предыдущем случае, осуществляется модификация частот: число 1 прибавляется к частотам всех рёбер пути, найденного на шаге 3. Кроме этого, увеличивается и максимальная частота в рёбрах: $F_{\max} = F_{\max} + 1$. Текущее выполнение основного этапа завершено; переходим к очередному выполнению основного этапа.

Проиллюстрируем некоторые детали основного этапа. Выбранная пара вершин $\{a, b\}$ соединена путём (показанным жирной линией), частоты в рёбрах которого затем увеличиваются на 1 (см. рисунок 4.13):

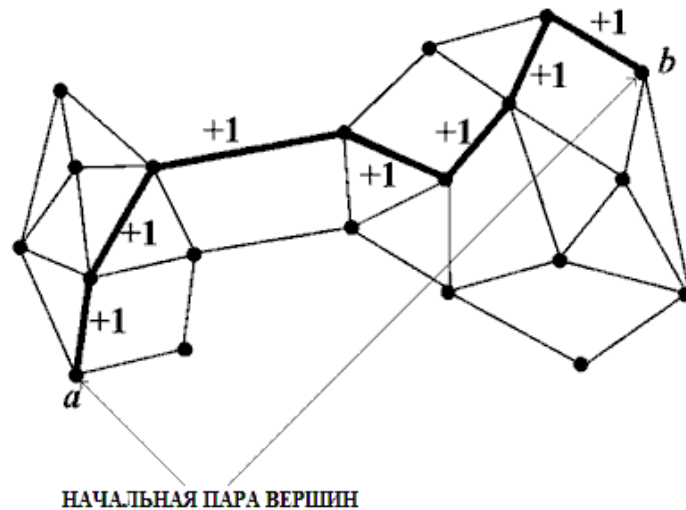


Рисунок 4.13 – Иллюстрация основного шага

Непосредственно перед выполнением шага 2 основного этапа может иметь место одна из трёх ситуаций, обозначенных как случаи А, В, С. На рисунках 4.14 – 4.16 жирные линии представляют рёбра с максимальной (по всем рёбрам графа) частотой, а тонкие линии представляют собой все остальные рёбра графа.

В случае А (рисунок 4.14) множество всех рёбер с максимальной частотой не содержит ни одного разреза графа. Поэтому найденный (алгоритмом Дейкстры на минимакс) путь не содержит ни одного ребра с максимальной частотой, в силу минимаксного определения длины пути. Следовательно, после модификации частот максимальная из частот рёбер не изменится.

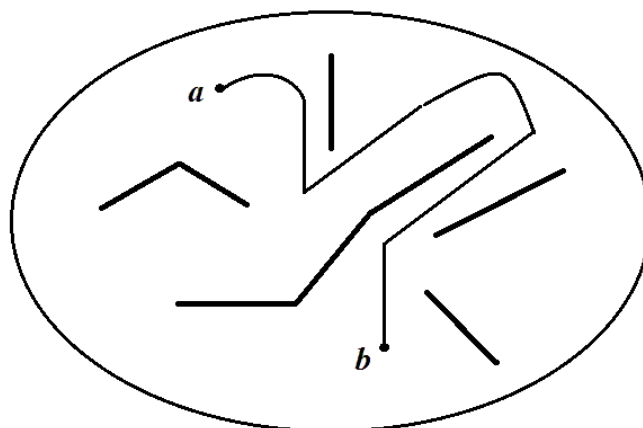


Рисунок 4.14 – Случай А

В случае В (рисунок 4.15) множество всех рёбер с максимальной частотой содержит разрез графа, но найденный путь не содержит ни одного ребра с максимальной частотой (обе вершины расположены по одну сторону от разреза). Следовательно, и в этом случае максимальная из частот рёбер не изменится.

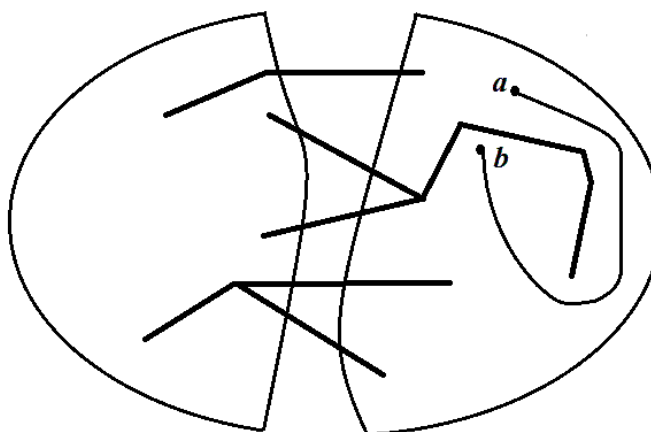


Рисунок 4.15 – Случай В

В случае С (рисунок 4.16) множество всех рёбер с максимальной частотой содержит разрез графа, и обе вершины расположены по разные стороны разреза. Поэтому путь, соединяющий эти две вершины, должен содержать

хотя бы одно ребро с максимальной частотой. Следовательно, после выполнения основного этапа максимальная из частот рёбер в графе должна увеличиться на 1.

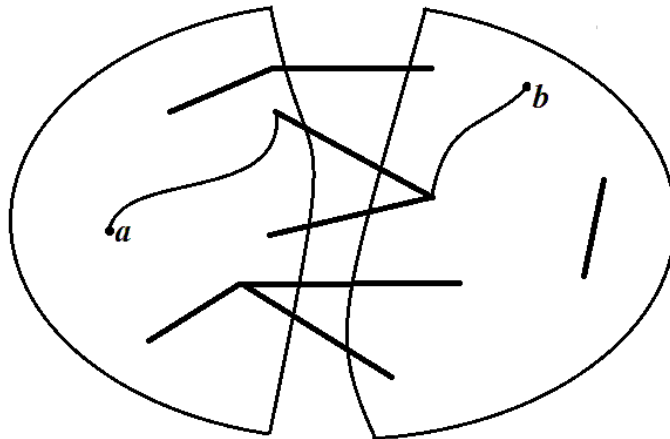


Рисунок 4.16 – Случай С

Понятно, что в любом случае после выполнения основного этапа частота увеличится на 1, по крайней мере, в одном ребре, и ни в каком ребре не уменьшится. Это означает, что после некоторого числа выполнений основного этапа максимальная частота в рёбрах обязательно возрастет на единицу. При этом перед выполнением шага 3 рёбра с максимальной частотой содержат разрез исходного графа.

Вернёмся к общей структуре алгоритма. Перед его выполнением задаётся число M требуемых дихотомий. Алгоритм останавливается, когда построено заданное число дихотомий. Рассуждения из предыдущего абзаца ясно показывают, что алгоритм остановится при любом заранее заданном числе дихотомий. Понятно также, что продолжительность работы алгоритма приблизительно пропорциональна желаемому числу дихотомий.

4.4.3 Сцепленность графа

Для каждого значения единственного параметра M , результатом работы описанного в разделе 1 алгоритма будет семейство M дихотомий, т.е. разбиений исходного графа на две части. Для многих графов все найденные дихотомии могут оказаться совпадающими, но это неверно в общем случае. На-

помним, что в алгоритме многократно используется датчик случайных чисел для последовательного выбора пар вершин. Поэтому в некоторых случаях, особенно практически важных, различные прогоны алгоритма производят не всегда совпадающие дихотомии. Эти случаи не являются досадными ошибками или недостатками метода. Более того, можно утверждать, что различные дихотомии естественно возникают при исследовании многих сложных систем, особенно тех, чьё функционирование определяется человеческой деятельностью. Следовательно, представляется целесообразным рассмотреть – как важное обобщение обычной задачи дихотомии – построение семейства дихотомий вместо единственной дихотомии. Построенное семейство характеризует исходную систему, моделируемую графом. Более того, в таких ситуациях сами по отдельные дихотомии, формирующие указанное семейство, не вызывают интереса. Оказалось более целесообразным сосредоточить внимание на свойствах построенного семейства дихотомий, как целостного объекта.

Введём необходимые понятия и определения, позволяющие описать интересующие нас свойства построенного семейства дихотомий. Семейство всех дихотомий, найденных алгоритмом, состоит из M элементов. Некоторые из них могут совпадать, а некоторые – быть различными. Пусть среди M дихотомий дихотомия d_p встречается m_p раз ($p = 1, \dots, t$), где

$$\sum_{p=1}^t m_p = M. \quad (4.4)$$

Числа m_p ($p = 1, \dots, t$), входящие в (4.4) подсчитываются непосредственно по семейству M дихотомий, найденных ЧАД. Для этого достаточно воспользоваться очевидным алгоритмом сравнения двух дихотомий. Расположим числа m_p ($p = 1, \dots, t$) в невозрастающем порядке. Этот упорядоченный набор представляет собой распределение дихотомий для данного графа G при заданном параметре M .

Интуитивно понятно, что свойства этих распределений отражают свойства исходных графов и, следовательно, свойства рассматриваемых систем, представленных этими графами. Они имеют различные содержательные ин-

терпретации, но в общем описывают сложность, запутанность, неясность, неопределенность и другие трудно формализуемые, хотя и весьма важные свойства разнообразных систем.

Предположим, что исходный граф имеет две связные компоненты. В этом случае ЧАД M раз строит единственную дихотомию, части которой соответствуют этим двум компонентам, и датчик случайных чисел совсем не используется. Поэтому все построенные дихотомии совпадают друг с другом. Распределение состоит из единственного числа M .

Теперь рассмотрим другой крайний случай. Пусть G является полным графом с n вершинами. Из симметричности этого графа следует, что каждый новый прогон ЧАД строит новую дихотомию на две части с равным числом вершин для чётных n и на две части с k и $k+1$ вершиной ($n = 2k+1$) для нечётных n . При достаточно больших n , например, при $n > 50$, шанс получить две совпадающие дихотомии практически равен нулю. Поэтому все M построенных дихотомий различны, и распределение состоит из M единиц.

Можно сказать, что крайние ситуации 1-го типа, в которых может быть построена только одна дихотомия, встречаются в графах, имеющих специальную «бинарную» структуру – есть два подмножества вершин, соединённых сравнительно небольшим числом рёбер. Вышеописанный ЧАД как раз выявляет эту структуру (конечно, если она есть).

Крайняя ситуация 2-го типа соответствует графам, имеющим «хаотическую» структуру. Это значит только то, что все возможные выходы вышеописанного ЧАД равноправны и равновероятны. Эта структура также выявляется при использовании ЧАД.

Имеются и другие, менее очевидные, примеры обоих крайних случаев, также выявляемые с помощью ЧАД. Но в большинстве графов, которые возникают в практических и теоретических задачах, обе структуры представлены частично и одновременно. Введём формальный показатель, который количественно выражает присутствие структур обоих типов в одном графе. Положим

$$E = -\sum_{p=1}^t \mu_p \ln(\mu_p), \text{ где } \mu_p = m_p/M. \quad (4.5)$$

Число E представляет собой обычную энтропию разбиения M элементов на t частей, состоящих из совпадающих элементов. Очевидно, что минимально возможное значение E равно 0, в то время как максимально возможное значение равно $\ln(M)$. В первом случае имеется только одна группа совпадающих элементов, во втором случае имеется M групп, состоящих из одного элемента каждый.

Положим

$$I = E/\ln(M),$$

где E определено формулой (4.5). Величина I является случайной, поскольку случайный датчик многократно использовался на каждом из M основных этапов ЧАД. По построению, $0 \leq I \leq 1$. Обратим внимание на то, что I является случайной величиной, но не средним значением. Формула (4.5) ясно показывает, что E не равно никакому среднему значению, относящемуся к отдельному этапу ЧАД, и то же самое верно и для I . Чтобы подчеркнуть зависимость I от числа основных этапов M будет использоваться обозначение $I(M)$. Математическое ожидание $\bar{I}(M)$ этой случайной величины можно назвать сцепленностью (англ. “entanglement”) рассматриваемого графа. В цитированных выше работах аналогичное понятие называлось декомпозиционной или дихотомической сложностью исходного графа G на уровне M . Однако это название не кажется удачным в связи с широкой распространенностью термина «сложность» в теории графов и в других разделах чистой и прикладной математики.

Предполагается, что введенное понятие сцепленности графа отражает важные свойства, упомянутые ранее в этом разделе. Но само это значение неизвестно и предложенный частотный алгоритм построения семейства дихотомий (ЧАД) даёт только его аппроксимацию $I(M)$. Оценка дисперсии этой аппроксимации была рассмотрена в [241]. Сравнительно низкие значения дисперсии обеспечивают воспроизводимость и надёжность эксперименталь-

ных результатов, описывающих поведение фондовых рынков. Более подробно эта тематика, связанная с фондовыми рынками, рассмотрена в следующем пункте.

4.4.4 Анализ фондовых рынков

Хотя все введенные выше понятия носят общий характер и относятся к произвольным неориентированным графам, специальные – так называемые рыночные графы, порожденные фондовыми рынками, в этом разделе рассмотрены подробнее.

Рассмотрен рынок S&P-500 (500 крупнейших компаний в США). Прежде всего, дадим описание хорошо известной графовой модели произвольного фондового рынка. Объекты соответствуют рассматриваемым (в течение некоторого периода) акциям. Расстояние между двумя акциями определяется следующим образом.

1. Определим базисный минимальный период, состоящий из l последовательных дней. Все данные, найденные для периода $x, x-1, \dots, x-l+1$, сопоставляются дню x . Положим длину l рассматриваемого периода равной 16. Такой выбор определен содержательными соображениями: для короткого периода данные слишком вариативны, для длинного периода – слишком сглажены.

2. Цены всех акций на момент закрытия биржи рассмотрены для дней $x, x-1, \dots, x-l+1$. Матрица R попарных коэффициентов корреляции считается, исходя из этих цен.

3. Расстояния d_{ij} между двумя акциями (например, i и j) определяется формулой $d_{ij} = 1 - r_{ij}$, где r_{ij} – соответствующий элемент матрицы R . Указанное расстояние d близко к 0 для «очень похожих» акций и близко к 2 для «очень непохожих» акций. Поэтому матрица $D = (d_{ij})$ рассматривается как матрица непохожести.

4. Устанавливается взаимно-однозначное соответствие между вершинами графа и рассматриваемыми акциями. Для каждого объекта a все осталь-

ные вершины упорядочиваются в порядке возрастания (неубывания) расстояния до a . Определяются первые четыре вершины в этом списке, а также все вершины, расстояния от которых до a совпадают с расстоянием от четвертого в этом списке объекта. Вершина, соответствующая объекту a , соединяется ребрами со всеми этими вершинами. Легко видеть, что при указанном построении граф не зависит от конкретной нумерации объектов в упорядоченном указанным образом списке. Число ближайших вершин (здесь 4) является параметром при построении графа.

На первый взгляд, описанная хорошо известная конструкция не вызывает никаких вопросов. Предположим теперь, что начальный день x – это сегодняшнее число. Это означает, что ничего о поведении биржи в последующем неизвестно. Заметим, что это относится не только к курсам акций, но и к самому составу S&P-500. Зато о прошлом, естественно, известно всё.

Поэтому для построения рыночного графа можно взять все акции из сегодняшнего списка S&P-500, которые существуют на этом рынке не менее 16 дней. Поскольку изменение списка S&P-500 случается достаточно редко (5 – 15 раз в год), то таких акций бывает 499 или 500. Обозначим построенный указанным выше способом граф через G_0 , а нормированную энтропию этого графа – через I_0 .

Аналогично определим рыночные графы G_j , построенные на курсах тех акций, которые были на рынке в исходный день x и существовали на нём до того, как на рынке появились j акций, просуществовавших на нём до дня x включительно. Таким образом, граф G_{20} содержит 480 вершин (из 500, входящих в S&P-500 в день x), соответствующих тем 480 акциям, которые присутствовали в списке S&P-500 раньше, чем 20 других акций, также входящих в S&P-500 в день x .

Теперь приведём некоторые результаты вычислений энтропий для разных значений j для двух начальных дней – 08.09.26 и 01.03.07, накануне активной фазы двух больших кризисов, а также для нескольких предшествующих дней, рассматриваемых как начальные.

Таблица 4.6 – Нормированная энтропия перед кризисом 2008 года при разном числе вершин у рыночного графа

j	16.09	17.09	18.09	19.09	20.09	21.09	22.09	23.09	24.09	25.09	26.09
0	0.9694	0.9764	0.9769	0.9767	0.9739	0.9753	0.9740	0.9775	0.9753	0.9770	0.9781
25	0.0699	0.4191	0.3098	0.7519	0.5258	0.5704	0.5790	0.3910	0.5308	0.3280	0.4584
29	0.9732	0.9760	0.9735	0.9759	0.9650	0.9736	0.9718	0.9711	0.9755	0.9743	0.8227

Таблица 4.7 – Нормированная энтропия перед кризисом 2001 года при разном числе вершин у рыночного графа

j	25.02	26.02	27.02	28.02	01.03	02.03	03.03	04.03	05.03	06.03	07.03
25	0.5014	0.9108	0.8420	0.6389	0.6474	0.7595	0.5309	0.1124	0.0244	0.3606	0.4834
28	0.1186	0.0216	0.3667	0.6064	0.3137	0.1017	0.1459	0.2926	0.1788	0.4669	0.5411

Таблицы 4.6 и 4.7 показывают, что как при малых, так и при больших значениях j , т. е. при разном числе рассматриваемых акций – значения нормированной энтропии близки к максимально возможному значению 1 и не дают сколько-нибудь содержательного представления о рассматриваемой ситуации. В то же время промежуточные значения параметра j , соответствующие примерно 3-х или 4-х летнему минимальному сроку непрерывного присутствия акций на рынке, дают более упорядоченную картину. Указанные явления присутствуют как вблизи, так и вдали (по времени) от больших кризисов.

4.4.5 Некоторые выводы

Приведенные выше соображения позволяют сформулировать некоторые направления дальнейших исследований, в конечном счете, направленных к прогнозированию больших кризисов. Следует проанализировать связь между добавлением/удалением некоторых акций из списков S&P-500 и типичными значениями энтропии соответствующих графов. Следует проанализировать список акций, хотя и входящих в список S&P-500 на данный день, но включенных в него на основе информации о будущем поведении этих акций. Энтропии именно таких акции и давали паттерны, которые позволяли бы судить о близости большого кризиса, если бы упомянутая информация была

известна a priori. Можно сравнивать эти списки, со списками, известными до исходного дня, полученными при изменении параметра j . Обратим внимание, что эти, как и многие другие исследования в этом направлении, становятся возможными только благодаря новому алгоритму построения семейства дихотомий, описанному в этом подразделе отчёта

С более «технической» точки зрения представляет интерес обобщение предложенного варианта ЧАД на случай декомпозиций более чем на два подграфа. Такие декомпозиции могут быть полезными при анализе голосований в различных органах.

4.5 Исследования по кластерному анализу данных

4.5.1 Введение и постановка задачи

Кластерный анализ (кластеризация) — процесс суммаризации неразмеченных данных, основанный на выделении наиболее репрезентативных объектов выборки с помощью меры схожести между объектами. С точки зрения машинного обучения кластеризация — это типичный пример задачи обучения без учителя (unsupervised learning). Как правило, алгоритмы кластеризации оптимизируют заданный критерий с помощью подбора параметров, описывающих кластеры (таких как расположения, форма и размер кластера), с последующим отнесением каждого из объектов к определенному кластеру. В общем случае задача кластеризации в вычислительном плане крайне сложна и прямая оптимизация целевой функции невозможна. Поэтому в большинстве алгоритмов используются приближенные методы оптимизации, которые часто ведут к субоптимальному решению. Кроме того, задача кластеризации важна по причине того, что огромное количество накопленных данных являются неразмеченными и классические алгоритмы обучения с учителем к ним неприменимы. В этом случае обучение без учителя позволит извлечь полезную информацию из данных.

Существует два подхода к определению центров кластеров. Во-первых, в качестве центра может выбираться фактический центр кластера в пространстве признаков. Наиболее известный алгоритм, работающий по этом принципу — k -means [242]. Во-вторых, в качестве центра может выбираться наиболее репрезентативный объект, характеризующий кластер в целом (алгоритмы k -medoids [243], Affinity Propagation [244]). Такие объекты называются представителями кластера (exemplars).

Таким образом, задача кластеризации, основанной на выделении представителей кластера, состоит в том, чтобы определить K наиболее репрезентативных объектов и для каждого объекта назначить соответствующего ему представителя. В качестве входных данных используется множество попарных схожестей между объектами $s(i, k)$ и желаемое количество кластеров (либо вес каждого из объектов, характеризующий его как потенциального представителя).

Данная работа построена следующим образом. В пункте 4.5.2 приведен краткий аналитический обзор существующих алгоритмов кластеризации. В пункте 4.5.3 детально описывается алгоритм аффинного распространения (Affinity Propagation) и наиболее популярные методы выбора параметров этого алгоритма. В пункте 4.5.4 с помощью сгенерированных синтетических данных и численного моделирования алгоритм АР сравнивается с классическими методами кластеризации. В пункте 4.5.5 приводятся основные результаты работы.

4.5.2 Краткий аналитический обзор алгоритмов кластеризации

В настоящее время разработано большое количество алгоритмов для кластерного анализа. Наиболее популярный из них — k -means, который с помощью «жадного» подхода минимизирует критерий квадратичной ошибки:

$$W(S, c) = \sum_{k=1}^K \sum_{i \in S_k} \sum_{v=i}^V (y_{iv} - c_{kv})^2, \quad (4.6)$$

где $S = \{S_1, \dots, S_K\}$ — разбиение множества объектов на непересекающиеся кластеры, $c_k = \{c_{k1}, \dots, c_{kV}\}$, $k = 1, \dots, K$ — множество центров кластеров, V — размерность пространства признаков. Преимущества алгоритма: простота и быстрая сходимость. Недостатки алгоритма: необходимость указания желаемого числа кластеров и начального положения центров кластеров. На практике кластеризация выполняется большое количество раз со случайной инициализацией центров и в качестве результата используется вариант с наименьшим значением критерия.

Другой популярный метод кластеризации базируется на иерархическом принципе и известен в литературе как аггломерация Уорда (Ward Agglomeration). На первом шаге того метода каждый объект рассматривается как отдельный кластер. Далее происходит итеративное попарное объединение кластеров исходя из принципа минимизации расстояния Уорда между ними. Расстояние Уорда определяется как изменение значения критерия (4.6) при объединении двух кластеров. Это расстояние вычисляется по формуле:

$$\Delta(k, l) = \frac{N_k N_l}{N_k + N_l} \|c_k - c_l\|^2,$$

где N_k, N_l — число элементов в объединяемых кластерах k и l , c_k, c_l — центры кластеров. Процесс останавливается, когда достигнуто желаемое число кластеров. К преимуществам алгоритма относится независимость от начальных данных. Недостатки: низкая вычислительная эффективность, необходимость априорных знаний о числе кластеров.

Другой метод нахождения кластеров базируется на поочередном извлечении из данных «аномальных» кластеров с последующим запуском алгоритма k -means с инициализацией в найденных центрах. Этот метод получил название ik -means (intelligent k -means) [245]. Преимущество его в том, что он не требует априорных знаний о числе кластеров и результат его работы детерминирован.

Еще одним методом иерархической кластеризации является метод dePDDP (Principal Direction Divisive Projection) [246]. Он основан на делении кластеров с помощью оценивания плотности распределения объектов в проекции на главную компоненту. Кластер разделяется на две части, если плотность распределения объектов имеет локальный минимум. Процедура останавливается, когда локальных минимумов плотности распределения ни в одном кластере не остается.

Алгоритм аффинного распространения (Affinity Propagation [244]) — это алгоритм кластеризации, основанный на выделении наиболее репрезентативных представителей кластера. Эксперименты показали, что методы, основанные на AP, могут приводить к более хорошим результатам, чем классические алгоритмы (такие как k -means, иерархическая кластеризация, k -medoids). AP был успешно использован в различных областях, в частности для распознавания лиц, сегментации и категоризации изображений, майнинга текстов и т.д. [244, 247, 248, 249, 250, 251]. Оригинальный алгоритм AP в деталях будет рассмотрен в следующем пункте.

4.5.3 Метод аффинного распространения

4.5.3.1 Общие сведения о методе

Метод аффинного распространения (Affinity Propagation, AP) был предложен в [244]. В этом методе данные рассматриваются как полный взвешенный граф, где вершины графа — это объекты выборки, а ребра характеризуются весами $s(i, k)$, которые характеризуют степень схожести между объектами. Веса могут быть как симметричными, так и несимметричными, в том числе не удовлетворяющими неравенству треугольника. В результате работы алгоритма определяется набор объектов, каждый из которых является наиболее репрезентативным представителем своего кластера. Количество этих объектов не задано заранее и определяется

в процессе итеративной передачи сообщений по ребрам графа. Целью работы алгоритма является максимизация суммы схожестей между объектами и их представителями.

Алгоритм принимает на вход набор попарных значений схожести между объектами (similarity), где $s(i, k)$ определяет, насколько хорошо объект с индексом k подходит на роль представителя для объекта с индексом i . Кроме того, для каждого объекта задается величина $s(k, k)$ (preference), характеризующая априорные представления о том, насколько объект подходит на роль представителя. Чем больше значение preference, тем больше вероятность того, что объект будет избран представителем кластера. В оригинальной работе предлагается использовать в качестве стандартного значения preference медиану всех значений $s(i, k)$.

В процессе работы алгоритма используется два типа сообщений.

1. Responsibility $r(i, k)$ пересылается от объекта i к объекту k и отражает, насколько объект k подходит на роль представителя для объекта i , учитывая других потенциальных представителей для объекта i . Правило передачи сообщений этого типа:

$$r(i, k) \leftarrow s(i, k) - \max_{j: j \neq k} \{a(i, j) + s(i, j)\}. \quad (4.7)$$

2. Availability $a(i, k)$ пересылается от потенциального представителя k к объекту i и отражает, насколько для объекта i будет выгодно выбрать объект k в качестве представителя, учитывая то, насколько хорошо k подходит в качестве представителя другим объектам. Правило передачи сообщений этого типа:

$$a(i, k) \leftarrow \begin{cases} \min \{0, r(k, k) + \sum_{j: j \notin \{i, k\}} \max\{0, r(j, k)\}\}, & i \neq k, \\ \sum_{j: j \neq k} \max\{0, r(j, k)\}, & i = k. \end{cases} \quad (4.8)$$

Для определения представителей используются накопленные значения $r(i, k)$ и $a(i, k)$. Пусть индекс k максимизирует сумму $r(i, k) + a(i, k)$. Тогда, если $k = i$, то объект i — это представитель одного из кластеров. В противном случае, если $k \neq i$, то объект k является представителем для

объекта i . Количество представителей определяет число кластеров, полученных в результате работы алгоритма.

В процессе работы алгоритма все значения $a(i, k)$ инициализируются нулями, после чего поочередно в цикле вычисляются новые значения $r(i, k)$ и $a(i, k)$ и представители для каждого из объектов. Алгоритм завершается, когда представители не меняются на протяжении заданного числа итераций (в оригинальной работе использовалось 10 итераций), либо при достижении заданного максимального числа итераций. Кроме того, для предотвращения осцилляций обновленные сообщения демпфируются с параметром λ , то есть в качестве нового значения используется сумма старого сообщения с весом λ и нового сообщения с весом $1 - \lambda$. Алгоритм метода Afinity Propagation следующий:

Вход: Матрица схожестей размера $N \times N$, максимальное количество итераций T , $\lambda \in [0, 1]$, $n = 10$.

Выход: Вектор размера N , содержащий индексы представителей для каждого из объектов.

$a[i, k] \leftarrow 0$

for $t = 1$ to T **do**

$\rho[i, k] \leftarrow s[i, k] - \max_{j: j \neq k} \{a[i, j] + s[i, j]\}$

$r[i, k] \leftarrow \lambda r[i, k] + (1 - \lambda) \rho[i, k]$

$\alpha[i, k] \leftarrow \begin{cases} \min \{0, r[k, k] + \sum_{j: j \notin \{i, k\}} \max\{0, r[j, k]\}\}, & i \neq k, \\ \sum_{j: j \neq k} \max\{0, r[j, k]\}, & i = k. \end{cases}$

$a[i, k] \leftarrow \lambda a[i, k] + (1 - \lambda) \alpha[i, k]$

$c[i] \leftarrow \operatorname{argmax}_{j=1, \dots, N} \{r[i, j] + a[i, j]\}$

if $c[i], i = 1, \dots, N$ не изменялись на протяжении n итераций **then**

return c

return c

4.5.3.2 Вывод алгоритма AP

Вывод метода аффинного распространения основан на представлении целевой функции как фактор-графа и применении к полученному графу алгоритма распространения доверия (belief propagation), так же известного как алгоритм max-product [252, 253]. Несколько разных вариантов вывода со всеми математическими преобразованиями представлены в [244, 247, 254].

Пусть $\{c_{ij}\}_{i,j=1}^N$ — набор из N^2 бинарных переменных. Если $c_{ij} = 1$, то объект j является представителем для объекта i . Введем несколько ограничений на значения этих переменных.

Во-первых, для каждого объекта должен существовать лишь один представитель, поэтому $\sum_{j=1}^N c_{ij} = 1$. Введем штрафную функцию:

$$I_i(c_{i1}, \dots, c_{iN}) = \begin{cases} -\infty, & \text{если } \sum_{j=1}^N c_{ij} \neq 1 \\ 0, & \text{в противном случае} \end{cases}$$

Во-вторых, если объект i выбирает в качестве представителя объект j , то j сам должен быть своим представителем. Штрафная функция примет вид:

$$E_j(c_{1j}, \dots, c_{Nj}) = \begin{cases} -\infty, & c_{jj} = 0 \text{ и } \exists i \neq j : c_{ij} = 1, \\ 0, & \text{в противном случае} \end{cases}$$

Функция S_{ij} определяет величину схожести между объектами и их потенциальными представителями:

$$S_{ij}(c_{ij}) = \begin{cases} s(i, j), & \text{если } c_{ij} = 1, \\ 0, & \text{в противном случае} \end{cases}$$

Тогда целевая функция и соответствующая оптимизационная задача записываются следующим образом:

$$F(c_{11}, \dots, c_{NN}) = \sum_{i,j} S_{ij}(c_{ij}) + \sum_i I_i(c_{i1}, \dots, c_{iN}) + \sum_j E_j(c_{1j}, \dots, c_{Nj}), \quad (4.10)$$

$$\min_{c_{ij} \in \{0,1\}} F(c_{11}, \dots, c_{NN}).$$

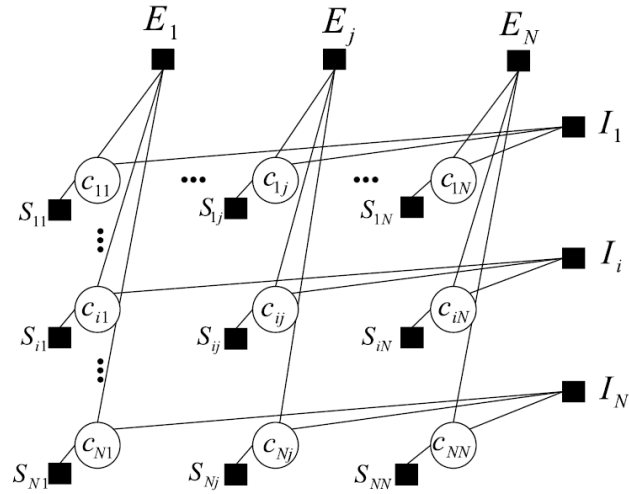


Рисунок 4.17 – Фактор-граф функции (4.10)

Представление этой функции в виде фактор-графа представлено на рисунке 4.17. Фактор-графом [252] называется графическое представление функции, которая может быть представлена в виде произведения (суммы) более простых функций. Вершинами графа являются простые функции и переменные, от которых они зависят. Между переменной и функцией есть ребро, если функция зависит от этой переменной. Таким образом, фактор-граф — это двудольный граф. С помощью такого представления функции можно эффективно подсчитывать маргинальные значения глобальной функции, в частности ее максимум или сумму по всем переменным, кроме одной. Соответствующие алгоритмы называются max-product и sum-product соответственно. Доказано [253], что при отсутствии циклов в графе эти алгоритмы сходятся к истинным маргинальным значениям функции за конечное число шагов. Если же в графе есть циклы (как в случае фактор-графа AP), то аналогичные алгоритмы известны под именем loopy belief propagation и их сходимость не гарантирована. Несмотря на это, некоторые задачи, такие как расшифровка турбокодов, решаются этими алгоритмами крайне эффективно [253]. Кроме того, алгоритм loopy belief propagation может интерпретироваться как минимизация методом градиентного спуска [255] и его сходимость может зависеть от выбранного шага. Поэтому

проблема расходимости алгоритма может быть решена с помощью демпфирования передаваемых сообщений.

Таким образом, суть алгоритма AP есть вычисление маргинальных функций

$$M(c_{ij}) = \max_{c_{kl}: k \neq i, l \neq j} F(c_{11}, \dots, c_{NN})$$

с последующим выбором значения c_{ij} , которое максимизирует значение этой маргинальной функции:

$$c_{ij}^* = \arg \max_{c_{ij} \in \{0,1\}} M(c_{ij}), \quad i, j = 1, \dots, N.$$

Полученные значения $\{c_{ij}^*\}$ максимизируют глобальную функцию $F(c_{11}, \dots, c_{NN})$.

Для вывода формул для передачи сообщений используется алгоритм max-sum [243]. Правила передачи сообщений записываются следующим образом:

$$\begin{aligned} \mu_{x \rightarrow f}(x) &= \sum_{\{l: f_l \in \text{ne}(x) \setminus f\}} \mu_{f_l \rightarrow x}(x), \\ \mu_{f \rightarrow x}(x) &= \max_{x_1, \dots, x_m} \{f(x_1, \dots, x_m) + \sum_{\{m: x_m \in \text{ne}(f) \setminus x\}} \mu_{x_m \rightarrow f}(x)\}. \end{aligned} \quad (4.11)$$

Здесь $\mu_{x \rightarrow f}(x)$ — сообщение от переменной x к функции f , $\mu_{f \rightarrow x}(x)$ — сообщение от функции f к переменной x , $\text{ne}(x) \setminus f$ — множество соседей переменной x за исключением функции f , $\text{ne}(f) \setminus x$ — множество соседей функции f за исключением переменной x соответственно.

Применяя правила (4.11) к фактор-графу AP, после нетривиальных математических преобразований можно получить правила передачи сообщений (4.7) и (4.8) (см. [254]).

4.5.3.3 Выбор параметров алгоритма AP

Несмотря на свою успешность, у алгоритма AP есть существенные недостатки.

Во-первых, необходимо определить меру схожести объектов $s(i, k)$. Авторы предлагают использовать в качестве меры схожести квадрат евклидова расстояния с отрицательным знаком, т.е. $s(i, k) = -\|x_i - x_k\|^2$. Подобное решение дает хорошие результаты на данных с простой структурой (к примеру, кластеры имеют нормальное распределение), но не справляется, когда кластеры имеют сложную нелинейную структуру. Примеры неверной кластеризации на нелинейных данных см. в [256].

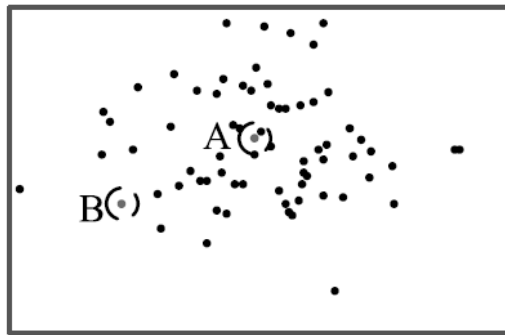


Рисунок 4.18 – Данные с нормальным распределением

Во-вторых, результаты кластеризации и количество кластеров существенно зависит от параметра preference $p(k) = s(k, k)$. Авторы рекомендуют использовать для всех объектов одинаковое значение, лежащее между максимальным и минимальным значениям $s(i, k)$. В своих экспериментах они использовали значение, равное медиане всех схожестей $s(i, k)$. Однако такой выбор ничем не обоснован и в некоторых случаях приводит к неверным результатам кластеризации. К тому же, при наличии априорных знаний о количестве кластеров невозможно напрямую включить эти знания в алгоритм. Единственный способ получить желаемое число кластеров — запускать алгоритм для нескольких значений $p(k)$ и затем выбирать то, которое приводит к нужному количеству кластеров. Кроме того, предположение, что все объекты могут с одинаковой вероятностью стать представителями кластера, неверно. К примеру, на рисунке 4.18 объект, находящийся в точке А подходит на роль представителя больше, чем объект находящийся в точке В.

Таким образом, перед исследователями встает задача поиска оптимальных параметров для метода AP. В литературе представлены разные способы выбора параметров, перечислим несколько из них. Кроме того, предложим несколько новых методов выбора параметров preference и similarity.

1. Евклидова мера схожести: $s(i, k) = -\|x_i - x_k\|^2$. Варианты выбора параметра preference:

– $p(k) = \text{median}\{s(i, k)\}$ — предложено авторами;

– $p(k) = \text{mean}\{s(i, k)\}$;

– на основе плотности распределения объектов в пространстве [256]

$$p(k) = \tau \frac{\sum_{i \neq k} s(i, k) [s(i, k) > \varepsilon]}{\sum_{i \neq k} [s(i, k) > \varepsilon]}, \quad (4.12)$$

где ε — порог, $\min s(i, k) \leq \varepsilon \leq \max s(i, k)$, $\tau \geq 0$ — коэффициент; авторами [257] использовались значения $\tau = 100$, $\varepsilon = 0.5 \cdot \text{median}\{s(i, k)\}$;

– на основе стабильности порождаемого разбиения — Stability-based Affinity Propagation (SAP) [258];

– динамическое изменение $p(k)$ в процессе обмена сообщениями — Adjustable preference Affinity Propagation (APAP) [256];

– на основе центров, порождаемых методом аномальных кластеров [258]

$$p(k) = -\frac{1}{\min_c \|x_k - c\| + \varepsilon}, \quad (4.13)$$

где c — множество центров аномальных кластеров, $\varepsilon \geq 0$ — параметр. В данной работе будем использовать значение $\varepsilon = 0.01$.

2. Мера схожести, основанная на скалярном произведении

$$s(i, k) = \langle y_i, y_k \rangle,$$

где $Y = (y_i)$ — центрированная и стандартизованная матрица признаков.

В этом случае возникает естественное определение параметров $p(k) = \|y_k\|^2$. Такое определение отражает тот факт, что наиболее подходящие

кандидаты в представители должны находиться как можно дальше от центра всех данных. Этот подход, предложенный Б. Миркиным, называется методом аномальных кластеров [259].

Таким образом, проблема выбора *preference* в данном случае элегантно решается. Несмотря на это, есть возможность варьировать разные варианты стандартизации матрицы наблюдений и использовать эвристики для определения $p(k)$, предложенные в предыдущем пункте.

4.5.4 Описание экспериментов

4.5.4.1 Генератор данных

Для экспериментов используем синтетические данные. В качестве генератора данных используем генератор, предложенный недавно Миркиным и Ковалевой [260]. В качестве входных данных генератор требует указать общее количество объектов, количество кластеров, минимальное количество объектов в кластере, размерность пространства и параметр α , влияющий на степень смешивания кластеров. Псевдокод генератора данных представлен алгоритмом 6. Примеры данных, полученных при помощи генератора, приведены на рисунке 4.19. Видно, что меньшим значениям α соответствуют большая степень перемешивания между кластерами. При $\alpha = 0.25$ кластеры становятся практически неразличимы в проекции на две главные компоненты. При $\alpha = 0.5$ и 0.75 кластеры становятся более различимыми.

Выберем для эксперимента следующие параметры генерации данных:

- количество кластеров $K = [7, 15, 25]$;
- размерность пространства $V = [15, 30]$;
- количество объектов $N = 1000$;
- минимальное количество объектов в кластере $n_{min} = 25$;
- параметр смешивания $\alpha = [0.25, 0.50, 0.75]$.

Всего выбрано 18 различных комбинации параметров. Для каждой комбинации сгенерируем $R = 10$ различных наборов данных, по которым будем усреднять результаты численного эксперимента.

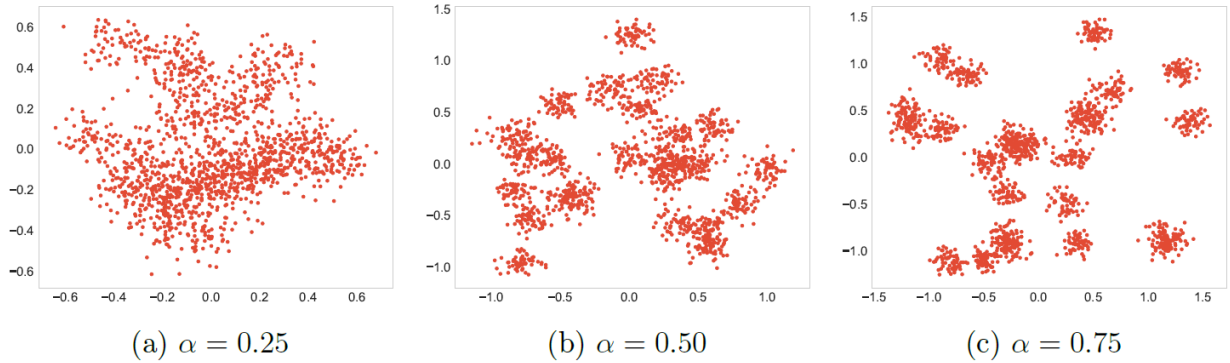


Рисунок 4.19 – Проекция датасета из $N = 1500$ объектов, $K = 25$ кластеров, $V = 25$ на две главные компоненты SVD для различных значений параметра α

Алгоритм 6: Генератор данных Миркина и Ковалевой

Вход: Число объектов N , число кластеров K , мин. численность кластера n_{min} , размерность V , параметр α .

Выход: Матрица объектов размера $N \times V$ и вектор меток длины N .

$n \leftarrow$ Массив из $K - 1$ случайных чисел, р.р. на $[0,1]$, в возрастающем порядке.

$n \leftarrow [0, n, 1]$ дополнить нулем и единицей

$d \leftarrow$ Разности между соседними числами в n .

for $i = 1, \dots, K$ **do**

$n[i] \leftarrow n_{min} + \text{round}\{(N - Kn_{min})d[i]\}$ - размер i -го кластера

$d_1, d_2 \leftarrow -1, 1$

$C, L \leftarrow [], []$

for $i = 1, \dots, K$ **do**

$\mu \leftarrow$ Случайный вектор размерности V , р.р. в кубе $\alpha \cdot [d_1, d_2]^V$.

$\Sigma \leftarrow$ Случайный вектор размерности V , р.р. в кубе $(d_2 - d_1) \cdot [0.025, 0.05]^V$.

$\Sigma \leftarrow \text{diag}(\Sigma^2)$ - диагональная матрица

$C \leftarrow [C, \mathcal{N}(\mu, \Sigma) \times n[i]]$ - сгенерировать $n[i]$ новых объектов

$L \leftarrow [L, i \times n[i]]$ - добавить метки для $n[i]$ объектов

Return C, L

4.5.4.2 Метрики качества кластеризации

Для оценки качества работы алгоритмов необходимо задать метрики качества. В качестве первой метрики используем разницу между истинным количеством кластеров K и значением K^* , полученным в результате работы алгоритма.

Во-вторых, для оценки качества полученного разбиения по сравнению с истинным разбиением, используем метрику Adjusted Rand Index (ARI). Метрика ARI вычисляется по формуле

$$ARI = \frac{\sum_{i,j} \binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}{\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] / 2 - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}},$$

где $n_{ij} = |S_i \cap S_j|$, $a_i = \sum_{j=1}^K |S_i \cap S_j|$, $b_j = \sum_{i=1}^K |S_i \cap S_j|$. Максимальное значение ARI равно 1. ARI обращается в ноль, если одно из разбиений состоит из одного кластера, содержащего все объекты выборки.

В-третьих, важной характеристикой алгоритма является время его работы t . Зачастую на практике не используются алгоритмы с наилучшим качеством по причине того, что они вычислительно неэффективны.

4.5.4.3 Алгоритмы, участвующие в эксперименте

В численных экспериментах будем сравнивать результаты кластеризации методом аффинного распространения при использовании различных параметров preference и similarity. Кроме того, для полноты эксперимента проведем вычисления с помощью классических алгоритмов k -means, ik -means, dePDDP. Наилучшее количество кластеров в методе k -means

может быть определено с помощью различных эвристических методов (см. [261]). Для определения числа кластеров в данном эксперименте будем использовать метод максимизации силуэта (silhouette).

Силуэт (silhouette) позволяет оценить качество кластеризации, используя только саму выборку и результат кластеризации. Сначала силуэт определяется отдельно для каждого объекта. Обозначим через $a(i)$ — среднее расстояние от объекта i до объектов из того же кластера, через $b(i)$ — среднее расстояние от объекта i до объектов из ближайшего кластера (отличного от того, в котором лежит сам объект). Тогда силуэтом данного объекта называется величина:

$$\text{Silhouette}(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}.$$

Силуэтом выборки называется средняя величина силуэта объектов данной выборки. Таким образом, силуэт показывает, насколько среднее расстояние до объектов своего кластера отличается от среднего расстояния до объектов других кластеров. Данная величина лежит в диапазоне $[-1, 1]$. Значения, близкие к -1 , соответствуют разрозненным кластеризациям, значения, близкие к нулю, говорят о том, что кластеры пересекаются и накладываются друг на друга, значения, близкие к 1 , соответствуют четко выделенным кластерам. Таким образом, чем больше силуэт, тем более четко выделены кластеры, и они представляют собой компактные, плотно сгруппированные облака точек.

Оптимальному числу кластеров k , соответствующих наилучшему разбиению, соответствует максимальное значение силуэта. Назовем такой алгоритм S - k -means.

Список всех алгоритмов, использованных в работе, приведен в Таблице 4.8.

Таблица 4.8 – Алгоритмы, участвующие в эксперименте и их параметры

№	Алгоритм	$s(i, k)$	$s(k, k)$	Обозначение
1	S- k -means			
2	ik -means			
3	dePDDP			
4	AP	$-\ x_i - x_k\ ^2$	$\text{median}\{s(i, k)\}$	APE median
5			density-based (4.12)	APE dens.
6			anomalous (4.13)	APE anomal.
7		$\langle y_i, y_k \rangle$	$\ y_k\ ^2$	APP
8			$\text{median}\{s(i, k)\}$	APP median
9			density-based (4.12)	APP dens.
10			anomalous (4.13)	APP anomal.

4.5.4.4 Результаты

В Таблицах 4.9 и 4.10 представлены результаты кластеризации для выбранных ранее параметров. Для каждого набора параметров было сгенерировано 10 наборов данных, на всех наборах были запущены десять перечисленных алгоритмов, рассчитаны средние значения и стандартные отклонения метрик, а так же среднее время работы алгоритмов. Полужирным шрифтом отмечены алгоритмы, которые дают самое большое значение ARI для заданных K, α .

Основные выводы:

- алгоритм APP, несмотря на то, что матрица схожести и параметр preference в нем вводится естественным образом на основе скалярных произведений, показывает самое низкое качество кластеризации среди всех алгоритмов. Количество определяемых кластеров оказывается намного больше истинного, причем чем меньше параметр смешивания, тем больше K ; достигаемые значения ARI так же неудовлетворительны.
- при большой степени смешивания кластеров (малые значения α) наилучшим образом показали себя алгоритмы S- k -means, APE anomal., APP anomal.; остальные алгоритмы, основанные на AP, показывают низкое качество при большом смешивании; причем при увеличении размерности качество, достигаемое алгоритмами на основе AP, деградирует, в то время как качество остальных алгоритмов возрастает;

- с точки зрения восстановления истинного числа кластеров наилучшим образом себя показывают алгоритмы *S-k-means*, *APE anomal.* и *APP anomal.*; *AP* с константными значениями *preference* правильно восстанавливают число кластеров только при низкой степени смешивания кластеров;
- алгоритмы, основанные на *AP* с *preference* на основе плотности распределения объектов так же показывают неудовлетворительное качество кластеризации;
- алгоритм *APP median* показывает лучший результат по сравнению с *APE median*, что подтверждает обоснованность выбора метрики схожести в форме скалярного произведения;
- алгоритмы *APP anomal.* и *APE anomal.* показывают сравнимые результаты, поэтому вывод о превосходстве одной метрики схожести над другой в данном случае сделать нельзя;
- наилучший результат среди всех алгоритмов дает *S-k-means*; во всех случаях он правильно восстанавливает количество кластеров, а его значения *ARI* превышают 0.9.

Таблица 4.9 – Результаты кластеризации с помощью десяти алгоритмов для 1500 объектов с $V = 15$, $K = 7, 15, 25$ кластеров и $\alpha = 0.25, 0.50, 0.75$. Каждая запись в таблице — среднее значение величины и ее стандартное отклонение по 10 наборам данных. t — среднее время работы алгоритма

K, α	S- k -means ($t = 3.99$)		ik -means ($t = 0.02$)		dePDDP ($t = 0.32$)		APE median ($t = 9.10$)	
	K^*	ARI	K^*	ARI	K^*	ARI	K^*	ARI
7, 0.25	6.5/0.8	0.967/0.074	7.2/1.6	0.843/0.209	9.6/1.2	0.946/0.110	24.8/3.5	0.289/0.067
7, 0.50	6.8/0.4	0.987/0.036	6.6/0.5	0.969/0.066	7.5/0.8	0.995/0.014	9.8/0.9	0.686/0.090
7, 0.75	6.9/0.3	1.000/0.001	6.5/0.5	0.981/0.025	7.2/0.4	0.999/0.001	7.0/0.0	1.000/0.000
15, 0.25	14.4/0.8	0.981/0.026	10.9/1.5	0.812/0.080	18.9/5.9	0.807/0.252	29.6/1.3	0.588/0.050
15, 0.50	15.0/0.0	1.000/0.000	10.9/1.7	0.793/0.109	17.6/1.3	0.981/0.011	15.5/0.7	0.955/0.059
15, 0.75	15.0/0.0	1.000/0.000	12.7/1.4	0.896/0.061	15.8/0.7	0.993/0.008	15.0/0.0	1.000/0.000
25, 0.25	24.8/0.4	0.992/0.014	13.4/0.9	0.606/0.041	19.1/14.9	0.338/0.347	34.8/1.6	0.813/0.028
25, 0.50	24.9/0.3	0.999/0.004	15.0/2.1	0.639/0.084	27.5/9.0	0.872/0.291	25.0/0.0	1.000/0.000
25, 0.75	25.0/0.0	1.000/0.000	19.2/3.0	0.784/0.108	28.6/2.1	0.975/0.017	25.0/0.0	1.000/0.000
K, α	APE dens. ($t = 8.99$)		APE anomal. ($t = 9.52$)		APP ($t = 0.92$)		APP median ($t = 1.39$)	
	K^*	ARI	K^*	ARI	K^*	ARI	K^*	ARI
7, 0.25	6.2/0.9	0.957/0.075	7.9/0.7	0.852/0.133	261.2/61.6	0.196/0.119	18.6/3.5	0.469/0.170
7, 0.50	6.8/0.4	0.985/0.036	7.9/0.8	0.845/0.139	71.2/11.0	0.446/0.139	9.7/1.3	0.718/0.164
7, 0.75	6.7/0.5	0.992/0.019	8.1/0.9	0.918/0.102	39.6/7.6	0.610/0.149	8.1/1.0	0.866/0.119
15, 0.25	9.1/1.1	0.710/0.069	15.1/0.3	0.988/0.010	300.9/37.7	0.281/0.087	23.9/1.8	0.708/0.095
15, 0.50	14.4/0.7	0.975/0.039	15.8/0.6	0.964/0.040	100.2/14.5	0.605/0.076	15.8/0.4	0.951/0.037
15, 0.75	14.3/0.6	0.979/0.021	16.5/1.1	0.961/0.046	60.7/9.6	0.733/0.089	15.1/0.3	0.989/0.024
25, 0.25	8.1/1.5	0.307/0.083	24.7/0.5	0.976/0.021	349.1/21.2	0.326/0.073	29.7/2.0	0.857/0.064
25, 0.50	20.7/1.7	0.843/0.064	25.6/0.8	0.992/0.012	137.1/9.3	0.677/0.055	25.1/0.3	0.995/0.010
25, 0.75	21.3/1.2	0.863/0.037	25.9/1.2	0.990/0.019	85.0/8.9	0.790/0.040	24.9/0.3	0.998/0.005
K, α	APP dens. ($t = 1.42$)		APP anomal. ($t = 1.30$)					
	K^*	ARI	K^*	ARI				
7, 0.25	24.4/5.0	0.399/0.155	8.9/0.8	0.725/0.157				
7, 0.50	12.8/2.2	0.619/0.166	7.5/0.5	0.896/0.109				
7, 0.75	10.1/1.5	0.760/0.132	7.1/0.3	0.986/0.043				
15, 0.25	28.6/2.8	0.653/0.108	15.2/0.9	0.927/0.058				
15, 0.50	16.9/1.0	0.912/0.051	15.0/0.0	1.000/0.001				
15, 0.75	15.4/0.7	0.962/0.060	14.8/0.4	0.995/0.011				
25, 0.25	35.2/3.0	0.802/0.070	23.4/0.8	0.893/0.035				
25, 0.50	25.4/0.7	0.984/0.031	24.8/0.4	0.994/0.011				
25, 0.75	24.9/0.3	0.998/0.005	24.7/0.6	0.991/0.018				

Таблица 4.10 – Результаты кластеризации с помощью десяти алгоритмов для 1500 объектов с $V = 30$, $K = 7, 15, 25$ кластеров и $\alpha = 0.25, 0.50, 0.75$. Каждая запись в таблице — среднее значение величины и ее стандартное отклонение по 10 наборам данных. t — среднее время работы алгоритма

K, α	S- k -means ($t = 3.60$)		ik -means ($t = 0.01$)		dePDDP ($t = 0.40$)		APE median ($t = 8.97$)	
	K^*	ARI	K^*	ARI	K^*	ARI	K^*	ARI
7, 0.25	6.9/0.3	0.995/0.014	6.7/0.5	0.985/0.027	8.6/1.2	0.984/0.014	17.9/0.8	0.426/0.051
7, 0.50	7.0/0.0	1.000/0.000	7.0/0.0	1.000/0.000	7.4/0.7	0.998/0.004	8.7/0.6	0.751/0.076
7, 0.75	7.0/0.0	1.000/0.000	6.9/0.3	0.999/0.004	7.3/0.6	0.995/0.013	7.1/0.3	0.978/0.067
15, 0.25	14.9/0.3	0.999/0.004	11.8/1.4	0.909/0.036	17.3/5.3	0.881/0.259	24.3/1.9	0.650/0.080
15, 0.50	15.0/0.0	1.000/0.000	14.6/0.7	0.987/0.023	16.7/1.6	0.987/0.009	15.0/0.0	1.000/0.000
15, 0.75	15.0/0.0	1.000/0.000	14.5/0.8	0.986/0.024	16.4/1.3	0.993/0.009	15.0/0.0	1.000/0.000
25, 0.25	24.6/0.5	0.990/0.013	18.2/2.1	0.751/0.074	21.3/16.4	0.481/0.436	28.5/1.2	0.911/0.019
25, 0.50	25.0/0.0	1.000/0.000	22.8/1.4	0.926/0.049	26.0/6.2	0.887/0.265	25.0/0.0	1.000/0.000
25, 0.75	25.0/0.0	1.000/0.000	24.1/0.5	0.973/0.017	28.6/1.7	0.975/0.011	25.0/0.0	1.000/0.000
K, α	APE dens. ($t = 8.87$)		APE anomal. ($t = 9.43$)		APP ($t = 0.57$)		APP median ($t = 1.19$)	
	K^*	ARI	K^*	ARI	K^*	ARI	K^*	ARI
7, 0.25	6.8/0.4	0.991/0.018	11.0/1.0	0.650/0.058	546.6/41.8	0.086/0.026	16.2/1.7	0.503/0.076
7, 0.50	7.0/0.0	1.000/0.000	11.9/0.5	0.552/0.061	174.1/40.7	0.416/0.188	8.5/1.2	0.821/0.163
7, 0.75	7.0/0.0	1.000/0.000	11.5/1.1	0.622/0.047	69.4/20.3	0.631/0.156	7.6/0.7	0.898/0.121
15, 0.25	12.2/1.5	0.899/0.064	17.4/1.4	0.857/0.096	598.8/40.9	0.123/0.054	22.0/2.0	0.735/0.095
15, 0.50	14.9/0.3	0.997/0.009	16.5/1.0	0.935/0.059	196.7/28.9	0.516/0.091	15.3/0.6	0.974/0.053
15, 0.75	14.9/0.3	0.998/0.005	17.1/2.2	0.888/0.098	99.0/18.1	0.712/0.090	15.0/0.0	1.000/0.000
25, 0.25	13.4/1.6	0.519/0.088	25.9/0.7	0.987/0.011	642.3/35.4	0.143/0.037	27.9/0.8	0.931/0.040
25, 0.50	24.4/0.7	0.983/0.018	27.2/1.4	0.967/0.026	243.6/27.5	0.559/0.050	25.0/0.0	1.000/0.000
25, 0.75	24.3/0.8	0.975/0.034	26.2/1.2	0.991/0.009	136.6/13.1	0.737/0.026	25.0/0.0	1.000/0.000
K, α	APP dens. ($t = 1.46$)		APP anomal. ($t = 1.10$)					
	K^*	ARI	K^*	ARI				
7, 0.25	22.0/2.9	0.415/0.077	10.6/0.7	0.694/0.061				
7, 0.50	12.4/3.4	0.664/0.217	7.5/0.7	0.901/0.133				
7, 0.75	9.1/1.5	0.821/0.135	7.2/0.4	0.948/0.105				
15, 0.25	25.9/3.5	0.671/0.110	15.7/0.8	0.931/0.088				
15, 0.50	15.8/1.0	0.942/0.073	15.0/0.0	1.000/0.000				
15, 0.75	15.2/0.4	0.975/0.056	15.0/0.0	1.000/0.000				
25, 0.25	30.9/1.0	0.884/0.031	25.0/0.0	0.998/0.002				
25, 0.50	25.1/0.3	0.998/0.007	25.0/0.0	1.000/0.000				
25, 0.75	25.0/0.0	1.000/0.000	25.0/0.0	1.000/0.000				

4.5.5 Некоторые выводы

При выполнении работы решены следующие задачи:

- проведен аналитический обзор методов кластеризации;
- приведен обзор метода аффинного распространения (AP), включая его математический вывод и способы задания его параметров;

- разработаны программы для генерации и визуализации данных, все упомянутые алгоритмы реализованы на языке Python 3;
- поставлен численный эксперимент для сравнения производительности алгоритма AP и классических алгоритмов кластеризации;
- проведено краткое обсуждение результатов численного моделирования.

5 Исследование моделей распространения и развития

В этом разделе отчета представлены результаты по исследованию некоторых моделей распространения и развития. Наряду с традиционным подходом динамического моделирования, основанного на построении системы дифференциальных уравнений, в этом разделе будет широко применено и агентное моделирование.

Раздел состоит из двух подразделов:

- исследование моделей распространения;
- исследование моделей специализации в абстрактных системах.

5.1 Исследование моделей распространения

5.1.1 Введение и постановки задач

В настоящее время основным подходом к моделированию процесса распространения веществ в пространстве являются динамические математические модели, базирующиеся на системах дифференциальных уравнений.

Так, например, различные модели используют в своей основе уравнения неразрывности, закон Дарси [262, 263], описывающий ток жидкости через пористую среду, а также различные системы уравнений Навье-Стокса [264].

Однако ни одна из моделей не отвечает всем потенциально возможным ситуациям, поскольку всегда остаётся определенное количество физических параметров, таких как температура, влияющих на процесс, которые остаются неучтёнными. При этом, для реализации моделей и подсчётов результатов необходимы значительные вычислительные ресурсы.

В связи с этим, в литературе встречаются модели, в которых изначально упоминается, что в некоторых ситуациях (например, распространение нефтяных разливов на поверхности моря) скорость вычислений заметно важнее абсолютной точности результата.

Одним из таких подходов является использование сетевой структуры для моделирования среды. И по рёбрам этой сети происходит распростране-

ние вещества с учётом различных правил, которые приближает данный процесс к реальному.

В свою очередь, теория устойчивости разностных схем для модельных задач газовой динамики хорошо представлена в литературе [265, 266, 267]. В основном, рассматриваются конечно-разностные схемы решения одномерных уравнений баротропной газовой динамики. Эти схемы являются явными двухуровневыми по времени и используют симметричный трехточечный шаблон в пространстве. Их конструкция основана на специальной квазигазодинамической и квазигидродинамической [268, 269] регуляризации исходного уравнения (без регуляризации схемы нестабильны). Схемы такого рода успешно применяются в многочисленных практических приложениях, в частности, см. [270, 271], однако их теория до сих пор недостаточно развита.

5.1.2 Основная часть

Как упоминалось в введении, для построения модели быстрого моделирования процесса распространения вещества лучше подходят варианты без использования систем дифференциальных уравнений. В связи с этим для моделирования распространения нефти на поверхности моря было решено использовать имитационную модель.

Также как и в работе [272], была использована сеть, моделирующая море, разделённое на одинаковые по размеру элементарные территории, между которыми будет перемещаться нефтяное пятно. Для характеристики рёбер между соседними участками была сконструирована функция, подобная функции из [273], которая учитывает скорость течения, скорость ветра в конкретном месте, а также температура воды на рассматриваемом участке моря.

После того, как построена сеть происходит динамическое моделирование распространения нефти по следующей схеме (рисунок 5.1).

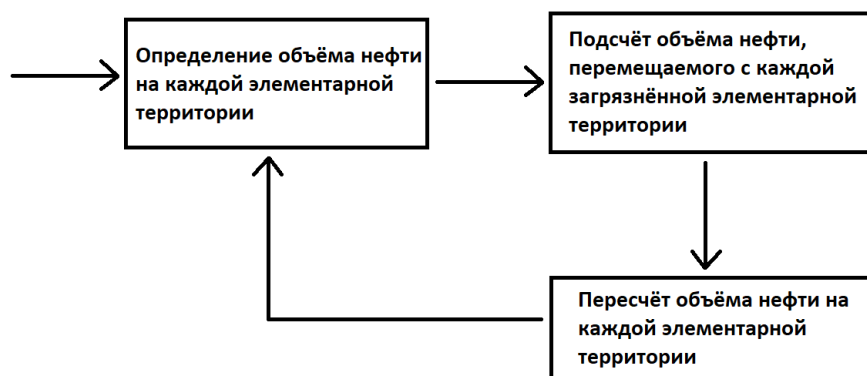


Рисунок 5.1 – Схема работы динамической модели распространения нефти на поверхности моря

Для пересчёта объёма нефти на каждой элементарной территории используется простая функция, которая учитывает все входящие и исходящие потоки для кадого участка моря:

$$Q_{ik} = Q_{(i-1)k} + \sum_{j=1}^N \hat{Q}_{ijk} - \sum_{j=1}^N \hat{Q}_{ikj}, \quad \forall k = 1, \dots, N,$$

где Q_{ik} – объём нефти, находящийся на участке k на i -й итерации, а \hat{Q}_{ikj} – объём нефти, который перешёл с участка k на участок j во время i -й итерации.

В результате был получен аппарат, позволяющий моделировать процесс распространения нефти на поверхности моря [274]. Однако в реальной жизни могут возникнуть различные сценарии разлива: авария на производстве, авария при транспортировке кораблём, прокол нефтепровода. Поэтому для того, чтобы правильно моделировать различные сценарии был реализован дополнительный блок, который может помещать вещество в имитационную модель (сеть элементарных участков моря) нефть разными режимами. Для этого пользователю позволено задавать общий объём разлива, а также скорость, с которой нефть попадает в море.

В рамках исследования динамических моделей процессов распространения на основе анализа систем дифференциальных уравнений была рассмотрена задача одномерной баротропной газовой динамикой:

$$\begin{cases} \partial_t \rho + \partial_x(\rho u) = 0, \\ \partial_t(\rho u) + \partial_x p(\rho) = 0. \end{cases}$$

где $\rho > 0$, u и p – плотность, скорость газа (искомые функции) и давление.

В результате исследований были линеаризованы регуляризованные явные конечно-разностные схемы на постоянном решении для данной задачи [275].

После этого было выведено необходимое условие типа фон Неймана для выполнения слабой консервативности ($\sup_{m \geq 0} \|y^m\|_H \leq \|y^0\|_H \quad \forall y^0 \in H$) [276].

Также был доказан критерий типа CFL для слабой консервативности в L^2 для соответствующей начальной задачи. Слабая консервативность в L^2 означает равномерность временной границы для нормы масштабированного решения вместо закона сохранения энергии для линеаризованной исходной системы, т.е. акустической системы уравнений [277].

Численный эксперимент показывает, что справедливость свойства слабой консервативности важно, поскольку оно предотвращает численные решения от известных возможных мешающих колебаний. Это свойство гарантирует равномерную по времени стабильность L^2 относительно исходных данных.

Поскольку на практике часто используются необходимые условия (выведение достаточных условий в целом гораздо сложнее), важно понимать, в какой степени это законно. В результате подробного анализа было выведено, что критерий значительно превосходит по применимости необходимое и достаточное условия, полученные в частном случае [278]. Более того, именно этот критерий наиболее хорошо соответствует результатам численных экспериментов для исходной газодинамической системы. Поэтому именно критерий является наиболее адекватным и полезным для практических целей.

5.1.3 Некоторые выводы

По итогам работы можно сказать, что поставленные задачи выполнены в полном объёме. Во-первых, в результате практических исследований дина-

мического моделирования без использования систем дифференциальных уравнений был разработан вычислительный комплекс, позволяющий моделировать процесс распространения нефти на поверхности Баренцева моря без необходимости использования серьёзных вычислительных мощностей. Более того, поскольку модель достаточно универсальна – при небольшом изменении параметров данный комплекс может быть применим и в других регионах.

Что касается теоретических исследований, связанных с использованием дифференциальных уравнений, были исследованы явные двухуровневые конечно-разностные схемы решения одномерных уравнений баротропной газовой динамики. Были линеаризованы схемы на постоянном решении и выведено необходимое условие типа фон Неймана и критерий типа CFL для слабой консервативности в L^2 для соответствующей задачи на всей прямой. Выведенный критерий оказался существенно уже необходимого условия и шире достаточного.

5.2 Исследование моделей специализации в абстрактных системах

5.2.1 Введение и постановка задачи моделирования специализации

Абстрактные системы, которые состоят из конечного числа элементов, каждый из которых может выполнять одну или несколько функций повсеместно возникают в различных областях знания, в частности в эволюционной биологии. В качестве подобного рода системы здесь рассматривается колония клеток, элементами системы служат непосредственно клетки. Каждая клетка может выполнять одну или несколько биологических функций, среди которых могут быть упомянуты репродуктивная, соматическая функция, функция фотосинтеза или производства различного рода метаболитов. В результате действия различного рода внутренних физиологических ограничений, клетка не может увеличить свой вклад в выполнение одной из функций, не уменьшив выполнения хотя бы одной другой функции. Таким образом,

возникают так называемые внутренние структурные ограничения. Важный вопрос, который ставится здесь – каким образом и при каких условиях в колонии клеток возникает функциональная специализация. Будем говорить, что клетка специализируется на выполнении какой-либо функции, если уровень выполнения данной функции данной клеткой максимален, а уровень выполнения клеткой прочих функций равен нулю.

Вопрос возникновения специализации очень важен в теории эволюции [279]. Данный вопрос тесно связан с понятием биологической сложности и увеличения этой сложности по мере эволюционирования живых систем. Существует множество различных подходов для определения биологической сложности организма. В частности, биологическая сложность может быть положена равной количеству различных специализированных типов внутри рассматриваемого организма. Таким образом, можно считать, что возникновение клеточной специализации неизбежно приводит к увеличению биологической сложности.

В рамках эволюционной биологии можно выделить два основных типа специализации: репродуктивно-соматическую специализацию и специализацию на различных соматических функциях. Репродуктивно-соматическая специализация предполагает в общих чертах, что рассматривается колония родственных клеток, где каждая клетка может выполнять одну из двух функций – репродуктивную и соматическую. Специализация означает, что происходит выделение отдельной репродуктивной линии клеток или отдельной соматической линии клеток, или этих двух линий одновременно. Специализация на различных соматических функциях может быть рассмотрена в двух различных контекстах. Первый: рассматривается группа клеток, которые существуют вместе, каждая из которых способна размножаться. В результате взаимодействия эти клетки могут обмениваться производимыми метаболитами и, при выполнении некоторых условий, начать специализироваться на производстве тех или иных полезных веществ. Второй: рассматривается колония клеток с уже сформировавшейся и выделившейся репродуктивной ли-

нией. Остальные клетки являются соматическими и могут начать специализироваться на различных не репродуктивных функциях, которые необходимы организму для поддержания жизнедеятельности и обеспечения ресурсами и энергией репродуктивных клеток. В нашей работе мы исследуем вопрос возникновения репродуктивно-соматической специализации.

Существует ряд работ [280, 281, 282, 283, 284], в которых предпринимается попытка теоретического объяснения процесса возникновения репродуктивно-соматической специализации [285, 286, 287, 288, 289]. В данных работах рассматриваются различные факторы, которые могут оказать влияние на возможность возникновения специализации в колониальных организмах. Среди таких факторов называются: размер колонии [284], первоначальные издержки воспроизводства [284], работа различных регуляторных генов [290], различные физические законы [283], которые действуют на конкретную колонию, форма организма. Одним из важнейших факторов, оказывающих влияние не жизнедеятельность колонии, может считаться фактор окружающей среды. Однако моделей, показывающих, каким образом различные условия внешней среды и уровень доступного ресурса влияют на процесс возникновения репродуктивно-соматической специализации, в сущности, не так уж и много.

Мы обобщаем фундаментальные теоретические модели, существующие в данной области знаний и предлагаем новую фитнес-оптимизационную модель, которая позволяет описать влияние факторов окружающей среды на возможности возникновения репродуктивно-соматической специализации в колониях клеток. Кроме того, мы пытаемся качественно оценить это влияние в совокупности с различными иными факторами: размером колонии и позиционными эффектами.

5.2.2 Основная часть

В нашей работе рассматривается колония, состоящая из конечного числа клеток. Индексы клеток задаются следующим образом: $i = 1, \dots, N$. Каждая клетка может выполнять одну из двух функций: репродуктивную и соматическую. Обозначим b_i выполнение репродуктивной функции клеткой i , а v_i – выполнение соматической функции данной клеткой. Введем в рассмотрение векторы $b = (b_1, \dots, b_N)$ и $v = (v_1, \dots, v_N)$.

Для каждой клетки i рассмотрим число $b_i^{\max} \in R$, $0 < b_i^{\max} < \infty$, которое задает максимальный уровень репродуктивной функции, который может быть достигнут клеткой с индексом i , и рассмотрим отображение $\varphi_i : [0, b_i^{\max}] \rightarrow R$. Внутренние структурные ограничения клетки зададим следующим образом:

$$v_i = \varphi_i(b_i), \quad i = 1, \dots, N.$$

Мы предполагаем, что все функции φ_i являются непрерывными на множествах $[0, b_i^{\max}]$, дважды непрерывно-дифференцируемыми на соответствующих интервалах $(0, b_i^{\max})$. Кроме того, если клетка начинает концентрироваться на выполнении одной функции, то выполнение ею другой функции должно уменьшаться. Это свойство может быть формализовано следующим образом:

образом: $\forall i = 1, \dots, N : \frac{d\varphi_i}{db_i} < 0$. Также предположим, что

$\forall i = 1, \dots, N : \varphi_i(b_i^{\max}) = 0$, $\varphi_i(0) = v_i^{\max}$, где $v_i^{\max} \in R$, $0 < v_i^{\max} < \infty$. Данные формулы говорят нам о том, что если клетка максимально концентрируется на выполнении одной функции, то уровень выполнения другой функции должен быть минимально возможным (то есть, должен равняться нулю).

Общий вклад колонии в репродуктивную функцию, B есть аддитивная функция переменных b_i , Соматическая функция колонии, V – есть аддитивная функция переменных v_i .

$$B = \sum_{i=1}^N b_i, \quad V = \sum_{i=1}^N v_i.$$

Аддитивность репродуктивной функции организма можно объяснить тем, что репродуктивная функция клетки может быть вычислена как количество новых колоний, получающихся в результате размножения данной клетки. Аддитивность соматической функции может быть объяснена тем, что в данной работе мы рассматриваем флагеллярное движение в качестве аппроксимации для соматической функции [284].

Общая жизнеспособность организма (или его фитнес-функция) W есть функция от двух неотрицательных аргументов, B и V , удовлетворяющая следующим свойствам: $W(B, V)$ есть неотрицательная функция, такая, что $W(B, V) = 0$ тогда и только тогда, когда либо $B = 0$, либо $V = 0$. Кроме того, функция жизнеспособности колонии должна быть неубывающей функцией по обоим аргументам при всех допустимых значениях этих аргументов [291, 292, 293]. В качестве жизнеспособности колонии в нашей работе мы будем рассматривать следующую мультипликативную функцию:

$$W = B^\alpha V^\beta, \text{ где } \alpha > 0 \text{ и } \beta > 0.$$

Данный вид функции жизнеспособности отражает тот факт, что и репродуктивная и соматическая функции являются жизненно необходимыми для нормального функционирования колонии. Выделим еще и важнейшее синергетическое свойство, которым обладает данная функция. Рассмотрим две клетки. Предположим, что первая клетка имеет относительно большое значение репродуктивной функции и относительно маленькое значение соматической. Напротив, предположим, что вторая клетка имеет относительно большое значение соматической функции и относительно маленькое значение репродуктивной. Каждая клетка в отдельности имеет низкое значение фитнес-функции, но вместе, внутри колонии, клетки могут взаимодействовать друг с другом и благодаря этому достичь большего уровня жизнеспособности [284].

Параметры α и β могут быть интерпретированы как относительные важности репродуктивной и соматической составляющей соответственно для жизнедеятельности организма. Данные коэффициенты также могут быть рассмотрены как некоторые параметры окружающей среды.

Далее мы введем в модель общее ресурсное ограничение. Предположим, что количество ресурса, доступного организму ограничено величиной C . Тогда запишем общее ресурсное ограничение в линейно виде:

$$k_1 B + k_2 V \leq C .$$

Параметр k_1 представляет собой количество ресурса, необходимое колонии для производства единицы репродуктивной функции, а параметр k_2 есть количество ресурса, необходимое для производства единицы соматической функции. Предположим, что $k_1 > 0$, $k_2 > 0$ и $C > 0$.

Таким образом, задача фитнес-оптимизации может быть сформулирована в виде задачи математического программирования:

$$\left\{ \begin{array}{l} W = \left(\sum_{i=1}^N b_i \right)^\alpha \left(\sum_{i=1}^N v_i \right)^\beta \rightarrow \max_{b,v} \\ v_i = \varphi_i(b_i), \quad i = 1, \dots, N, \\ k_1 \sum_{i=1}^N b_i + k_2 \sum_{i=1}^N v_i \leq C, \\ b_i \geq 0, \quad v_i \geq 0, \quad i = 1, \dots, N. \end{array} \right.$$

Наша задача – выяснить, каким образом и при каких условиях возникает специализация в рассматриваемой модели. Будем говорить, что клетка i специализируется на соматической функции тогда и только тогда, когда $v_i = v_i^{max}$, и клетка i специализируется на репродуктивной функции тогда и только тогда, когда $b_i = b_i^{max}$.

5.2.3 Основные результаты

Нами были рассмотрены три различные модификации модели. Прежде чем описать их, определим еще одно понятие. Будем говорить, что две клетки имеют разные типы, если они имеют различные структурные ограничения.

Различные типы клеток могут возникать из-за различных позиционных эффектов и действия различного рода регуляторных генных механизмов. В первой модификации модели мы рассматривали колонии, состоящие из клеток одного типа: мы хотели выявить чистый эффект влияния среды на жизнедеятельность колонии. Во второй модификации мы уже добавляли в модель возможность существования в группе клеток различных типов. В третьей модификации мы обобщали понятие структурных ограничений и рассматривали эти ограничения в форме неравенств. Основные новые результаты нашей модели можно сформулировать следующим образом.

1. В колониях большого размера введение ресурсного ограничения может привести как к неспециализированным состояниям, так и к возникновению частичной специализации. В частности, в условиях среды, достаточно сильно ограниченной в ресурсах, колонии большого размера будут неспециализированными.
2. Введение ресурсного ограничения в модель может привести к возникновению специализации в колониях маленького размера. В частности, были выделены два основных случая возникновения специализации в колониях маленького размера. Первый случай может быть characterized следующим образом: большая важность одной из компонент фитнес-функции в совокупности с достаточно большими издержками производства этой функции. Таким образом, в данной ситуации возникает некоторый компромисс, между большими издержками производства и большой важностью, который заставляет некоторые клетки колонии специализироваться на данной функции. Второй случай может быть описан как случай, в котором количество ресурса, доступного колонии в заданной среде, достаточно невелико. В данном случае жесткие условия, в которых оказывается рассматриваемый организм, вынуждают его специализироваться на различных функциях.
3. Представим, что в модели нет ресурсного ограничения. Тогда все клетки в колонии большого размера будут специализированными, кроме, возможно, одного типа клеток.

4. Представим, что в модели нет ресурсного ограничения. Тогда специализация в случае колонии маленького размера может возникнуть при наличии позиционных эффектов. Кроме того, положительный эффект, оказываемый на возможность возникновения специализации в колониях маленького размера позиционными эффектами, может быть усилен вследствие наличия существенной разницы между важностями различных компонент фитнес-функции в рассматриваемой среде.
5. При рассмотрении модели с внутренними структурными ограничениями в виде неравенств была построена простейшая динамика оптимальных стратегий колонии в рассматриваемом фитнес-ландшафте при простейшей динамике ресурсов. Было выявлено, какие оптимальные состояния соответствуют среде с высоким содержанием ресурса, а какие средам с низким содержанием ресурсов.

ЗАКЛЮЧЕНИЕ

При выполнении проекта получены следующие основные результаты, могут быть сделаны следующие выводы и рекомендации:

- математически поставлены задачи об оценке окупаемости требуемых инвестиций в строительство транспортного узла в какой-либо конкретной точке региона и об оценке срока окупаемости этих инвестиций; показано, что обе задачи, которые формулируются как задачи отыскания максимума функции максимина, сводятся к задаче квадратичного программирования, в то время как отыскание наиболее неблагоприятного сценария формирования рыночных транспортных тарифов на перевозки грузов сводится к решению задач линейного программирования, образующих двойственную пару;
- сформулированы и доказаны проверяемые достаточные условия конкурентоспособности тарифов на услуги транспортного узла в условиях рынка, которые использованы при формулировке систем ограничений в обеих задачах;
- в рамках исследования теории расписаний в турнирах для случаев трех и четырех участников в одном матче построены обобщения стандартного расписания и расписания равных преимуществ; показано, что данные случаи являются наиболее значимыми с практической точки зрения;
- представлена первая версия концепции платформы для исследования проблем и возможностей по кооперации в Арктике на базе моделей конкурентной кооперации в идеологии многоуровневого подхода к декомпозиции сложных кибер-физических систем;
- в рамках теории многокритериального выбора предложен и исследован подход к анализу чувствительности недоминируемых альтернатив к изменению границ интервалов неопределенности замещения, где альтернативы выбираются на основе интервальных данных оценок замещения критериев;
- в рамках анализа способов выбора оптимальных вариантов и процедур ранжирования альтернатив выполнено эмпирическое исследование устой-

- чивости агрегированных рейтингов, строящихся с помощью ординальных методов, основанных на парных сравнениях по правилу большинства;
- в рамках теории функций доверия исследованы меры конфликта, индуцированных конъюнктивными и дизъюнктивными правилами комбинирования свидетельств; рассмотрены условия согласования правил комбинирования с мерами конфликта;
 - в рамках теории функций доверия показано, что в конъюнктивных мерах конфликта можно выделить метрическую и энтропийную составляющие; исследованы свойства изменений этих составляющих при комбинировании свидетельств;
 - в рамках исследования степени манипулируемости правил агрегирования предпочтений предложена и исследована модель индивидуального и коалиционного манипулирования правил агрегирования предпочтений в одномерной модели формирования предпочтений; для трех альтернатив выявлены минимально манипулируемые группы мажоритарных правил;
 - в рамках анализа в банковской сфере рассмотрено применение модели биномиального теста коррелированных случайных бинарных величин для задачи проверки точности прогноза количества дефолтов;
 - проведена оценка влияния мошеннических операций на уязвимость всей финансовой системы; предложен агентно-ориентированный подход к моделированию критических уровней мошеннической деятельности, которые делают финансовую систему уязвимой; этот подход применен для исследования таких типов мошенничества, как финансовая пирамида и скрытый риск;
 - на основе разработанных ранее моделей влияния в сетях предложен эффективный алгоритм вычисления индексов ближних и дальних взаимодействий; оценена вычислительная сложность этого алгоритма;
 - с помощью сетевых моделей проанализированы отношения между странами на различных уровнях взаимодействия, а именно, выполнен сетевой ана-

- лиз: а) международной миграции с 1970 по 2014 гг. на основе данных по потокам миграции между странами; б) торговли продовольствием; в) террористических сообществ;
- проанализирована неоднородность сообщества российских экономистов по результатам анкетирования; предложены несколько методов составления рейтингов научных журналов по экономике с учетом выявленной неоднородности сообщества российских экономистов;
 - в рамках исследования паттернов сформулированы и доказаны свойства порядково-инвариантной паттерн-кластеризации. Эти свойства показывают, что для практической реализации порядково-инвариантной и диффузионно-инвариантной паттерн-кластеризаций могут быть использованы быстрые алгоритмы сортировки данных;
 - в рамках исследования структурной сложности графа предложен новый, вычислительно более мощный вариант частотного алгоритма построения семейства дихотомий произвольного графа, что позволяет вычислить такой показатель сложности графа, как сцепленность; этот алгоритм применен к данным фондового рынка S&P-500 для анализа согласованности его функционирования;
 - проведено экспериментальное сравнение алгоритмов кластеризации, основанных на методе аффинного распространения (Affinity Propagation) с другими алгоритмами кластеризации с точки зрения правильности определения количества кластеров и оценки качества полученного разбиения по сравнению с истинным разбиением; численные эксперименты проведены на данных с различной степенью смешивания кластеров; определена иерархия качества алгоритмов кластеризации, основанных на методе аффинного распространения, для разных типов данных;
 - разработан вычислительный комплекс, позволяющий моделировать процесс распространения нефти на поверхности Баренцева моря на основе агентно-сетевого подхода;

- исследованы явные двухуровневые конечно-разностные схемы решения одномерных уравнений баротропной газовой динамики; линейаризованы схемы на постоянном решении и выведено необходимое условие типа фон Неймана и критерий типа CFL для слабой консервативности в L^2 для соответствующей задачи на всей прямой;
- предложена и исследована новая фитнес-оптимизационная модель, описывающая влияние факторов окружающей среды на возможности возникновения репродуктивно-соматической специализации в колониях клеток.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Отчет МНУЛ анализа и выбора решений, ТЗ-72 за 2017г.

Раздел 1

- 2 Belenky A., Fedin G., Kornhauser A. Finding optimal competitive transshipment tariffs in a regional cargo transport hub// Working paper, Series WP7, Mathematical Methods for Decision Making in Economics, Business, and Politics. - 2018 (prepared for publishing).
- 3 Беленький А., Юнусова А. Теоретико-игровой подход к проблеме регулирования тарифов на транзитные контейнерные перевозки на железных дорогах России// Транспорт: Наука, Техника, Управление. – 2011. - N 6. С. 3 -8.
- 4 Shindin S. Parumasur N. Numerical simulation of a transport fragmentation coagulation mode// Applied Mathematics and Computation. - 2014. - 246. P. 192 - 198.
- 5 Belenky A. A 3-person game on polyhedral sets// Computers & Mathematics with Applications. – 1994. N 28(5). P. 53 - 56.
- 6 SteadieSeifi M., Dellaert N., Nuijten W., Van Woensel T., Raoufi R. Multimodal freight transportation planning: a literature review// European Journal of Operations Research. – 2014. - N 23(1). - P. 1 - 15.
- 7 Kreutzberger E., Konings R. The challenge of appropriate hub terminal and hub-and-spoke network development for seaports and intermodal rail transport in Europe// Research in Transportation Business & Management. – 2016. - N 19. - P. 83 - 96.
- 8 Bingxiang Z., Mingsheng S., Jimin L. Hub-and-spoke construction study of air Ciana// Energy Procedia. Part B. – 2012. - N 16. – P. 971 - 976.
- 9 Serper E., Almur S. The design of capacitated intermodal hub networks with different vehicle types// Transportation Research. Part B: Methodological. – 2016. - N 86. – P. 51 - 65.

- 10 Беленький А. Минимаксные задачи планирования с линейными ограничениями и методы их решения// Автоматика и Телемеханика. - 1981. - N 10. – С. 157 - 170.
- 11 Hotelling H. Stability in competition// The Economic Journal. - 1929. - N 39. – P. 41 - 57.
- 12 Hakimi S.L. On locating new facilities in a competitive environment// European Journal of Operational Research. - 1983. - N 12. – P. 29 - 35.
- 13 Plastria F. Static competitive facility location: An overview of optimization approaches// European Journal of Operational Research. – 2001. - N 129(3). - P. 461 - 470.
- 14 Kress D., Pesch E. Sequential competitive location on networks// European Journal of Operational Research. - 2012. - N 217(3). - P. 483 - 499.
- 15 Ahn H.K., Cheng S.W., Cheong O., Golin M., Van Oostrum R. Competitive facility location: The Voronoi game// Theoretical Computer Science. - 2004. - N 310(1). – P. 457 - 467.
- 16 Ashtiani M.G., Makui A., Ramezani R. A robust model for a leader-follower competitive facility location problem in a discrete space// Applied Mathematical Modelling. - 2013. - N 37. – P. 62 - 71.
- 17 Kucukaydin H., Aras N., Kuban Altinel I. Competitive facility location problem with attractiveness adjustment of the follower: A bilevel programming model and its solution// European Journal of Operational Research. - 2011. - N 208(3). – P. 206 - 220.
- 18 Saidani N., Chu F., Chen H. Competitive facility location and design with reactions of competitors already in the market// European Journal of Operational Research. - 2012. - N 219(1). – P. 9–17.
- 19 Wang, X., Ouyang Y. A continuum approximation approach to competitive facility location design under facility disruption risks// Transportation Research: Part B: Methodological. - 2013. - N 50. – P. 90 - 103.

- 20 Fernandez J., Salhi S., Toth B.G. Location equilibria for a continuous competitive facility location problem under delivered pricing// Computers \& Operations Research. – 2012. - N 41. – P. 185 - 195.
- 21 Fleckinger P., Lafay T. Product flexibility and price competition in Hotelling's duopoly// Mathematical Social Sciences. - 2010. - N 60. – P. 61 - 68.
- 22 Beladi H., Chakrabarti A., Marjit S. Sequential spatial competition in vertically related industries with different product varieties// Economics Letters. - 2010. - N 106. – P. 112 - 114.
- 23 Dasci A., Laporte G. A continuous model for multistore competitive location// Operations Research. - 2005. - N 53. – P. 263 - 280.
- 24 Dasci A. Laporte G. Erratum to "A continuous model for multistore competitive location"// Operations Research. - 2007. - N 55. – P. 1188 - 1190.
- 25 Adjiman C.S., Androulakis I.P., Floudas C.A. Global optimization of MINLP problems in process synthesis and design// Computers and Chemical Engineering. - 1997. - N 21. - P. 445 - 450.
- 26 Zhang Y. Designing a retail store network with strategic pricing in a competitive environment// International Journal of Production Economics. - 2015. - N 159. – P. 265 - 273.
- 27 Marianov V., Serra D., ReVelle C. Location of hubs in a competitive environment// European Journal of Operational Research. - 1999. - N 114. – P. 363 - 371.
- 28 Wagner B. A note on "Location of hubs in a competitive environment"// European Journal of Operational Research. - 2008. - N 184. - P. 57 - 62.
- 29 Eiselt H.A., Marianov V. A conditional p-hub location problem with attraction functions. Computers & Operations Research, - 2009. - N 36. – P. 3128 - 3135.
- 30 Sasaki M. Hub network design model in a competitive environment with flow threshold// Journal of the Operations Research Society of Japan. - 2005. - N 48. - P. 158 - 171.
- 31 Mahmutogullari A.I., Kara B.Y. Hub location under competition// European Journal of Operational Research. - 2016. - N 250. P. 214 - 225.

- 32 Luer-Villagra A., Marianov V. A competitive hub location and pricing problem// European Journal of Operational Research. – 2013. - N 231. - P. 734 - 744.
- 33 Karpov A. A new knockout tournament seeding method and its axiomatic justification// Oper. Res. Let. – 2016. - N 44(6). – P. 706 - 711.
- 34 Vassilevska Williams V. Knockout Tournaments// In H. Moulin, F. Brandt, V. Conitzer, U. Endriss, J. Lang, A. Procaccia (Eds.), Handbook of Computational Social Choice (P. 453 - 474). - 2016. - Cambridge: Cambridge University Press.
- 35 Groh C., Moldovanu B., Sela A., Sunde U. Optimal seedings in elimination tournaments// Econ. Theory. – 2012. – N 49(1). – P. 59 - 80.
- 36 Kräkel M. Optimal seedings in elimination tournaments revisited// Econ. Theory Bull. - 2014. - N 2(1). - P. 77 - 91.
- 37 Rosen S. Prizes and incentives in elimination tournaments// Am. Econ. Rev. – 1986. – N 76(4). – P. 701 - 715.
- 38 Stracke R., Höchtl W., Kerschbamer R., Sunde U. Incentives and selection in promotion contests: Is it possible to kill two birds with one stone?// Man. Dec. Econ. – 2015. - N 36(5). – P. 275 - 285.
- 39 Wei G., Bary B., Qin Y. Sorted or pooled? Optimal tournament design for heterogeneous contestants// Cluster Comp. – 2018. doi: 10.1007/s10586-018-1714-1
- 40 Dagaev D., Suzdaltsev A. Competitive intensity and quality maximizing seedings in knock-out tournaments// J. Comb. Optim. – 2018. - N 35(1). – P. 170 - 188.
- 41 Karpov A. Generalized knockout tournament seedings// Int. J. Comp. Sci. Sport. – 2018. - N 17(2). - in press.
- 42 OEIS. The On-Line Encyclopedia of Integer Sequences. [Электронный ресурс]. - 2018. - URL: <http://oeis.org> (дата обращения 26.10.2018).

- 43 Wright M. OR analysis of sporting rules – A survey// Eur. J. Oper. Res. – 2014. - N 232(1). – P. 1 - 8.
- 44 Schwenk A.J. What is the Correct Way to Seed a Knockout Tournament?// Am. Math. Mon. – 2000. - N 107(2). – P. 140 - 150.
- 45 Dagaev D., Sonin K. Winning by Losing: Incentive Incompatibility in Multiple Qualifiers// J. Sport. Econ. – 2018. – N 19(8). - P. 1122 - 1146.
- 46 Brandenburger A. M., Nalebuff B. J. Co-Opetition. – Crown Business, 2011. – 304 p.
- 47 Sørensen C.T.N. The New Eurasian Geopolitics: Regional Multilateral Initiatives, Cooperation, and Competition// Policy. – 2017. – N 28.
- 48 Young O.R. et al. The Arctic in world affairs: a North Pacific dialogue on international cooperation in a changing Arctic: 2014 North Pacific Arctic Conference proceedings. – Seoul, Korea: Korea Maritime Institute, 2014. – 424 p.
- 49 Pezard S. et al. Maintaining Arctic Cooperation with Russia: Planning for Regional Change in the Far North. – Rand Corporation, 2017. – 100 c.
- 50 Gong K. The Cooperation and Competition between China, Japan, and South Korea in the Arctic// Asian Countries and the Arctic Future. – 2016. – P. 237 - 254.
- 51 Troyakova T. The Cooperation of Russia and Northeast Asian Countries in the Arctic: Challenges and Opportunities// East Asia-Arctic Relations: Boundary, Security and International Politics. – McGill-Queen's Press-MQUP, - 2016. – P. 117 - 125
- 52 Stephenson S.R., Pincus R. Challenges of sea-ice prediction for Arctic marine policy and planning// Journal of Borderlands Studies. – 2018. – N 33(2). – P. 255 - 272.
- 53 Hamilton H.A. et al. Investigating cross-sectoral synergies through integrated aquaculture, fisheries, and agriculture phosphorus assessments: A case study of Norway// Journal of Industrial Ecology. – 2016. – N 20(4). – P. 867 - 881.

- 54 K  p  l   J., Mikkola H. The promise of the geoeconomic Arctic: a critical analysis// Asia Europe Journal. – 2016. – N 14(2). – P. 203 - 220.
- 55 Shvydun S., Aleskerov F. A Mathematical Approach to Conflict Resolution in the Arctic Region// Cybernetics (CYBCONF), 2017 3rd IEEE International Conference. – 2017. – P. 1 - 6.
- 56 d'Aspremont C., Jacquemin A. Cooperative and noncooperative R & D in duopoly with spillovers// The American Economic Review. – 1988. – N 78(5). – P. 1133 - 1137.
- 57 Roelofs M. R.,   stbye S. E., Heen E. E. Asymmetric firms, technology sharing and R&D investment// Experimental Economics. – 2017. – N 20(3). – P. 574 - 600.
- 58 Chirkin V. et al. Gaming Experiments for Analysis of Pricing Mechanisms at Electricity Markets// IFAC-PapersOnLine. – 2016. – N 49(32). – P. 13 - 18.
- 59 Lee E.A. Cyber physical systems: Design challenges// 11th IEEE Symposium on Object Oriented Real-Time Distributed Computing (ISORC). – IEEE. - 2008. – P. 363 - 369.
- 60 Utne I.B. et al. Risk Management in Aquaculture: Integrating Sustainability Perspectives// ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering. – American Society of Mechanical Engineers. - 2017. – C. V07BT06A054-V07BT06A054.
- 61 F  re M. et al. Precision fish farming: A new framework to improve production in aquaculture// Biosystems Engineering. – 2018. – N 173. – P. 176 - 193.
- 62 Buanes A. et al. In whose interest? An exploratory analysis of stakeholders in Norwegian coastal zone planning// Ocean & Coastal Management. – 2004. – N 47(5-6). – P. 207 - 223.
- 63 St  rkersen K.V. Fish first: Sharp end decision-making at Norwegian fish farms// Safety science. – 2012. – N 50(10). – P. 2028 - 2034.
- 64 Pan M., Huntington H.P. A precautionary approach to fisheries in the Central Arctic Ocean: Policy, science, and China// Marine Policy. – 2016. – N 63. – P. 153 - 157.

- 65 Korgin N.A., Korepanov V.O. Experimental Gaming Comparison of Resource Allocation Rules in Case of Transferable Utilities// International Game Theory Review. – 2017. – N 19(02). – P. 1750006.
- 66 Emelyanov S.V., Korovin S.K. Control of complex and uncertain systems: new types of feedback. – Springer Science & Business Media, 2012. – 322 p.
- 67 Burkov V.N. et al. Mechanism Design and Management: Mathematical Methods for Smart Organizations: (for Managers, Academics and Students). – Nova Publishers, 2013. – 190 p.
- 68 Burkov V.N., Novikov D.A., Shchepkin A.V. Control Mechanisms for Ecological-Economic Systems. – Springer International Publishing, 2015. – 163 p.
- 69 Дорри М.Х., Рошин А.А. Инструментальная программно-алгоритмическая система для разработки исследовательских комплексов // Мехатроника, автоматизация, управление. - 2008. – N 12(93). - С. 12 - 17.
- 70 Korgin, N., Rojdestvenskaya S. Concordant Approach for R&D Projects' Evaluation and Ranking// 2017 IEEE 11th International Conference on Application of Information and Communication Technologies (AICT). - N 2. – 2017. - P. 358 - 362.
- 71 Korgin N. A. Introduction to theory of control in organizations for kids via interactive games //IFAC-PapersOnLine. – 2015. – N 48(29). – P. 289 - 294.
- 72 Korgin N., Østbye S. Coopetition in the Arctic: Theory and Experiments// Proceedings of the 18th International Conference on Group Decision and Negotiation. NUA Discussion Papers in Economics and Management. – 2018. - P. 356 - 369.

Раздел 2

- 73 Impact assessment guidelines// http://ec.europa.eu/smart-regulation/impact/commission_guidelines/docs/iag_2009_en.pdf

- 74 Handbook of sensitivity analysis// A. Saltelli, K. Chan, M. Scott (Eds.). - New York: Wiley, 2009.
- 75 Ашманов С.А. Линейное программирование. - М.: Физматлит, 1981.
- 76 Rios-Insua D., French S. A Framework for Sensitivity Analysis in Discrete Multi-Objective Decision-Making// European Journal of Operational Research. - 1991. - N 54. - P. 176 - 190.
- 77 Wolters W.T.M., Mareschal B. Novel Types of Sensitivity Analysis for Additive MCDM Methods// European Journal of Operational Research. - 1995. - N 81. - P. 281 - 290.
- 78 French S. Mathematical programming approaches to sensitivity calculations in decision analysis// Journal of the Operational Research Society. - 1992. - N 43(8) - P. 813 - 819.
- 79 Triantaphyllou E., Sánchez A. A Sensitivity Analysis Approach for Some Deterministic Multi-Criteria Decision Making Methods// Decision Sciences. - 1997. - N 28. - P. 151 - 194.
- 80 Erkut E., Tarimcilar M. On Sensitivity Analysis in the Analytic Hierarchy Process// IMA Journal of Management Mathematics. - 1991. - N 3. - P. 61 - 83.
- 81 Triantaphyllou E. Sensitivity analysis approach for MCDM methods. Ch. 8. Multi-Criteria Decision Making Method: a Comparative Study. - 2000. - New York: Kluwer, 2000.
- 82 Huang Y.- F. Enhancement on Sensitivity Analysis of Priority in Analytic Hierarchy Process// International Journal of General System. - 2002. - N 31. - P. 531 - 542.
- 83 Genc T. Sensitivity Analysis on PROMETHEE and TOPSIS Weights// International Journal of Management and Decision Making. - 2014. - N 13. - P. 403 - 421.
- 84 Vetschera R. Sensitivity Analysis for the ELECTRE Multicriteria Method// Zeitschrift für Operations Research. - 1986. - N 30(4). - P. B99 - B117.
- 85 Tervonen T., Figueira J.R., Lahdelma R., Salminen P. SMAA-III. A Simulation-Based Approach for Sensitivity Analysis of ELECTRE III// In: Real-

Time and Deliberative Decision Making: Application to Emerging Stressors / Linkov I., Ferguson E., Magar V.S. (Eds.). - Springer, 2008. - P. 241 - 253.

- 86 Подиновский В.В. Аксиоматическое решение проблемы оценки важности критериев в многокритериальных задачах принятия решений// Современное состояние теории исследования операций / Под ред. Н.Н. Моисеева. - М.: Наука, 1979. - С.117 - 145.
- 87 Подиновский В.В. Введение в теорию важности критериев в многокритериальных задачах принятия решений: Учебное пособие. - М.: Физматлит, 2007.
- 88 Подиновский В.В., Потапов М.А. Важность критериев в многокритериальных задачах принятия решений: теория, методы, софт и приложения// Открытое образование. - 2012. - N 2. - С. 55 - 61.
- 89 Подиновский В.В. Об относительной важности критериев в многокритериальных задачах принятия решений// Многокритериальные задачи принятия решений / Под ред. Емельянова С.В. - М.: Машиностроение, 1978. - С. 48 - 82.
- 90 Подиновский В.В. Параметрическая важность критериев и интервалы неопределенности замещений в анализе многокритериальных задач// Журнал вычислительной математики и математической физики. - 2008. - N 48(11). - С. 1979 - 1998.
- 91 Подиновский В.В. Анализ устойчивости результатов выбора при частичном отношении предпочтения// Искусственный интеллект и принятие решений. - 2009. - N 4. - С. 45 - 52.
- 92 Podinovski V.V. Sensitivity Analysis for Choice Problems with Partial Preference Relations// European Journal of Operational Research. - 2012. - N 221. - P. 198 - 204.
- 93 Нелюбин А.П. Анализ устойчивости многокритериального выбора методами теории важности критериев при изменении интервальных оценок важности// Открытое образование. - 2012. - N 2. - С. 47 - 51.

- 94 Nelyubin A.P. Criteria importance theory: Sensitivity Analysis of Multicriterial Choice Using Interval Importance Information// American Journal of Control System and Information Technology. - 2013. - N 1. - P. 13 - 17.
- 95 Подиновский В.В. Чувствительность многокритериального выбора к изменению оценок важности неоднородных критериев// Информационные технологии в науке, образовании и управлении. - 2017. - N 4. - С. 23 - 27.
- 96 Passy U., Levanon Y. Analysis of multiobjective decision problems by the indifference band approach// Journal of Optimization Theory and Applications. - 1984. - N 43(2). - P. 205 - 235.
- 97 Меньшикова О.Р., Подиновский В.В. Построение отношения предпочтения и ядра в многокритериальных задачах с упорядоченными по важности неоднородными критериями// Журнал вычислительной математики и математической физики. - 1988. - N 5. - С. 647 - 659.
- 98 Berman V.P., Naumov G.Y., Podinovski V.V. Interval Value Tradeoffs Methodology and Techniques of Multi-Criteria Decision Analysis// User-Oriented Methodology and Techniques of Decision Analysis and Support. - Berlin: Springer, 1993. - P. 144 - 149.
- 99 Меньшикова О.Р., Подиновский В.В. Отношение предпочтения с интервалами неопределенности замещений// Автоматика и телемеханика. - 2007. - N 6. - С. 157 - 165.
- 100 Лотов А.В. Поспелова И.И. Многокритериальные задачи принятия решений: Учебное пособие. - М.: Издательский отдел факультета ВМиК МГУ; МАКС Пресс, 2008.
- 101 Подиновский В.В. Многокритериальные задачи с однородными равноценными критериями// Журнал вычислительной математики и математической физики. - 1975. - N 2. - С. 330 - 344.
- 102 Подиновский В.В., Гаврилов В.М. Оптимизация по последовательно применяемым критериям. - М.: Советское радио. - 1975. - 2-е изд. М.: ЛЕНАНД, 2015.

- 103 Vilkas E. An axiomatic definition of the leximin// European Journal of Political Economy. - 1986. - N 2/4. - P. 455 - 463.
- 104 Подиновский В.В. Задача оценивания коэффициентов важности как симметрически-лексикографическая задача оптимизации// Автоматика и телемеханика. - 2003. - N 3. - С. 150 - 162.
- 105 Podinovski V.V. Interval articulation of superiority and precise elicitation of priorities// European Journal of Operational Research.- 2007. - N 180. - P. 406 - 417.
- 106 Подиновский В.В. Анализ чувствительности многокритериального выбора к изменению интервальных оценок замещений критериев// Журнал вычислительной математики и математической физики. - 2018. - N 58(3). - С. 485 - 494.
- 107 Подиновский В.В. Анализ чувствительности выбора к изменению параметров частичных отношений предпочтений// IX Московская международная конференция по исследованию операций (ORM2018). Москва, 22 – 27 октября 2018 г. Труды. / Отв. ред. Ф.И. Ерешко – Москва: МАКС Пресс, 2018, Т. 2, С. 141 – 144.
- 108 Podinovski V.V. Sensitivity Analysis of Mulcriteria Choice to Changes in Intervals of Value Tradeoffs// Computational Mathematics and Mathematical Physics. - 2018. – N 58(3). - P. 461 - 469.
- 109 Aleskerov F.T., Pislyakov V., Subochev A. Ranking Journals In Economics, Management And Political Science By Social Choice Theory Methods// Working papers by NRU Higher School of Economics. 2014. Series WP BRP "Science, Technology and Innovation", No. WP BRP 27/STI/2014.
- 110 Subochev A., Pislyakov V. With or without h-index? Comparing aggregates of rankings based on seven popular bibliometric indicators// Proceedings of the 23D International Conference on Science and Technology Indicators. Leiden: Universiteit Leiden-CWTS. 2018. P. 1135 - 1143.

- 111 Subochev A., Aleskerov F., Pislyakov V. Ranking journals using social choice theory methods: A novel approach in bibliometrics// Journal of Informetrics. - 2018. - N 12(2). P. 416 - 429.
- 112 Dempster A.P. Upper and lower probabilities induced by multivalued mapping// Ann. Math. Statist. – 1967. N – 38. - P. 325 - 339.
- 113 Shafer G. A mathematical theory of evidence. – Princeton, N.J.: Princeton University Press, 1976.
- 114 Martin A. About Conflict in the Theory of Belief Functions// Belief Functions: Theory and Applications, Advances in Intelligent and Soft Computing, 2012, AISC 164. - P. 161 - 168.
- 115 Destercke S., Burger T. Toward an axiomatic definition of conflict between belief functions// IEEE Transactions on Cybernetics. – 2013. - N 43(2). - P. 585 - 596.
- 116 Bronevich A., Lepskiy A., Penikas H. The Application of Conflict Measure to Estimating Incoherence of Analyst's Forecasts about the Cost of Shares of Russian Companies// Procedia Computer Science. - 2015. N 55. - P. 1113 - 1122.
- 117 Jousselme A.-L., Grenier D., Bossé E. A new distance between two bodies of evidence// Information Fusion, 2001, 2, pp. 91-101.
- 118 Liu W. Analysing the degree of conflict among belief functions// Artificial Intelligence. -2006. - N 170. - P. 909 - 924.
- 119 Jousselme A.-L., Maupin P. Distances in evidence theory: Comprehensive survey and generalizations// International Journal of Approximate Reasoning. – 2012. - N 53. - P. 118 - 145.
- 120 Lepskiy A. About Relation between the Measure of Conflict and Decreasing of Ignorance in Theory of Evidence// Proceedings of the 8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13). Amsterdam - Beijing - Paris: Atlantis Press, 2013. - P. 355 - 362.

- 121 Daniel M. Conflict between Belief Functions: a New Measure Based on their Non-Conflicting Parts// In: F. Cuzzolin (eds.) BELIEF 2014. LNCS (LNAI) vol.8764. - P. 321 - 330. Springer, Heidelberg (2014).
- 122 Schubert J. The internal conflict of a belief function// In: T. Denoeux, M-H. Masson (eds.) BELIEF, 2012, AISC 164. - P. 169 - 177.
- 123 Lepskiy A. On internal conflict as an external conflict of a decomposition of evidence// Vejnarová J., Kratochvil V. (Eds.): BELIEF 2016, LNAI 9861, Springer-Verlag. - P. 25 - 34.
- 124 Bronevich A., Lepskiy A. Imprecision indices: axiomatic, properties and applications// Int. J. Gen. Syst. – 2015. N 44(7–8). - P. 812 - 832.
- 125 Smets Ph. The combination of evidence in the Transferable Belief Model// IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1990. N 12(5). - P. 447 - 458.
- 126 Dubois D., Prade H. On the combination of evidence in various mathematical frameworks// In: Flamm, J., Luisi, T. (eds.) Reliability Data Collection and Analysis. P. 213 - 241 Kluwer Acad. Publ., Dordrecht (1992)
- 127 Lepskiy A. General Schemes of Combining Rules and the Quality Characteristics of Combining // F. Cuzzolin (Ed.): BELIEF 2014, LNAI 8764, Springer-Verlag, 2014, P. 29 - 38.
- 128 Bronevich A.G., Rozenberg I.N. The choice of generalized Dempster-Shafer rules for aggregating belief functions// Int. J. Approx. Reason. – 2015. - N 56. - P. 122 - 136.
- 129 Dubois D., Prade H. A Set-Theoretic View of Belief Functions: Logical Operations and Approximations by Fuzzy Sets// International Journal of General Systems. – 1986. N 12(3). - P. 193 - 226.
- 130 Deza M.M., Deza E. Encyclopedia of distances. – Berlin Heidelberg: Springer; 2009.
- 131 Bronevich A.G., Rozenberg I.N. Conjunctive rules in the theory of belief functions and their justification through decisions models. In J. Vejnarová, V.

Kratochvíl eds. Belief Functions: Theory and Applications. Lecture Notes in Computer Science, 9861, Springer Verlag, Berlin, 2016. - P. 137 - 145.

- 132 Lepskiy A. On the Conflict Measures Agreed with the Combining Rules// in: Destercke S., Denoeux T., Cuzzolin F., Martin A. (eds) Belief Functions: Theory and Applications. BELIEF 2018. Lecture Notes in Computer Science, 2018, vol. 11069. Springer, Cham. - P. 172 - 180.
- 133 Алескеров Ф.Т., Иванов А.А., Карабекян Д.С., Якуба В.И. Манипулируемость мажоритарных правил выбора в модели Даунса формирования предпочтений/ IX Московская международная конференция по Исследованию Операций (ORM-2018 Germeyer-100) Москва, 22-27 октября 2018.
- 134 Aleskerov F., Karabekyan D., Ivanov A., Yakuba V. Individual manipulability of majoritarian rules for one-dimensional preferences/ The Sixth International Conference on Information Technology and Quantitative Management (ITQM 2018), October, 20-21, 2018, Omaha, Nebraska, USA.

Раздел 3

- 135 Bernanke B., Gertler M., Gilchrist S. The financial accelerator and the flight to quality// NBER Working Paper, 4789, 1996.
- 136 Khorasgani A. Evaluation of basel III capital requirement for UK SMEs: Is IRB approach efficient for small firms?// Nantes, France: FEBS 2015 Conference, 2015.
- 137 Engelmann B., Rauhmeier R. The Basel II Risk Parameters. - Berlin: Springer, 2006.
- 138 Hamerle A., Liebig T., Scheule H. Forecasting Credit Portfolio Risk, 2004.
- 139 Hsiao C. Analysis of Panel Data. - Cambridge: Cambridge University Press, 2003.
- 140 Basel II. International Convergence of Capital Measurement and Capital Standards, 2006.

- 141 Nagpal K., Bahar R. Measuring default correlation// Risk. - 2001. - N 14(3).
- P. 129 - 132.
- 142 WP14. Studies on the Validation of Internal Rating Systems, 2005.
- 143 Bahadur R. A representation of the joint distribution of responses to n dichotomous items. B S. H., Studies in Item Analysis and Prediction. - Stanford: Stanford University Press, 1961.
- 144 Lunn A.D., Davies S.J. A note on generating correlated binary variables// Biometrika. - 1998. - N 85(2). – P. 487 - 490.
- 145 Ruseckas J. Probabilistic model of N correlated binary random variables and non-extensive statistical mechanics// Physics Letters A. – 2015. N 379. - C. 654 - 659.
- 146 Vasicek O. Probability of Loss Loan Portfolio// Probability of Loss Loan Portfolio, San Francisco, CA: KMV Corporation, 1987.
- 147 Meng Q. Implications of PD-LGD Correlation in a Portfolio Setting// Moody's Analytics. - 2010.
- 148 Ermolova M.D., Penikas H.I. PD-LGD correlation study: Evidence from the Russian corporate bond market// Model Assisted Statistics and Applications. – 2017. - N 12(4). – P. 335 - 358.
- 149 Lozinskaia A., Merikas A., Merika A., Penikas H. Determinants of the probability of default: the case of the internationally listed shipping corporations// Maritime Policy & Management. - 2017. - N 4(7). – P. 837 - 858.
- 150 Huschens S., Stahl G. A general framework for IRBA backtesting// Bankarchiv. – 2005. N 53. – P. 241 - 248.
- 151 Ермолова М.Д., Пеникас Г.И. Биномиальный тест для коррелированных бинарных случайных величин для проверки точности рейтинговой модели// Управление финансовыми рисками. - 2018. - N 55(03). - С. 174 - 190.

- 152 Sikka P. The corrosive effects of neoliberalism on the UK financial crises and auditing practices: A dead-end for reforms// In Accounting Forum. Elsevier. – 2015. - N 39(1). - P. 1 - 18.
- 153 Ngai E.W.T., Hu Y., Wong Y.H., Chen Y., Sun X. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature// Decision Support Systems. – 2011. - N 50(3). – P. 559 - 569.
- 154 Мамонов М.Е. «Дыры» в капитале обанкротившихся российских банков: старые факторы и новые гипотезы// Экономическая политика. – 2017, N 12(1).
- 155 Мамонов М.Е. «Дыры» в капитале еще не обанкротившихся российских банков: попытка ранней идентификации и оценка масштаба возможных потерь [Электронный ресурс]. - 2017. - URL: http://www.forecast.ru/_ARCHIVE/Presentations/Mamonov/1832017.pdf
- 156 Blass A.A., Grossman R.S. Financial fraud and banking stability: The Israeli bank crisis of 1983 and trial of 1990// International Review of Law and Economics. – 1996. - N 16(4). – P. 461 - 472.
- 157 Niinimäki J.P. Evergreening in banking. Journal of Financial Stability. – 2007. - N 3(4). – P. 368 - 393.
- 158 Páscua M.R., Seghir A. Harsh default penalties lead to Ponzi schemes// Games and Economic Behavior. – 2009. - N 65(1). – P. 270 - 286.
- 159 Tennant D. Why do people risk exposure to Ponzi schemes? Econometric evidence from Jamaica// Journal of International Financial Markets, Institutions and Money. – 2011, - N 21(3). – P. 328 - 346.
- 160 Cortés D., Santamaría J., Vargas J.F. Economic shocks and crime: Evidence from the crash of Ponzi schemes// Journal of Economic Behavior & Organization. – 2016. - N 131. - P. 263 - 275.

Раздел 4

- 161 Freeman L.C. Centrality in social networks: conceptual clarification// Social Networks. – 1979. - N 1. - P. 215 – 239.
- 162 Bonacich P. Technique for Analyzing Overlapping Memberships// Sociological Methodology. – 1972. – N 4. - P. 176 – 185.
- 163 Brin S., Page L. The anatomy of a large-scale hypertextual Web search engine// Comput. Netw. – 1998. - N 30. - P. 107 – 117.
- 164 Katz L. A New Status Index Derived from Sociometric Index// Psychometrika. – 1953. - P. 39 – 43.
- 165 Freeman L.C. A set of measures of centrality based upon betweenness// Sociometry. – 1977. - N 40. - P. 35 – 41.
- 166 Freeman L.C., Borgatti S.P., White D.R. Centrality in valued graphs: A measure of betweenness based on network flow// Soc. Networks. – 1991. - N 13. - P. 141 – 154.
- 167 Newman M.E.J. A measure of betweenness centrality based on random walks// Soc. Networks. – 2005. - N 27. - P. 39 – 54.
- 168 Aleskerov F.T., Andrievskaya I.K., Permjakova E.E. Key borrowers detected by the intensities of their short-range interactions / Working papers by NRU Higher School of Economics. Series FE "Financial Economics". 2014. No. WP BRP 33/FE/2014.
- 169 Aleskerov F.T., Meshcheryakova N.G., Shvydun S.V. Power in Network Structures// in: Models, Algorithms, and Technologies for Network Analysis. Springer Proceedings in Mathematics & Statistics / Ed. by V.A. Kalyagin, A.I. Nikolaev, P.M. Pardalos, O. Prokopyev. N 197. Springer International Publishing, 2017. P. 79 – 85.
- 170 Aleskerov F.T., Meshcheryakova N.G., Shvydun S.V. Centrality Measures in Networks based on Nodes Attributes, Long-Range Interactions and Group Influence / Издательский дом ВШЭ. Series WP7 "Математические методы анализа решений в экономике, бизнесе и политике". 2016.

- 171 Page L., Brin S., Motwani R., Winograd T. The PageRank Citation Ranking: Bringing Order to the Web// Technical report, Stanford Digital Library Technologies Project. 1998.
- 172 Aleskerov F., Meshcheryakova N., Rezyapova A., Shvydun S. Network analysis of international migration// National Research University Higher School of Economics Working Paper, Series WP7 “Mathematical methods for decision making in economics, business and politics”, WP7/2016/06 – 2016.
- 173 Lucas E. Recreations Mathematiques. - Paris, 1882.
- 174 United Nations, Department of Economic and Social Affairs, Population Division, retrieved 30 November 2016. United Nations database, POP/DB/MIG/Flow/Rev.2008.
- 175 United Nations, Department of Economic and Social Affairs, Population Division, retrieved 5 April 2017 <http://www.un.org/en/development/desa/population/migration/data/estimates2/estimates15.shtml>.
- 176 Eurostat, Migration and migrant population statistics, retrieved 26 April 2017, http://ec.europa.eu/eurostat/statistics-explained/index.php/Migration_and_migrant_population_statistics;
- 177 OECD. Stat. International Migration Database. <http://stats.oecd.org/Index.aspx?DataSetCode=MIG>.
- 178 Van Mol C., De Valk H. Migration and immigrants in Europe: A historical and demographic perspective// Integration processes and policies in Europe. – Springer, Cham, 2016. – P. 31 - 55.
- 179 Durand J., Massey D.S., Parrado E.A. The new era of Mexican migration to the United States// The Journal of American History. – 1999. - N 86(2). – P. 518 – 536.
- 180 Fao.org. ESS Website ESS: Food balance sheets [online]. – 2018. - URL: <http://www.fao.org/economic/ess/fbs/en/> [дата обращения 25 октября 2018].
- 181 Wits.worldbank.org. WITS - About WITS. [online]. – 2010. - URL: https://wits.worldbank.org/about_wits.html [дата обращения 19 апреля 2018].

- 182 Unstats.un.org. United Nations Statistics Division - International Merchandise Trade Statistics [online]. – 2008. - URL:
https://unstats.un.org/unsd/tradereport/introduction_MM.asp [дата обращения 17 марта 2018].
- 183 Global Terrorism Database. START: National Consortium for the Study of Terrorism and Responses to Terrorism [online]. – 2009. - URL:
<http://start.umd.edu/gtd/>.
- 184 Aleskerov F. Power indices taking into account agents' preferences// Mathematics and democracy. Springer, Berlin, Heidelberg, P.1 – 18, 2006.
- 185 Aleskerov F., Kalyagin V., Pogorelskiy K. Actual voting power of the IMF members based on their political-economic integration// Mathematical and Computer Modelling. – 2008. - N 48(9-10). – P. 1554 - 1569.
- 186 Aleskerov F., Holler M.J., Kamalova R. Power distribution in the Weimar Reichstag in 1919—1933// Annals of Operations Research. – 2014. - N 215(1). – P. 25 - 37.
- 187 Banzhaf J.F. Weighted Voting Doesn't Work: A Mathematical Analysis// Rutgers Law Rev. – 1965. - N 19. - P. 317 - 343.
- 188 Laruelle A., Valenciano F. Shapley–Shubik and Banzhaf Indices Revisited// Math. Oper. Res. – 2000. – N 26(1). - P. 89 - 104.
- 189 Penrose L.S. The elementary statistics of majority voting// Journal of the Royal Statistical Society. – 1946. - N 109(1). – P. 53 - 57.
- 190 Shvarts D.A. Axiomatics for power indices that account for player preferences// Automation and Remote Control. – 2010. - N 71(1). – P. 128 - 141.
- 191 Khmelnitskaya A., Selcuk O., Talman D. The Shapley value for directed graph games// Operations research letters. – 2016. - N 44(1). – P. 143 - 147.
- 192 Myerson R. Graphs and Cooperative Games// Math. Operations Research. – 1977. - N 2. – P. 225 - 229.
- 193 Jackson M.O. Social and economic networks. – Princeton: Princeton university press, 2010.

- 194 Changat M., Mohandas S., Mulder H.M., Narasimha-Shenoi P.G., Powers R.C., Wildstrom D.J. Axiomatic characterization of the center function. The case of universal axioms// Discrete Applied Mathematics. – 2017. - N 227. – P. 44 - 57.
- 195 Changat M., Mohandas S., Mulder H.M., Narasimha-Shenoi P.G., Powers R.C., Wildstrom D.J. Axiomatic characterization of the center function. The case of non-universal axioms// Discrete Applied Mathematics. – 2018. - N 244. – P. 56 - 69.
- 196 McMorris F.R., Mulder H.M., Novick B., Powers R.C. An ABC-Problem for location and consensus functions on graphs// Discrete Applied Mathematics. – 2016. - N 207. P. 15 - 28.
- 197 Dubey P., Shapley L.S. Mathematical properties of the Banzhaf power index// Mathematics of Operations Research. – 1979. - N 4(2). – P. 99 - 131.
- 198 Балацкий Е.В., Екимова Н.А. Международные рейтинги университетов: практика составления и использования// Экономика образования. – 2012. – N 2. – С. 67 - 80.
- 199 Федорец О.В. Коллективная экспертиза научных журналов: методика агрегирования экспертных оценок и построения рейтинга// Управление большими системами. Сборник трудов. – 2009. – N 27. – С. 18 - 35.
- 200 Алескерев Ф.Т., Писляков В.В., Субочев А.Н., Чистяков А.Г. Построение рейтингов журналов по менеджменту с помощью методов теории коллективного выбора / Высшая школа экономики. Серия WP7 "Математические методы анализа решений в экономике, бизнесе и политике". - 2011. - N 04. – 44с.
- 201 Муравьев А. О научной значимости российских журналов по экономике и смежным дисциплинам// Вопросы экономики. – 2013. – N 4. – С. 130 - 151.
- 202 Алескерев Ф.Т., Писляков В.В., Субочев А.Н. Построение рейтингов журналов по экономике с помощью методов теории коллективного выбо-

- ра / Препринт WP7/2013/03, Нац. исслед. ун-т «Высшая школа экономики». – М.: Изд. дом Высшей школы экономики. – 2013. – N 03. – 48с.
- 203 Алескеров Ф.Т., Катаева Е.С., Писляков В.В., Якуба В.И. Оценка вклада научных работников методом порогового агрегирования // Управление большими системами: сборник трудов. - 2013. - N 44. - С. 172 - 189.
- 204 Балацкий Е.В., Екимова Н.А. Рейтингование участников российского рынка экономических исследований// Journal of Institutional Studies (Журнал институциональных исследований). – 2015. – N 7(3). – С. 102 - 121.
- 205 Балацкий Е.В., Екимова Н.А. Проблема манипулирования в системе РИНЦ// Вестник УрФУ. Серия экономика и управление. 2015. – N 14(2). – С. 166 - 178.
- 206 Подиновский В.В., Потапов М.А. Метод взвешенной суммы критериев в анализе многокритериальных решений: Pro et contra// Бизнес-информатика. – 2013. – N 3(25) . – С. 41 - 48.
- 207 Мазалов В.В., Печников А.А., Фалько И.А. О построении рейтинга научных журналов// Управление большими системами. Сборник трудов. – 2009. – N 27. – С. 47 - 52.
- 208 Чеботарев П.Ю. О рейтинге журналов и агрегировании неполных балльных оценок// Управление большими системами. Сборник трудов. – 2009. – N 27. – С. 81 - 90.
- 209 Миркин Б.Г. О статье О.В. Федорца «Коллективная экспертиза научных журналов: методика агрегирования экспертных оценок и построения рейтинга»// Управление большими системами. Сборник трудов. – 2009. – N 27. – С. 53 - 58.
- 210 Shvydun S.V. Normative properties of multi-criteria choice procedures and their superpositions / Издательский дом ВШЭ. Series WP7 "Математические методы анализа решений в экономике, бизнесе и политике". No. WP7/2015/07. – 2015.

- 211 Белянин А.В., Бессонов В.А. О российской экономической науке и научном сообществе// Экономический журнал ВШЭ. – 2011. – N 15(2). – С. 265 - 268.
- 212 Мальцев А. Российское сообщество экономистов: особенности и перспективы// Вопросы экономики. – 2016. – N 11. – С. 135 - 158.
- 213 Рубинштейн А.Я., Бураков Н.А., Славинская О.А. Сообщество экономистов и экономические журналы (социологические измерения VS библиометрии): Научный доклад. – М.: Институт экономики РАН, 2017. – 83 с.
- 214 Lazarsfeld P.F. The logical and mathematical foundation of latent structure analysis & The interpretation and mathematical foundation of latent structure analysis// S.A. Stouffer et al. (eds.), Measurement and Prediction. Princeton, NJ: Princeton University Press. – 1950. – P. 362 - 472.
- 215 Linzer D.A., Lewis J.B. polCA: An R Package for Polytomous Variable Latent Class Analysis// Journal of Statistical Software. – 2011. – N 42(10). – P. 1 - 29.
- 216 Inselberg A., Dimsdale B. Parallel Coordinates for visualizing multi-dimensional geometry// Computer Graphics. – 1987. – P. 24 - 44.
- 217 Мячин А.Л. Анализ паттернов: порядково-инвариантная паттерн-кластеризация// Управление большими системами. – 2016. – N 61. – С. 41 - 59.
- 218 Myachin A. New methods of pattern analysis in the study of Iris Anderson-Fisher Data // 6th International Conference on Computers Communications and Control. – 2016. – P. 97 - 102.
- 219 Copeland A.H. A reasonable social welfare function (mimeo) / University of Michigan, Ann Arbor (Seminar on Application of Mathematics to the Social Sciences). – 1951.
- 220 Структура российского сообщества экономистов и его отношение к российским экономическим журналам [Электронный ресурс]. – 2018. –

URL: <https://www.hse.ru/DeCAn/RussianEconomicCommunity> (дата обращения 23.11.2018)

- 221 Егорова Л.Г., Мячин А.Л. Структура российского сообщества экономистов и его отношение к российским экономическим журналам (часть 1): анализ методами латентных классов и теории коллективного выбора// (находится на повторной рецензии в журнале «Проблемы управления»).
- 222 Егорова Л.Г., Мячин А.Л. Структура российского сообщества экономистов и его отношение к российским экономическим журналам (часть 2): анализ паттернов респондентов// (находится на повторной рецензии в журнале «Проблемы управления»).
- 223 Egorova L., Myachin A. The Structure of the Russian Academic Economic Community and Its Opinion on Russian Economic Journals// (подан на рецензию в серию препринтов ВШЭ WP7).
- 224 Aleskerov F., Alper C. Inflation, Money, and Output Growths: Some Observations// Bogazici University Research Paper, No.SBE 96-06, 1996.
- 225 Aleskerov F., Alper C. A clustering approach to some monetary facts: a long-run analysis of cross-country data// The Japanese Economic Review. – 2000. – N 51(4). - P. 555 - 567.
- 226 Aleskerov F., Ersel H., Gundes C., et al. Environmental Grouping of the Bank Branches ant their Performances// Yapi Credit Bank. Discussion Paper Series No 97-03, 1997, 24 p.
- 227 Aleskerov F., Ersel H., Yolalan R. Multicriterial Ranking Approach for Evaluating Bank Branch Performance// International Journal of Information Technology and Decision Making. – 2004. – N 3(2). - P. 321 - 335.
- 228 Aleskerov F., Ersel H, Gundes C., et al. A Multicriterial Method for Personnel Allocation among Bank Branches// Yapi Kredi Discussion Paper Series, Istanbul, 1998, No 98–01.
- 229 Aleskerov F., Gokhberg L., Egorova L., Myachin A., Sagieva G. Pattern Analysis in the Study of Science, Education and Innovative Activity in Russian Regions// Procedia Computer Science. – 2013. – N 17. - P. 687 - 694.

- 230 Алескеров Ф.Т., Белоусова В.Ю., Егорова Л.Г., Миркин Б.Г. Анализ паттернов в статике и динамике, часть 1: обзор литературы и уточнение понятия// Бизнес-информатика. - 2013. – N 3(25). - С. 3 - 18.
- 231 Inselberg A. Parallel Coordinates: Visual Multidimensional Geometry and Its Applications. - Springer, 2009, 554 p.
- 232 Mirkin B.G. Clustering for Data Mining: A Data Recovery Approach. - Taylor and Francis Group, 2005, 350 p.
- 233 Mirkin B.G. Summary and semi-average similarity criteria for individual clusters// in: Models, Algorithms, and Technologies for Network Analysis, Ed. by B.I. Goldengorin, V.A. Kalyagin, P.M. Pardalos, Vol.59, NY: Springer, 2013, P. 101 - 126.
- 234 Мячин А.Л. Анализ паттернов: порядково-инвариантная паттерн-кластеризация// Управление большими системами: сборник трудов. - 2016. - N 61. - С. 41 - 59.
- 235 Программа секций XIX Апрельской конференции 2018, p.113. URL: <https://conf.hse.ru/pubs/share/direct/220103399> (дата обращения: 02.11.2018)
- 236 Мячин А.Л. Анализ паттернов в системе параллельных координат на базе парного сравнения показателей// Автоматика и телемеханика (в печати)
- 237 Rubchinsky A. Divisive-Agglomerative Classification Algorithm Based on the Minimax Modification of Frequency Approach// Working paper WP7/2010/07. – М.: NRU HSE, 2010. – 48 p.
- 238 Rubchinsky A. Divisive-Agglomerative Algorithm and Complexity of Automatic Classification Problems// Working paper WP7/2015/09. – М.: NRU HSE, 2015. – 44 p.
- 239 Rubchinsky A. Family of Graph Decompositions and Its Applications to Data Analysis// Working paper WP7/2016/09. – М.: NRU HSE, 2016. – 60 p.
- 240 Rubchinsky A. A New Approach to Network Decomposition Problems// In: Kalyagin V., Nikolaev A., Pardalos P., Prokopyev O. (eds) Models, Al-

- gorithms, and Technologies for Network Analysis. NET 2016. Springer Proceedings in Mathematics & Statistics, 197. Springer, Cham, 2017.
- 241 Rubchinsky A. Graph Dichotomy Algorithm and Its Applications to Analysis of Stocks Market// In: Kalyagin V., Pardalos P., Prokopyev O., Utkina I. (eds) Computational Aspects and Applications in Large-Scale Networks. NET 2017. Springer Proceedings in Mathematics & Statistics, 247. Springer, Cham, 2018.
- 242 Jain A.K. Data clustering: 50 years beyond k-means// Pattern recognition letters. – 2010. - N 31(8). – P. 651 - 666.
- 243 Bishop C.M.. Pattern Recognition and Machine Learning (Information Science and Statistics). - Springer-Verlag, Berlin, Heidelberg, 2006.
- 244 Frey B.J., Dueck D. Clustering by passing messages between data points// Science. – 2007. - N 315(5814). – P. 972 - 976.
- 245 Chiang M.M.-T., Mirkin B. Intelligent choice of the number of clusters in k-means clustering: an experimental study with different cluster spreads// Journal of classification. – 2010. - N 27(1). – P. 3 – 40.
- 246 Tasoulis S.K., Tasoulis D.K., Plagianakos V.P. Enhancing principal direction divisive clustering// Pattern Recognition. – 2010. - N 43(10). – P. 3391 - 3411.
- 247 Dueck D. Affinity propagation: clustering data by passing messages. - Citeseer, 2009.
- 248 Chehdi K., Soltani M., Cariou C. Pixel classification of large-size hyperspectral images by affinity propagation// Journal of applied remote sensing. – 2014. - N 8(1). – P. 083567.
- 249 Guan R., Shi X., Marchese M., Yang C., Liang Y. Text clustering with seeds affinity propagation// IEEE Transactions on Knowledge and Data Engineering. – 2011. - N 23(4). – P. 627 - 637.
- 250 Shang F., Jiao L.C., Shi J., Wang F., Gong M. Fast affinity propagation clustering: A multilevel approach// Pattern recognition. – 2012. - N 45(1). – P. 474 - 486.

- 251 Yang C., Bruzzone L., Guan R., Lu L., Liang Y. Incremental and decremental affinity propagation for semisupervised clustering in multispectral images// IEEE Transactions on Geoscience and Remote Sensing. – 2013. - N 51(3). – P. 1666 – 1679.
- 252 Kschischang F.R., Frey B.J., Loeliger H.-A. Factor graphs and the sum-product algorithm// IEEE Transactions on Information Theory. – 2001. - N 47(2). – P. 498 - 519.
- 253 Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. - Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.
- 254 Givoni I.E., Frey B.J. A binary variable model for affinity propagation// Neural Computation. – 2009. - N 21(6). – P. 1589 - 1600.
- 255 Heskes T. Stable fixed points of loopy belief propagation are local minima of the bethe free energy// In Advances in neural information processing systems. – 2003. – P. 359 - 366.
- 256 Jia H., Ding S., Meng L., Fan S. A density-adaptive affinity propagation clustering algorithm based on spectral dimension reduction// Neural Computing and Applications. – 2014. – N 25(7-8). – P. 1557 - 1567.
- 257 Li P., Ji H., Wang B., Huang Z., Li H. Adjustable preference affinity propagation clustering// Pattern Recognition Letters. – 2017. - N 85. – P. 72 - 78.
- 258 Chen D.-W., Sheng J.-Q., Chen J.-J., Wang C.-D. Stability-based preference selection in affinity propagation// Neural Computing and Applications. - 2014. - N 25(7-8). – P. 1809 – 1822.
- 259 Mirkin B. Clustering: A data recovery approach. - CRC Press Book, 2012.
- 260 Kovaleva E.V., Mirkin B.G. Bisecting k-means and 1d projection divisive clustering: A unified framework and experimental comparison// Journal of Classification. – 2015. – N 32(3). – P. 414 - 442.

- 261 Kodinariya T.M., Makwana P.R. Review on determining number of cluster in k-means clustering// Int. J. Adv. Res. Comput. Sci. Manage. Stud. - 2013. - N 1(6). – P. 90 - 95.

Раздел 5

- 262 Pavlenko V.I., Muangu J., Korobov V.B., Lokhov A.S. Actual problems of prevention and liquidation of oil spills in the Arctic and methods of environmental damage to coastal areas assessment (in Russian)// Arctic Ecol. Econ. – 2015 – N 3(19) – P. 4 - 11.
- 263 Stanovoy V.V., Lavrenov I.V., Neelov I.A. Oil Spill Simulation System for the Arctic Seas// Problems of the Arctic and Antarctic. – 2007 – N. 77.
- 264 Ovsienko S.N., Zatsepa S.N., Ivchenko A.A. Numerical Simulation as Data Support Feature for Decision Making when Selecting Strategy for Marine Environment Protection from Oil Pollution (in Russian)// Hydrometeoizdat – 2011 – N. 213.
- 265 Bakhvalov N.S., Zhidkov N.P., Kobelkov G.M. Numerical methods. – Binom, Moscow – 2011.
- 266 Coulombel J.-F. Stability of finite difference schemes for hyperbolic initial boundary value problems// HCDTE Lecture Notes. Part I. Dispersive and Transport Equations, G. Alberti, F. Ancona, S. Bianchini et al., eds // Amer. Inst. Math. Sci., Springfield. – 2013 – P. 97 - 226.
- 267 Ganzha V.G., Vorozhtsov, E.V. Computer-aided analysis of difference schemes for partial differential equations. – Wiley, New York. – 1996.
- 268 Elizarova T.G. Quasi-gas dynamic equations. – Springer, Dordrecht. – 2009.
- 269 Chetverushkin B.N. Kinetic schemes and quasi-gas dynamic system of equations. – CIMNE, Barcelona – 2008.
- 270 Elizarova T.G., Bulatov O.V. Regularized shallow water equations and a new method of simulation of the open channel flows// Comput. Fluids. – 2011. – N. 46. – P. 206 - 211.

- 271 Zlotnik A., Gavrilin V. On a conservative finite-difference method for 1D shallow water flows based on regularized equations// Math. Problems in Meteorological Model. Math. In Industry, A. B´atkai, P. Csom´os, A. Hor´anyi, etc., eds. – 2016. – N. 24. – P. 16 - 31.
- 272 Zhilyakova L.Y. The application of resource networks for modeling of distribution of substances in the aquatic environment (in Russian). – 2011. – N. 2. – P. 46 - 51.
- 273 Aleskerov F. Taxation for improving regional ecological situation// Ecological Economics of Sustainability Conference, Washington DC, World Bank. – 1990.
- 274 Demin S.S., Aleskerov F.T. Modelling Possible Oil Spills in the Barents Sea and their Consequences// Dynamics of Disasters, ed. by P.M. Pardalos, I. S. Kotsireas, A. Nagurney. Springer International Publishing. – 2018. – N. 2. – P. 47 - 56.
- 275 Злотник А.А., Ломоносов Т.А. Об условиях L^2 -диссипативности линеаризованных явных КГД-разностных схем для уравнений одномерной газовой динамики// Доклады Академии Наук. Математика. – 2018. – N. 482(4).
- 276 Zlotnik A., Lomonosov T. On conditions for weak conservativeness of regularized explicit finite-difference schemes for 1D barotropic gas dynamics equations// Differential and Difference Equations with Applications. – 2018. – N. 230. – P. 635 - 647.
- 277 Zlotnik A.A., Lomonosov T.A. On conditions for L^2 -dissipativity of linearized explicit QGD finite-difference schemes for one-dimensional gas dynamics equations// Doklady Mathematics. – 2018. – N. 98(2). – P. 458 - 463.
- 278 Suhomozgii A.A., Sheretov Yu.V. Stability analysis of a finite-difference scheme for solving the Saint-Venant equations in the shallow water theory// Appl. of functional anal. in approx. theory, Tver State Univ. – 2013. – P. 48 - 60.

- 279 Simpson C. The evolutionary history of division of labour// Proc. R. Soc. London B. - 2011. - N 279. - P. 116 - 121.
- 280 Rueffler C., Hermisson J., Wagner G.P. Evolution of functional specialization and division of labor// PNAS. - 2012. - N 109. - P. 326 - 335.
- 281 Ispolatov I., Ackermann M., Doebeli M. Division of labour and the evolution of multicellularity// Proc. R. Soc. London B. - 2012. - N 279. - P. 1768 - 1776.
- 282 Willensdorfer M. On the evolution of differentiated multicellularity// Evolution. - 2009. - N 63. - P. 306 - 323.
- 283 Solari C.A., Kessler J.O., Goldstein R.E. Motility, mixing, and multicellularity// Gen. Program. Evolv. Mach. - 2007. - N 8. P. 115 - 129.
- 284 Michod R.E., Viossat Y., Solari C.A., Hurand M., Nedelcu A.M. Life-history evolution and the origin of multicellularity// J. Theor. Biol. - 2006. - N 239. - P. 257 - 272.
- 285 Koufopanou V., Bell G. Soma and germ - an experimental approach using Volvox// Proc. R. Soc. London B. - 1993. - N 254. P. 107 - 113.
- 286 Koufopanou V. The evolution of soma in the Volvocales// Am. Nat. 1994. - N 143. - P. 907 - 931.
- 287 Miller S.M. Volvox, chlamydomonas, and the evolution of multicellularity// Nat. Educ. - 2010. - N 3. - P. 65.
- 288 Rossetti V., Schirrmeister B.E., Bernasconi M.V., Bagheri H.C. The evolutionary path to terminal differentiation and division of labor in cyanobacteria// J. Theor. Biol. - 2010. - N 262. - P. 23 - 34.
- 289 Rossetti V., Bagheri H.C. Advantages of the division of labour for the long-term population dynamics of cyanobacteria at different latitudes// Proc. R. Soc. London B. - 2012. - N 279. - P. 3457 - 3466.
- 290 Gavrillets S. Rapid Transition towards the Division of Labor via Evolution of Developmental Plasticity// PLoS Comput. Biol. - 2010. - N 6.

- 291 Bossert W., Qi C.X., Weymark J.A. Extensive social choice and the measurement of group fitness in biological hierarchies// Biol. Phil. - 2013. - N 28. - P. 75 - 98.
- 292 Okasha S. Evolution and the levels of selection. – Oxford: Oxford University Press, 2006.
- 293 Okasha S. Individuals, groups, fitness and utility: multi-level selection meets social choice theory// Biol. Phil. - 2009. - N 24. - P. 561 - 584.